The invasive mussel that could devastate the Amazon p. 390

How the elephant lost its tusks pp. 394 & 483

Debunking climate claims from oil and gas companies p. 405

Science

$15
22 OCTOBER 2021
science.org

AAAS

MOLDABLE WOOD

Transforming boards into versatile 3D shapes

pp. 400 & 465

Gordon Research Conferences

New Conference Application Process

Join the over 400+ GRC communities and experience a high-quality scientific conference experience while also advancing the frontiers in your field!

New Conferences in 2022

- Additive Manufacturing of Soft Materials
- Biology of Winter
- Calcium Signaling
- Cell-Cell Fusion
- Computational Materials Science and Engineering
- Electrochemical Interfaces in Energy Conversion and Storage
- Fluids in Disease Transmission and Contamination
- Frontal Cortex
- Genetic Biocontrol
- Graft Preservation in Heart Transplantation
- Hypothalamus
- Immunoengineering
- Immunometabolism in Health and Disease
- In Vivo Ultrasound Imaging
- Lipidomics
- Multiscale Mechanochemistry and Mechanobiology
- Neural Mechanisms of Acoustic Communication
- Radionuclide Theranostics for the Management of Cancer
- Single-Cell Cancer Biology
- Single-Cell Genomics
- Systems Aging
- Unconventional Semiconductors and Their Applications

Applications for 2022 conferences are being accepted at www.grc.org

Is your scientific community interested in gathering as a small group to discuss cutting-edge, frontier science while also experiencing high-quality, intimate and informal interactions and networking opportunities with peers from around the world? Consider submitting a proposal to initiate a brand-new GRC!

GRC is always interested in receiving new proposals for cutting-edge topics not already covered by the GRC portfolio.

The GRC Board is particularly interested in proposals from engineers and computer scientists!

Proposals are due on an annual basis on June 1 and interested individuals should reach out to Beth Mancuso (**bmancuso@grc.org**) for more information!

Workshop speakers. Top row (L to R): Yang Zhang, Qiangfeng Zhang, and Peilong Lu. Bottom row (L to R): Dek Woolfson, Nagarajan Vaidehi, Haipeng Gong.

Tsinghua-*Science* Workshops: Computational Structural Biology

Improvements in modeling and computer tools for predicting the shapes of proteins were the main topics of discussion in a recent series of online workshops organized by Tsinghua University and Science/*AAAS.*

The composition and shape of proteins determine their functions. From protein folding to receptor binding and beyond, scientists have searched for ways to predict such structures and interactions. More than ever, modern computational tools provide increasing accuracy in predicting the shape of biological structures.

On June 18, 2021, two scientists–computational experts Yang Zhang of the University of Michigan in the United States and Qiangfeng Zhang of Tsinghua University in China–discussed ways to predict molecular structures. Yang Zhang started by explaining how computer simulations can be used to predict the locations of the atoms in a protein based on its amino-acid sequence. He described some of the major achievements in this research, including advances in template-based and template-free modeling, especially with the use of neural networks.

Next, Qiangfeng Zhang expanded on the application of solutions based on artificial intelligence (AI). In particular, he described how to use AI in analyzing data from cryo-electron microscopy (cryo-EM). Most importantly, he explained how combining computational tools with the growing understanding of structural biology helped to create cryoNet, an AI-based method that automatically models data from cryo-EM analysis.

On July 16, biologist Peilong Lu of Westlake University in China opened a session on protein design with a talk about transmembrane proteins. As he noted, when a protein includes more than one region that crosses the membrane, the design gets especially challenging. He explained how he and his colleagues applied computational design to homodimers and tetramers that include transmembrane sections. In one example, he noted the similarity between the model and the actual transmembrane protein's crystal structure. He even showed functional activities from designed proteins that form a pore in the cell membrane.

Following Lu, chemist and biochemist Dek Woolfson of the University of Bristol in the United Kingdom explained how he takes inspiration from naturally existing protein structures to make novel proteins. Using computational tools, he designs unknown protein structures that fold into stable positions, and then tries to make them in the lab. Describing these novel proteins, he pointed out that chemically and physically it should be possible to create these structures, but nature just doesn't make them, as far as anyone knows. In the previous presentation, Lu described his lab's toolkit, which can be used to make these so-called "dark-matter" protein structures.

The last session of the workshops on computational structural biology took place on August 20. During this session, quantum chemist Nagarajan Vaidehi of U.S.-based Beckman Research Institute of the City of Hope, a National Cancer Institute-designated comprehensive cancer center, and biophysicist Haipeng Gong of Tsinghua University, talked about protein dynamics. Vaidehi kicked off this session by reviewing some of her work on G protein-coupled receptors (GPCRs). She explained the transient nature of the G protein or β-arrestin binding to a GPCR. Using data from fluorescence resonance energy transfer (FRET) experiments plus molecular-dynamics (MD) simulations, Vaidehi showed that she can design ligands biased to a specific GPCR-based signaling pathway, which might be used to make therapeutics with reduced side effects.

Gong gave the workshop's last presentation, which examined models to predict the distances and contacts between a protein's amino-acid residues. These models–AmoebaContact and GANProDist–rely on various methods, including the automated neural architecture search, to enhance the prediction of protein structures. Perhaps most intriguing, Gong's methods can be applied to conformational transitions–moving rather than static protein structures.

These experts unwrapped a variety of new approaches to modeling protein structures, as well as elucidating some newly discovered functional interactions between proteins. Beyond the basic biological intrigue of their work, it offers considerable promise for the future of medicine.

Sponsored by

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

IMAGES PROVIDED BY TSINGHUA UNIVERSITY

Managing neurological disorders with neuromodulation

Parkinson's disease and epilepsy, chronic pain, and overactive bladder: These disorders and more can be improved in some patients through precise adjustments to the activity of specific areas of the central nervous system. Beijing PINS Medical Co., Ltd. provides a range of neuromodulating tools to do just that.

Dr. James Wang
Vice President, Beijing Tsinghua Changgung Hospital

Dr. Chaoshi Niu
Head of Neurosurgery Department, University of Science and Technology of China

Dr. Xi Wu
Deputy Director, Neurosurgery Department, Shanghai Changhai Hospital

The activity of specific nerves working in networks determines a person's neurological capabilities and health. Abnormal or unbalanced activity in certain nuclei or pathways contributes to a range of neurological conditions, including chronic pain and neurodegeneration. More and more, neurosurgeons have made the transition from previous functional—but destructive—neurosurgery to neuromodulation. With recent advances in neuromodulation technology, clinicians can now treat a wide range of neurological disorders, thereby improving the lives of many patients.

"Neuromodulation is a technique that allows the in vitro adjustment of parameters through technical means," says Chaoshi Niu, head of the neurosurgery department at the University of Science and Technology of China (USTC) and director of the neurosurgery department of the First Affiliated Hospital of USTC. "Compared with the original destruction and excision [caused by neurosurgery], this technique is minimally invasive, nondestructive, modifiable, and does not cause permanent side effects or affect other subsequent treatments."

In addition, some technology allows remote control of neuromodulation. According to neurosurgeon Xi Wu of Changhai Hospital in Shanghai, remote programming of a neurostimulator provides "strong support for regular follow-ups." This feature also reduces patients' travel and "guarantees timely monitoring of symptom changes," Wu explains.

Methods of stimulation

Neurosurgeons use modulation in several ways: deep brain stimulation (DBS), vagus nerve stimulation (VNS), spinal cord stimulation (SCS), and sacral neuromodulation (SNM), among others.

Current surgical indications for DBS are Parkinson's disease, intractable essential tremor, and dystonia, says neurosurgeon James Wang, vice president of Beijing Tsinghua Changgung Hospital. "Another clinically useful DBS application is a closed-loop DBS for seizure control, where a biosensor detects the seizure spikes on an EEG [electroencephalogram], which will trigger the DBS—placed in or around the seizure focus—to desynchronize the seizure discharge," Wang notes.

In patients who experience severe seizures that cannot be addressed with a surgical resection and some pediatric patients, VNS—electrical stimulation of the left vagus nerve, which travels from the brain along the side of the neck—can be used. VNS can also be used to treat severe depression when medical treatments fail.

In patients with severe pain, SCS can be applied. Potential patients for this treatment, Wang says, include "failed back-surgery patients, postherpes patients, and post nerve or spinal-cord injury patients, among others." He adds, "Many experimental studies are currently ongoing to see whether SCS can play a role to facilitate waking comatose patients or improve neurofunctions after spinal-cord injury."

PHOTOS: PROVIDED BY NOMIS

For patients who suffer from nonobstructive urinary frequency, urgency, and incontinence, physicians could try SNM. As Wang notes, "Clinical trials using SNM to treat intractable constipation and perineal pain are underway."

The most common uses of neuromodulation vary around the world. "Half of the neurostimulator market in the U.S.A. involves SCS, whereas the most frequently used neurostimulator in China is DBS for Parkinson's disease," Wang states.

Transforming the technology

With hardware improvements in neurostimulation technologies, ever more accurate and precise modulation can be performed. "With the modern neurosurgical stereotaxic technique, we can basically place the DBS anywhere we want in the brain, accurately and safely, and give stimulation with different amplitudes and frequencies," Wang explains. "We can finely tune a specific function this way—much like on a radio where once you find a station, you can then turn the volume knob up and down to adjust the sound."

In addition to DBS devices, PINS Medical produces VNS, SCS, and SNM devices. These include implantable pulse generators, electrodes, external programming devices, and more.

When asked about the advantages of using technology from PINS Medical, Wang says, "The PINS stimulator has several key features that make it safer and more convenient to use for both physicians and patients." For one thing, he points out that PINS makes electrodes and pulse generators compatible with 3 Tesla magnetic resonance imaging (3T MRI). PINS' pulse generator can be turned on and off during a functional MRI study, and as a result, Wang explains, "stimulation-induced functional MRI changes can be revealed and studied."

The PINS stimulator includes a battery that can be recharged for over 20 years, which means most patients will probably never need a battery replacement.

The PINS neuromodulator line also comes with features that enhance some applications. For instance, the PINS pulse generator can provide variable frequency stimulation that "was found to be important for certain Parkinson's symptoms, such as relieving a patient's freezing gait," Wang says. The electronics in a PINS system can even detect electrocardiogram signals and sudden acceleration in three dimensions, which Wang says, "can be used to remotely monitor elderly patients for cerebral or cardiac events or falls."

Taming tremors

Recently, Wang treated a 65-year-old man who had suffered from tremors in both of his hands and arms for more than 20 years. The tremors got even worse when this man tried to eat, drink, or write. He had tried physical therapy and taking oral beta blockers with some—but very little—success.

Wang's team implanted bilateral PINS electrodes in the ventral intermediate nucleus of the patient's thalamus. "Once the stimulator was turned on, his hand tremor was nicely suppressed by more than 90 percent," Wang exclaims. "The patient is very satisfied with this outcome, and he has become more socially active, according to his wife."

Similarly, a woman from Qinghai Province suffered from involuntary movements after a car accident when she was 18 years old. After two decades of dystonia, she received implants with PINS electrodes in two areas of her brain. "With the stimulation of PINS DBS neuromodulation stimulator model G102RS, the patient's dystonia has been significantly improved," says Wu. "The patient and her family members were all excited and crying in front of me because of the remarkable clinical effect. They expressed regret that they did not know about this treatment earlier."

These patients and many others live higher-quality lives through neuromodulation, which will continue to improve. "The developmental direction of neuromodulation will focus mainly on two aspects: the exploration of surgical indications and targets and the upgrading of equipment hardware and software," Niu says. "These two aspects will also promote each other." For example, DBS could be extended to a range of new indications, possibly even Alzheimer's disease, according to Niu. Consequently, advances in the technology will drive even more applications of these methods—all with improving results.

Sponsored by

PINS 品驰

PHOTOS: PROVIDED BY NOMIS

CONTENTS

22 OCTOBER 2021 • VOLUME 374 • ISSUE 6566

394 & 483

NEWS

IN BRIEF

380 News at a glance

IN DEPTH

382 Polio vaccination to resume across Afghanistan
Success depends on shoring up the country's health system *By L. Roberts*

383 Red blood cells may be immune sentinels
Oxygen-carrying cells also capture DNA from pathogens and damaged cells *By M. Leslie*
SCIENCE TRANSLATIONAL MEDICINE
RESEARCH ARTICLE BY L. K. M. LAM *ET AL.*
10.1126/SCITRANSLMED.ABJ1008

384 Ancient DNA reveals long-sought homeland of modern horses
Bronze Age herders began riding the ancestor of modern horses in the grassy western Eurasian steppe *By A. Gibbons*

385 Misuse of funds leads to shutdown of top primate center
Kyoto University's move follows dismissal of director *By D. Normile*

386 Anthropologists grapple with identifying remains by race
Professional group is developing new standards for linking human remains to specific populations instead of races *By L. Wade*

387 What tore apart *The Lancet*'s COVID-19 origin probe?
Former members cite confidentiality dispute and challenge concerns over conflicts of interest that derailed effort *By J. Cohen*

388 Hunt begins for ancient Antarctic ice
Six teams plan to drill cores that could shed light on warm climates 1.5 million years ago
By P. Voosen

FEATURES

390 A golden menace
An invasive mussel is devastating cosystems as it spreads through South American rivers, threatening the Amazon basin
By S. Moutinho

INSIGHTS

PERSPECTIVES

394 Of war, tusks, and genes
Societal conflict leaves an evolutionary signature in wildlife
By C. T. Darimont and F. Pelletier
REPORT p. 483

396 Iron-catalyzed coupling of three reactants
Vinyl boronates couple under mild conditions to two partners through carbon-carbon bonds *By G. Lefèvre*
RESEARCH ARTICLE p. 432

397 Toward collective animal neuroscience
Social intelligence requires the study of both groups of brains and the individual brain
By J. Sliwa
RESEARCH ARTICLES pp. 421 & 422

399 Bacterial defense islands limit viral attack
Mobile phage-defense genes underpin rapid evolution of resistance in nature
By S. Meaden and P. C. Fineran
REPORT p. 488

400 Step aside, aluminum honeycomb
A two-step process can turn hardwood into a strong and flexible moldable material
By M. Tajvidi and D. J. Gardner
REPORT p. 465

402 An improved water-harvesting cycle
Tuning metal-organic frameworks at the atomic level helps release trapped water molecules
By L. Öhrström and F. M. Amombo Noa
REPORT p. 454

403 Tackling the pandemic with (biased) data
Data are crucial for understanding and addressing the pandemic, but there are pitfalls *By C. Pagel and C. A. Yates*

POLICY FORUM

405 How ambitious are oil and gas companies' climate goals?
Despite many claims, the sector is not on track *By S. Dietz* et al.

PHOTO: ELEPHANTVOICES

BOOKS *ET AL.*

410 The myth of healthy smoking
Two books examine timeworn tactics employed by the tobacco industry to attract new customers *By M. Parascandola*

411 Contextualizing avian origins and evolution
Recent genomic advances have transformed our understanding of birds *By A. Feduccia*

LETTERS

413 Accurately assessing children's asthma study
By J. C. Celedón

414 Misleading portrayal of children's asthma study
By J. M. Berg

414 Editor's note
By T. Appenzeller

RESEARCH

IN BRIEF

415 From *Science* and other journals

REVIEW

418 Nanophotonics
Multidimensional phase singularities in nanophotonics *J. Ni* et al.
REVIEW SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABJ0039

Making variant-resistant antibodies

419 Climate change
Socioeconomic impacts of marine heatwaves: Global issues and opportunities
K. E. Smith et al.
REVIEW SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABJ3593

RESEARCH ARTICLES

420 Coronavirus
Low-dose mRNA-1273 COVID-19 vaccine generates durable memory enhanced by cross-reactive T cells
J. Mateus et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABJ9853

Social neuroscience

421 Social agent identity cells in the prefrontal cortex of interacting groups of primates *R. Báez-Mendoza* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABB4149

422 Cortical representation of group social communication in bats *M. C. Rose* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABA9584
PERSPECTIVE p. 397

423 Coronavirus
A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa
E. Wilkinson et al.

432 Organic chemistry
General method for iron-catalyzed multicomponent radical cascades–cross-couplings *L. Liu* et al.
PERSPECTIVE p. 396

439 Gut inflammation
Identification of a choroid plexus vascular barrier closing during intestinal inflammation
S. Carloni et al.

REPORTS

448 Nanophotonics
Broadband electro-optic polarization conversion with atomically thin black phosphorus *S. Biswas* et al.

454 Water capture
Evolution of water structures in metal-organic frameworks for improved atmospheric water harvesting *N. Hanikel* et al.
PERSPECTIVE p. 402

459 Nanocatalysts
Sulfur-anchoring synthesis of platinum intermetallic nanoparticle catalysts for fuel cells *C.-L. Yang* et al.

465 Structural materials
Lightweight, strong, moldable wood via cell wall engineering as a sustainable structural material *S. Xiao* et al.
PERSPECTIVE p. 400; PODCAST

472 Coronavirus
Defining variant-resistant epitopes targeted by SARS-CoV-2 antibodies: A global consortium study *K. M. Hastie* et al.

478 Metallurgy
In situ design of advanced titanium alloy with concentration modulations by additive manufacturing *T. Zhang* et al.

483 Poaching impacts
Ivory poaching and the rapid evolution of tusklessness in African elephants
S. C. Campbell-Staton et al.
PERSPECTIVE p. 394

488 Microbiology
Rapid evolutionary turnover of mobile genetic elements drives bacterial resistance to phages *F. A. Hussain* et al.
PERSPECTIVE p. 399

DEPARTMENTS

377 Editorial
Preparing for "Disease X"
By Maria D. Van Kerkhove et al.

379 Editorial
Share vaccine know-how
By Priti Krishtel and Fatima Hassan

498 Working Life
When to say 'no'
By Mingde Zheng

ON THE COVER

Unlike polymers and metals, wood cannot be easily shaped. To close this processing gap, researchers have developed a method of cell wall engineering in which partial chemical delignification and a subsequent physical water-shock process make it possible to mold flat sheets of wood into versatile 3D structures (e.g., honeycombs) with substantially improved mechanical strength. See pages 400 and 465. *Photo: Jiaqi Dai*

Science Staff 376
Science Careers 494

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2021 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $2148; Foreign postage extra: Air assist delivery: $98. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling available from backissues.science.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

ILLUSTRATION: CHRISTINA CORBACI

Science AAAS

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

Executive Editor Monica M. Bradford

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

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

News Editor Tim Appenzeller

NEWS MANAGING EDITOR John Travis **INTERNATIONAL EDITOR** Martin Enserink **DEPUTY NEWS EDITORS** Elizabeth Culotta, Lila Guterman, David Grimm, Eric Hand (Europe), David Malakoff **SR. CORRESPONDENTS** Daniel Clery (UK), Jon Cohen, Jeffrey Mervis, Elizabeth Pennisi **ASSOCIATE EDITORS** Jeffrey Brainard, Kelly Servick, Catherine Matacic **NEWS REPORTERS** Adrian Cho, Jennifer Couzin-Frankel, Jocelyn Kaiser, Rodrigo Pérez Ortega (Mexico City), Robert F. Service, Erik Stokstad, Paul Voosen, Meredith Wadman **INTERNS** Menna Ibrahim, Sofia Moutinho, Anil Oza, Alex Viveros **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, Michael Price, Tania Rabesandratana (Barcelona), Joshua Sokol, Emily Underwood, Gretchen Vogel (Berlin), Lizzie Wade (Mexico City) **CAREERS** Rachel Bernstein (Editor), Katie Langin (Associate Editor) **COPY EDITORS** Julia Cole (Senior Copy Editor), Morgan Everett, Cyra Master (Copy Chief) **ADMINISTRATIVE SUPPORT** Meagan Weiland

Creative Director Beth Rakouskas

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

Chief Executive Officer and Executive Publisher Sudip Parikh

Publisher, Science Family of Journals Bill Moran

DIRECTOR, BUSINESS SYSTEMS AND FINANCIAL ANALYSIS Randy Yi **DIRECTOR, BUSINESS OPERATIONS & ANALYSIS** Eric Knott **DIRECTOR OF ANALYTICS** Enrique Gonzales **MANAGER, BUSINESS OPERATIONS** Jessica Tierney **SENIOR BUSINESS ANALYST** Cory Lipman **FINANCIAL ANALYST** Isacco Fusi **ADVERTISING SYSTEM ADMINISTRATOR** Tina Burks **DIGITAL/PRINT STRATEGY MANAGER** Jason Hillman **SENIOR MANAGER, PUBLISHING AND CONTENT SYSTEMS** Marcus Spiegler **ASSISTANT MANAGER DIGITAL/PRINT** Rebecca Doshi **SENIOR CONTENT AND PUBLISHING SYSTEMS SPECIALIST** Jacob Hedrick **SENIOR CONTENT SPECIALISTS** Steve Forrester, Antoinette Hodal, Lori Murphy **PRODUCTION SPECIALIST** Kristin Wowk **DIGITAL PRODUCTION MANAGER** Lisa Stanford **CONTENT SPECIALIST** Kimberley Oster **ADVERTISING PRODUCTION OPERATIONS MANAGER** Deborah Tompkins **DESIGNER, CUSTOM PUBLISHING** Jeremy Huntsinger **SR. TRAFFIC ASSOCIATE** Christine Hall **SPECIAL PROJECTS ASSOCIATE** Sarah Dhere

ASSOCIATE DIRECTOR, BUSINESS DEVELOPMENT Justin Sawyers **GLOBAL MARKETING MANAGER** Allison Pritchard **DIGITAL MARKETING MANAGER** Aimee Aponte **JOURNALS MARKETING MANAGER** Shawana Arnold **MARKETING ASSOCIATES** Ashley Hylton, Mike Romano, Tori Velasquez, Jenna Voris, Justin Wood **SENIOR DESIGNER** Kim Huynh

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

DIRECTOR, PRODUCT & PUBLISHING DEVELOPMENT Chris Reid **DIRECTOR, BUSINESS STRATEGY AND PORTFOLIO MANAGEMENT** Sarah Whalen **ASSOCIATE DIRECTOR, PRODUCT MANAGMENT** Kris Bishop **PRODUCT DEVELOPMENT MANAGER** Scott Chernoff **PUBLISHING TECHNOLOGY MANAGER** Michael Di Natale **SR. PRODUCT ASSOCIATE** Robert Koepke **PRODUCT ASSOCIATE** Anne Mason **SPJ ASSOCIATE MANAGER** Samantha Bruno Fuller **SPJ ASSOCIATE** Casey Buchta

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

DIRECTOR, GLOBAL SALES Tracy Holmes **US EAST COAST AND MID WEST SALES** Stephanie O'Connor **US MID WEST, MID ATLANTIC AND SOUTH EAST SALES** Chris Hoag **US WEST COAST SALES** Lynne Stickrod **ASSOCIATE DIRECTOR, ROW** Roger Goncalves **SALES REP, ROW** Sarah Lelarge **SALES ADMIN ASSISTANT, ROW** Victoria Glasbey **DIRECTOR OF GLOBAL COLLABORATION AND ACADEMIC PUBLISHING RELATIONS, ASIA** Xiaoying Chu **ASSOCIATE DIRECTOR, INTERNATIONAL COLLABORATION** Grace Yao **SALES MANAGER** Danny Zhao **MARKETING MANAGER** Kilo Lan **ASCA CORPORATION, JAPAN** Yoshimi Toda (Tokyo), Miyuki Tani (Osaka)

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

---

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

SCIENCE INTERNATIONAL
Clarendon House
Clarendon Road
Cambridge, CB2 8FH, UK

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

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

EDITORIAL
science_editors@aaas.org

NEWS
science_news@aaas.org

INFORMATION FOR AUTHORS
science.org/authors/
science-information-authors

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

MEDIA CONTACTS
scipak@aaas.org

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

INSTITUTIONAL SALES AND SITE LICENSES
science.org/librarian

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

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

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

MEMBERSHIP AND INDIVIDUAL SUBSCRIPTIONS
science.org/subscriptions

MEMBER BENEFITS
aaas.org/membership/benefits

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

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

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

Erin Adams, *U. of Chicago*
Takuzo Aida, *U. of Tokyo*
Leslie Aiello, *Wenner-Gren Fdn.*
Deji Akinwande, *UT Austin*
Judith Allen, *U. of Manchester*
Marcella Alsan, *Harvard U.*
Sebastian Amigorena, *Inst. Curie*
James Analytis, *UC Berkeley*
Trevor Archer, *NIEHS, NIH*
Paola Arlotta, *Harvard U.*
David Awschalom, *U. of Chicago*
Delia Baldassarri, *NYU*
Nenad Ban, *ETH Zürich*
Nandita Basu, *U. of Waterloo*
Franz Bauer, *Pontificia U. Católica de Chile*
Ray H. Baughman, *UT Dallas*
Carlo Beenakker, *Leiden U.*
Yasmine Belkaid, *NIAID, NIH*
Philip Benfey, *Duke U.*
Kiros T. Berhane, *Columbia U.*
Joseph J. Berry, *NREL*
Alessandra Biffi, *Harvard Med.*
Chris Bowler, *École Normale Supérieure*
Ian Boyd, *U. of St. Andrews*
Emily Brodsky, *UC Santa Cruz*
Ron Brookmeyer, *UCLA* **(S)**
Christian Büchel, *UKE Hamburg*
Dennis Burton, *Scripps Res.*
Carter Tribley Butts, *UC Irvine*
György Buzsáki, *NYU School of Med.*
Mariana Byndloss, *Vanderbilt U. Med. Ctr.*
Annmarie Carlton, *UC Irvine*
Simon Cauchemez, *Inst. Pasteur*
Ling-Ling Chen, *SIBCB, CAS*
M. Keith Chen, *UCLA*
Zhijian Chen, *UT Southwestern Med. Ctr.*
Ib Chorkendorff, *Denmark TU*
Amander Clark, *UCLA*
James J. Collins, *MIT*
Robert Cook-Deegan, *Arizona State U.*
Virginia Cornish *Columbia U.*
Carolyn Coyne, *Duke U.*
Roberta Croce, *VU Amsterdam*
Ismaila Dabo, *Penn State U.*
Jeff L. Dangl, *UNC*
Chiara Daraio, *Caltech*
Nicolas Dauphas, *U. of Chicago*
Christian Davenport, *U. of Michigan*
Frans de Waal, *Emory U.*
Claude Desplan, *NYU*
Sandra Díaz, *U. Nacional de Córdoba*
Ulrike Diebold, *TU Wien*
Stefanie Dimmeler, *Goethe-U. Frankfurt*
Hong Ding, *Inst. of Physics, CAS*
Dennis Discher, *UPenn*
Jennifer A. Doudna, *UC Berkeley*
Ruth Drdla-Schutting, *Med. U. Vienna*
Raissa M. D'Souza, *UC Davis*
Bruce Dunn, *UCLA*
William Dunphy, *Caltech*
Scott Edwards, *Harvard U.*
Todd Ehlers, *U. of Tübingen*
Andrea Encalada, *U. San Francisco de Quito*
Nader Engheta, *UPenn*
Karen Ersche, *U. of Cambridge*
Beate Escher, *UFZ & U. of Tübingen*
Barry Everitt, *U. of Cambridge*
Vanessa Ezenwa, *U. of Georgia*
Michael Feuer, *GWU*
Toren Finkel, *U. of Pitt. Med. Ctr.*
Gwenn Flowers, *Simon Fraser U.*
Peter Fratzl, *Max Planck Inst. Potsdam*
Elaine Fuchs, *Rockefeller U.*
Jay Gallagher, *U. of Wisconsin*
Daniel Geschwind, *UCLA*
Ramon Gonzalez, *U. of South Florida*
Sandra González-Bailón, *UPenn*
Nicolas Gruber, *ETH Zürich*
Hua Guo, *U. of New Mexico*
Taekjip Ha, *Johns Hopkins U.*
Sharon Hammes-Schiffer, *Yale U.*
Wolf-Dietrich Hardt, *ETH Zürich*
Louise Harra, *U. Coll. London*
Jian He, *Clemson U.*
Carl-Philipp Heisenberg, *IST Austria*
Ykä Helariutta, *U. of Cambridge*
Janet G. Hering, *Eawag*
Christoph Hess, *U. of Basel & U. of Cambridge*
Heather Hickman, *NIAID, NIH*
Hans Hilgenkamp, *U. of Twente*
Kai-Uwe Hinrichs, *U. of Bremen*
Deirdre Hollingsworth, *U. of Oxford*
Randall Hulet, *Rice U.*
Auke Ijspeert, *EPFL*
Akiko Iwasaki, *Yale U.*
Stephen Jackson, *USGS & U. of Arizona*
Erich Jarvis, *Rockefeller U.*
Peter Jonas, *IST Austria*
Matt Kaeberlein, *U. of Wash.*
William Kaelin Jr., *Dana-Farber Cancer Inst.*
Daniel Kammen, *UC Berkeley*
Kisuk Kang, *Seoul Nat. U.*
Sabine Kastner, *Princeton U.*
V. Narry Kim, *Seoul Nat. U.*
Robert Kingston, *Harvard Med.*
Nancy Knowlton, *Smithsonian Institution*
Etienne Koechlin, *École Normale Supérieure*
Alex L. Kolodkin, *Johns Hopkins U.*
Julija Krupic, *U. of Cambridge*
Paul Kubes, *U. of Calgary*
Gabriel Lander, *Scripps Res.* **(S)**
Mitchell A. Lazar, *UPenn*
Wendell Lim, *UCSF*
Luis Liz-Marzán, *CIC biomaGUNE*
Omar Lizardo, *UCLA*
Jonathan Losos, *Wash. U. in St. Louis*
Ke Lu, *Inst. of Metal Res., CAS*
Christian Lüscher, *U. of Geneva*
Jean Lynch-Stieglitz, *Georgia Inst. of Tech.*
David Lyons, *U. of Edinburgh*
Fabienne Mackay, *QIMR Berghofer*
Anne Magurran, *U. of St. Andrews*
Asifa Majid, *U. of York*
Oscar Marín, *King's Coll. London*
Charles Marshall, *UC Berkeley*
Christopher Marx, *U. of Idaho*
David Masopust, *U. of Minnesota*
Geraldine Masson, *CNRS*
Jason Matheny, *Georgetown U.*
Heidi McBride, *McGill U.*
C. Robertson McClung, *Dartmouth*
Rodrigo Medellín, *U. Nacional Autónoma de México*
C. Jessica Metcalf, *Princeton U.*
Baoxia Mi, *UC Berkeley*
Tom Misteli, *NCI, NIH*
Alison Motsinger-Reif, *NIEHS, NIH* **(S)**
Suresh Naidu, *Columbia U.*
Danielle Navarro, *U. of New South Wales*
Daniel Nettle, *Newcastle U.*
Daniel Neumark, *UC Berkeley*
Beatriz Noheda, *U. of Groningen*
Helga Nowotny, *Vienna Sci. & Tech. Fund*
Rachel O'Reilly, *U. of Birmingham*
Pilar Ossorio, *U. of Wisconsin*
Andrew Oswald, *U. of Warwick*
Isabella Pagano, *Istituto Nazionale di Astrofisica*
Elizabeth Levy Paluck, *Princeton U.*
Jane Parker, *Max Planck Inst. Cologne*
Giovanni Parmigiani, *Dana-Farber Cancer Inst.* **(S)**
Daniel Pauly, *U. of British Columbia*
Ana Pêgo, *U. do Porto*
Samuel Pfaff, *Salk Inst.*
Julie Pfeiffer, *UT Southwestern Med. Ctr.*
Philip Phillips, *UIUC*
Matthieu Piel, *Inst. Curie*
Kathrin Plath, *UCLA*
Martin Plenio, *Ulm U.*
Katherine Pollard, *UCSF*
Elvira Poloczanska, *Alfred-Wegener-Inst.*
Julia Pongratz, *Ludwig Maximilians U.*
Philippe Poulin, *CNRS*
Jonathan Pritchard, *Stanford U.*
Lei Stanley Qi, *Stanford U.*
Trevor Robbins, *U. of Cambridge*
Joeri Rogelj, *Imperial Coll. London*
Amy Rosenzweig, *Northwestern U.*
Mike Ryan, *UT Austin*
Miquel Salmeron, *Lawrence Berkeley Nat. Lab*
Nitin Samarth, *Penn State U.*
Erica Ollmann Saphire, *La Jolla Inst.*
Joachim Saur, *U. zu Köln*
Alexander Schier, *Harvard U.*
Wolfram Schlenker, *Columbia U.*
Susannah Scott, *UC Santa Barbara*
Anuj Shah, *U. of Chicago*
Vladimir Shalaev, *Purdue U.*
Jie Shan, *Cornell U.*
Beth Shapiro, *UC Santa Cruz*
Jay Shendure, *U. of Wash.*
Steve Sherwood, *U. of New South Wales*
Brian Shoichet, *UCSF*
Robert Siliciano, *JHU School of Med.*
Lucia Sivilotti, *U. Coll. London*
Alison Smith, *John Innes Ctr.*
Richard Smith, *UNC* **(S)**
Mark Smyth, *QIMR Berghofer*
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*
Wim van der Putten, *Netherlands Inst. of Ecology*
Ivo Vankelecom, *KU Leuven*
Henrique Veiga-Fernandes, *Champalimaud Fdn.*
Reinhilde Veugelers, *KU Leuven*
Bert Vogelstein, *Johns Hopkins U.*
Julia Von Blume, *Yale School of Med.*
David Wallach, *Weizmann Inst.*
Jane-Ling Wang, *UC Davis* **(S)**
Jessica Ware, *Amer. Mus. of Natural Hist.*
David Waxman, *Fudan U.*
Chris Wikle, *U. of Missouri* **(S)**
Terrie Williams, *UC Santa Cruz*
Ian A. Wilson, *Scripps Res.* **(S)**
Hao Wu, *Harvard U.*
Wei Xie, *Tsinghua U.*
Yu Xie, *Princeton U.*
Jan Zaanen, *Leiden U.*
Kenneth Zaret, *UPenn School of Med.*
Bing Zhu, *Inst. of Biophysics, CAS*
Xiaowei Zhuang, *Harvard U.*
Maria Zuber, *MIT*

EDITORIAL

Preparing for "Disease X"

The past 30 years have exposed the global public health and economic threats posed by the emergence of infectious pathogens with epidemic and pandemic potential. Severe acute respiratory syndrome (SARS), middle east respiratory syndrome (MERS), influenza, Ebola, Marburg, Lassa, Nipah, Zika, and now SARS coronavirus 2 (SARS-CoV-2) each have been the "Disease X" of their time. The risk of future emergence is driven by multiple forces, including climate change, ecosystem changes, and increasing urbanization. The next Disease X could appear at any time, and the world needs to be better prepared.

The newly established World Health Organization (WHO) Scientific Advisory Group on the Origins of Novel Pathogens (SAGO) presents an unprecedented opportunity to better guide studies that specifically investigate high-threat pathogens. Its mandate is to advise the WHO on developing a framework to define comprehensive studies on the origins of such pathogens, including SARS-CoV-2—information that is essential for developing policies and enhancing preparedness to reduce the possibility of future zoonotic spillover events (transmission of a pathogen from animals to humans) and the chance that those events become major outbreaks.

"...SAGO will quickly assess the status of SARS-CoV-2 origin studies..."

This is not the first time that international studies on the origin of a new virus have been conducted. Yet each time, scientists at the WHO and elsewhere faced challenges—not only scientific, but also logistical and political. These hurdles have also hampered efforts to understand the origins of COVID-19.

Since the beginning of this pandemic, scientists from around the world have worked together to understand the events that led to the first human infections. In May 2020, WHO's member states passed a unanimous resolution giving WHO a mandate to bring together international experts for scientific and collaborative studies on the virus origins. But it's clear that the scientific processes have been hurt by politicization, which is why the global scientific community must redouble efforts to drive the scientific process forward. In forming SAGO, experts were selected (from an open call for applicants) with diverse technical expertise from countries in all six WHO regions.

Building on the March 2021 findings from the Joint WHO-China study, as well as other findings published since then, SAGO will quickly assess the status of SARS-CoV-2 origin studies and advise the WHO on what is known, the outstanding gaps, and next steps. All hypotheses must continue to be examined and, as WHO has said from the outset, a fully open and transparent scientific process is essential. Recent findings on the potential for zoonotic spillover of SARS-CoV-2 to humans, either directly from bats or through other animals, include, but are not limited to, studies of wildlife sold in markets in and around Wuhan, China (where cases of COVID-19 were first reported in December 2019); studies of SARS-like coronaviruses circulating in bats in China and Southeast Asia; studies on prepandemic biological sampling around the world; and other animal susceptibility studies. However, detailed investigations of the earliest known and suspected cases in China prior to December 2019 are still urgently needed, including analyses of stored blood samples from 2019 in Wuhan and surrounding areas and retrospective searches of hospital and mortality data for earlier cases.

As well, laboratory hypotheses must be examined carefully, with a focus on labs in the location where the first reports of human infections emerged in Wuhan. A lab accident cannot be ruled out until there is sufficient evidence to do so and those results are openly shared.

COVID-19 will not be the last Disease X. We need scientific collaboration, data sharing, and implementation of a robust "one health" approach that brings together the human, animal, and environmental spheres to boost risk identification, reduction, and surveillance in animals and at the human-animal-environment interface. This must be linked to early action to investigate, characterize, and contain threats. In parallel, the world needs systematic processes to study the emergence of these pathogens and their routes of transmission from natural reservoirs to humans. Laboratory protocols around the world must be monitored and strengthened.

Globally, at least 4.8 million people have died from COVID-19. They and their families are owed answers as to where and how the virus originated. It's in everyone's interest to better prepare for the next Disease X.

–Maria D. Van Kerkhove, Michael J. Ryan, Tedros Adhanom Ghebreyesus

Maria D. Van Kerkhove
is the COVID-19 technical lead and head of Emerging Diseases and Zoonoses in the Health Emergencies Programme at the World Health Organization, Geneva, Switzerland.

Michael J. Ryan
is the executive director of the Health Emergencies Programme of the World Health Organization, Geneva, Switzerland.

Tedros Adhanom Ghebreyesus
is the director general of the World Health Organization, Geneva, Switzerland.
sago@who.int

Published online 13 October 2021; 10.1126/science.abm7796

BII | Science | Prize for Innovation

Apply for the new BII & Science Prize for Innovation today

Behind every life-changing solution is an entrepreneurial scientist–a creative mind who proved an idea in the lab and dared to carry it out in the world.

To encourage more scientists to translate their research, BioInnovation Institute (BII) & Science present a new annual award.

Our three winners will have their essays published in Science magazine and will be invited into BII's entrepreneurial ecosystem. In addition, the Grand Prize winner will receive a prize of USD 25,000 and each runner-up will receive USD 10,000 at a grand award show celebration in Copenhagen, Denmark.

The call for applications has just opened.
Apply before November 1, 2021.
www.bii.dk/scienceprize

Presented by BII & Science

BII BioInnovation Institute Science

Apply before November 1, 2021
www.bii.dk/scienceprize

EDITORIAL

Share vaccine know-how

Leaders of the world's major economies, representing 80% of the world's gross domestic product, will meet at the end of this month at the G20 meeting in Rome to address issues of global importance, including how to increase the worldwide supply of COVID-19 vaccines. How the world addresses the current state of vaccine inequity will affect how we meet future global disease challenges.

It has been 10 months since a vaccine became available, but only 27% percent of people in low- and lower–middle-income countries—home to the majority of the world's population—have received a first dose, compared with over 70% in many high- and upper–middle-income countries. The push for booster shots in wealthy countries will further stratify vaccination access. It will take almost 1 billion doses annually to offer boosters to people age 50 and over in wealthier countries, which is enough to vaccinate nearly 40% of the entire continent of Africa.

To achieve equity, intellectual property (IP) relating to COVID-19 vaccines should be temporarily waived to bolster their production. IP includes patents, trademark, copyright, and any trade secrets that are not in a patent. A waiver will ensure that drugmakers will not enforce their IP against other manufacturers for a time-limited period. Pharmaceutical companies will also need to share knowledge, including the "secret sauce," to accelerate vaccine production by other manufacturers.

> "...companies will also need to share...the 'secret sauce,' to accelerate vaccine production by other manufacturers."

Pharmaceutical companies have been prioritizing wealthy countries. Moderna and Pfizer have supplied the vast majority of vaccine to high-income countries. This imbalance has been attributed primarily to preorders and stockpiling of existing supply by wealthy nations, who are likely to have 1.2 billion extra doses available that they do not need by the end of this year.

Reliance on charity is not working as envisioned. The main donation mechanism, COVID-19 Vaccines Global Access (COVAX), purchases vaccines for distribution to 92 countries using money donated by governments and charitable foundations. To date, only 371.1 million doses have been shipped, falling far short of a targeted 2 billion doses by the end of 2021. It remains to be seen whether this goal will be reached in 2022.

Waiving IP rights and sharing knowledge will allow middle- and low-income countries to produce their own vaccines rather than rely on donations and COVAX. A time-limited IP waiver would provide legal certainty so that prospective manufacturers across the world can initiate plans to produce. Last year, Moderna said it would not enforce its vaccine patents during the pandemic. However, without knowledge transfer, including trade secrets, other manufacturers cannot produce this vaccine on an accelerated timeline. Moderna's refusal to license technology to other manufacturers has been criticized because its vaccine was developed primarily with public funds. According to the former director of chemistry at Moderna, if knowledge were shared, supply could increase in as little as 3 months.

An incorrect claim often made is that the capability to produce the vaccines does not exist outside of the United States and Europe. Manufacturers such as Teva, Incepta, Biolyse, and several Korean companies have the capacity and facilities to commence production but have reported being turned down by vaccine makers. Over 25 manufacturers from low- and middle-income countries have also indicated their ability and willingness to produce vaccines through the World Health Organization's COVID-19 mRNA vaccine technology transfer hub. Waiving IP and sharing knowledge with these manufacturers could accelerate production and save lives. If Johnson & Johnson (J&J) licensed its IP and shared knowledge, multiple manufacturers currently making the Sputnik V vaccine could potentially also make the J&J vaccine within 6 months because they rely on similar technologies. Instead, J&J contracted a South African company to bottle and package its vaccine for shipment to Europe but did not license the necessary technology for the company itself to produce the vaccine in South Africa—despite the urgent need in Africa, where 95% of people are yet to be fully vaccinated.

The shared goal should be to save as many lives as possible, which means swiftly addressing obstacles that have been revealed during this pandemic. Going forward, our collective survival will depend on both sharing knowledge and ensuring equity in accessing medicine.

–Priti Krishtel and Fatima Hassan

Priti Krishtel is a founder and an executive director of the Initiative for Medicines, Access and Knowledge, Oakland, CA, USA. priti@i-mak.org; @pritikrishtel and @IMAKGlobal

Fatima Hassan is the founder and director of the Health Justice Initiative, Cape Town, South Africa. fatima@healthjusticeinitiative.org.za; @_HassanF and @HealthJusticeIn

10.1126/science.abm8724

NEWS

> “Many thousands of deaths ... could have been avoided.”

A U.K. parliamentary report that faulted the government for being slow to start COVID-19 lockdowns and testing. The result, it said, was one of the United Kingdom’s “most important public health failures.”

IN BRIEF

Edited by **Jeffrey Brainard**

COVID-19

Roster for origin panel 2.0 unveiled

The World Health Organization (WHO) last week proposed the 26-member lineup of its new, permanent Scientific Advisory Group on the Origins of Novel Pathogens (SAGO), one of several groups that will continue to explore what caused the COVID-19 pandemic. The panel retains six of the 10 members chosen by WHO for an earlier, ad hoc team that sought the source of SARS-CoV-2; that group drew criticism after it concluded, following a fact-finding trip to China, that a laboratory origin was “highly unlikely.” The SAGO roster, chosen from more than 700 applications, includes representatives from 26 countries. One member is from China, which has challenged WHO’s call to more fully probe the lab-leak hypothesis. SAGO will also recommend and coordinate efforts to track down other novel pathogens and prevent major outbreaks. Supporters hope constituting it as a standing panel focused on science will standardize and depoliticize its work. WHO plans to finalize the roster later this month after a comment period.

U.S. academy boots archaeologist

#METOO | The U.S. National Academy of Sciences (NAS) this month expelled prominent archaeologist Luis Jaime Castillo Butters of Peru, a former minister of culture, for sexual harassment. The ouster is the academy’s third for sexual harassment since May and the first of an international member. It followed a 2020 investigation of Castillo Butters, an expert on the Moche culture, by his university, the Pontifical Catholic University of Peru (PUCP). That probe was triggered by a blog post that year by independent journalist Michael Balter containing allegations from several unnamed women. Unnamed professors corroborated two of the accounts. Castillo Butters told *Science*, “I am completely and absolutely innocent. ... The NAS process is not fair.” The university’s report found evidence of harassment but stated PUCP could not initiate disciplinary proceedings because the alleged offenses occurred prior to 2016, when the university adopted sexual harassment regulations. In 2019, NAS adopted bylaw changes that for the first time allowed it to expel members for misconduct.

Viking presence is 1000 years old

ARCHAEOLOGY | In the 1960s, scientists discovered what is still the only confirmed Viking settlement in North America, at L’Anse aux Meadows in Newfoundland. But researchers couldn’t determine exactly when the Norse had lived there: Radiocarbon dating of artifacts suggested it was within a 275-year window starting in the 790s. Now, using chemical clues left by ancient solar storms, scientists report in this week’s issue of *Nature* that they’ve determined the Vikings were present in 1021. When big solar storms hit Earth, the particles cause a spike in the creation of carbon-14 atoms, which are incorporated into growing trees. One such storm hit in 993, previous research found, and researchers used tree rings containing the telltale carbon spike to calculate that three pieces of cut wood found at L’Anse aux Meadows had all come from trees felled 28 years later.

Vikings built structures at L’Anse aux Meadows in Canada like this reconstructed example.

PHOTO: GLENN NAGEL PHOTOGRAPHY

Large charity bets on diversity

WORKFORCE | The Howard Hughes Medical Institute (HHMI) last week said it will spend $2.2 billion on a slew of programs aimed at eliminating racial disparities among biomedical faculty at U.S. universities and medical centers. The philanthropy is banking on continued robust growth in its $22 billion endowment and a significant contraction in recent years of its flagship investigators program to finance the initiative, which stretches from the undergraduate years to the ranks of senior faculty members. More than half the money will be funneled to 125 early-career scientists, to be chosen for their scientific excellence and commitment to diversity. (Federal law prohibits HHMI from limiting eligibility for the awards to those from groups underrepresented in science.) Other efforts include mentorship training for advisers to HHMI-funded graduate students and postdocs. HHMI officials hope the spending, budgeted through 2040, will trigger additional investments by other private entities and government agencies to address the persistent underrepresentation of minorities in the biomedical workforce.

Biden targets nonstick chemicals

ENVIRONMENT | The U.S. Environmental Protection Agency (EPA) this week said it will propose enforceable standards

IN FOCUS A striped fishing spider spinning an egg sac was among the winners in the Wildlife Photographer of the Year photo competition sponsored by the Natural History Museum in London. Members of this species (*Dolomedes scriptus*), native to eastern North America, carry the sacs, which can contain more than 750 eggs, until the eggs hatch. This spider's leg span is 6 centimeters.

PHOTO: GIL WIZEN/WILDLIFE PHOTOGRAPHER OF THE YEAR

for levels in drinking water of a class of long-lasting, toxic chemicals known as per- and polyfluoroalkyl substances (PFAS). These compounds, which feature nonstick, stain-repellent properties, break down extremely slowly, have been found in many drinking water supplies, and are linked to health effects such as cancer and low birth weight. Among other new measures, EPA will require companies to provide it with toxicity data on many types of PFAS, a step that the chemical industry has long resisted. Environmental advocates have criticized previous commitments by the administrations of former Presidents Barack Obama and Donald Trump to control PFAS as coming up short.

COVID-19 reverses TB progress

PUBLIC HEALTH | The coronavirus pandemic has greatly hurt the worldwide fight against tuberculosis (TB), according to the World Health Organization's *Global Tuberculosis Report,* published last week. As many countries shifted resources to combating COVID-19 and overwhelmed health care systems treated fewer patients with other conditions, an estimated 1.5 million people died from TB's bacterial infection, up from 1.4 million in 2019. This marks the first time TB deaths globally have risen since 2005. At the same time, TB diagnoses dropped from 7.1 million in 2019 to 5.8 million last year, suggesting a growing number of people sick with TB remain untreated. The number of people who received preventive medication because they were exposed to TB or are HIV positive dropped by one-fifth in 2020. The impact may not show up as increased TB case numbers and deaths until next year, the report's authors warn.

Senate bills would boost research

FUNDING | Health, climate, and defense research would get major funding increases under bills released this week by Democrats in the U.S. Senate. But Republicans are demanding major changes, including equal boosts for civilian and military spending, rather than the proposed ones of 13% and 5%, respectively. The nine-bill package would give 12% more to the National Institutes of Health (NIH) and the National Science Foundation (NSF), bringing their appropriations to $48 billion and $9.5 billion. NIH would get a new Advanced Research Projects Agency-Health while NSF would gain a technology directorate. Legislators also proposed an 8% boost for NASA's science programs, a 16% hike for in-house labs at the National Institute of Standards and Technology, and 10% more for the Defense Advanced Research Projects Agency. For the first time in a decade, the spending bills also included earmarks, projects sought by individual lawmakers to benefit their state.

A health worker gives polio vaccine to a child in Kabul, Afghanistan, in 2018.

PHOTO: AP PHOTO/RAHMAT GUL

GLOBAL HEALTH

Polio vaccination to resume across Afghanistan

Success depends on shoring up the country's health system

By **Leslie Roberts**

The Taliban government of Afghanistan has OK'd house-to-house polio vaccination across the country starting on 8 November. The 18 October announcement is "huge," says Hamid Jafari, who directs operations in the region for the Global Polio Eradication Initiative (GPEI). It will give the program access to 3.3 million Afghan children who have been out of reach for 3.5 years—and might bring the world closer to the ultimate goal of polio eradication.

A decadeslong campaign against the disease has left Afghanistan as one of just two countries, along with neighboring Pakistan, to still harbor the wild polio virus. But the fight has been hamstrung there since 2018, when the Taliban, suspicious that vaccinators were helping the U.S. government target drone strikes, banned house-to-house vaccination in its stronghold in southern Afghanistan. Since then, roughly 85% of the country's polio cases have occurred in areas that are off limits to vaccinators.

The new government's decision comes at an opportune time, polio officials say. After years of struggle and fluctuating case counts, only one case of wild polio has been detected this year in Afghanistan, and one in Pakistan. (Despite the Taliban's ban, polio surveillance across both countries has remained strong—as has vaccination outside of southern Afghanistan.) Cases resulting from vaccine-derived polio viruses are down, too. "It's a unique opportunity," says Aidan O'Leary, who directs GPEI at the World Health Organization (WHO) in Geneva. "We have to be sure we grab it."

GPEI has quietly negotiated with the Taliban since 2018, and those talks intensified when the group began its dizzying blitz this summer and assumed power on 15 August.

Key discussions occurred in September, when WHO Director-General Tedros Adhanom Ghebreyesus and Ahmed Al-Mandhari, WHO's regional director for the Eastern Mediterranean, toured the country. In those talks, says Jafari, who was part of WHO's delegation, Taliban officials made clear that their first priority was to shore up the Sehatmandi Project, the backbone of Afghanistan's health system. (*Sehatmandi* means "health" in Urdu.) Its major donors, including the World Bank, the European Union, and the U.S. Agency for International Development, froze funds when the Taliban took over. By 1 month ago, just 17% of the system's more than 2300 health facilities were fully functional, and many of its 20,000 health workers had not been paid. The system was on "the brink of collapse," Tedros and Al-Mandhari said in a 22 September statement, warning of "an imminent humanitarian catastrophe."

Routine childhood immunization has fallen off, says Godwin Mindra, a senior immunization specialist and deputy polio team lead for UNICEF in Kabul, and measles outbreaks are raging. COVID-19 vaccine is sitting unused and near expiration. With local facilities closed, people are crowding into regional hospitals, Mindra says. "I went to one regional hospital in Kandahar, and the patient load had increased three to four times. In the children's ward, there were three to four children in one bed."

On 20 September, the Global Fund threw Afghanistan a lifeline: $15 million to support Sehatmandi in October. Two days later, the United Nations kicked in $45 million to carry Sehatmandi through January 2022. Then on 24 September, the U.S. Department of the Treasury issued licenses that will allow organizations to provide humanitarian assistance to Afghanistan without violating U.S. sanctions. However, no new money has yet come through.

The November polio campaign will target 10 million children. As usual, the vaccinators will be women—under Afghan custom, only women can enter a home. This time they must wear hijabs, but most do anyway, Jafari says. He adds the program intends to help address other health needs, too. Polio vaccinators will also administer vitamin A, and the program hopes to help deliver measles and COVID-19 vaccines when feasible.

GPEI will conduct a second campaign in December, synchronized with one in Pakistan. Polio has to be tackled in both countries at once, Jafari says. "If not, one will surely reinfect the other country."

Six more polio vaccination campaigns are planned for Afghanistan in 2022. But with support for Sehatmandi assured only through January, their prospects are uncertain. "The global community needs to find a sustainable solution or Afghanistan will go into the next level of crisis," Jafari says. And as for polio, "The risk of resurgence remains very high." ■

Leslie Roberts is a science journalist in Washington, D.C.

BIOLOGY

Red blood cells may be immune sentinels

Oxygen-carrying cells also capture DNA from pathogens and damaged cells

By **Mitch Leslie**

As the roughly 30 trillion human red blood cells cruise the bloodstream and supply tissues with oxygen, they may also be doing something biologists did not expect: scanning for signs of infection and injury. According to new findings that would dramatically expand their roles, red blood cells snare suspect DNA from microbial invaders or damaged tissue and warn the immune system of danger.

"This is a new immune function for red blood cells that is quite exciting," says pathologist Steven Spitalnik of Columbia University. "I'm thinking of 17 different applications." The findings suggest, for instance, the immune role is linked to the anemia that often afflicts people with sepsis, COVID-19, and other conditions—and point to ways to combat it.

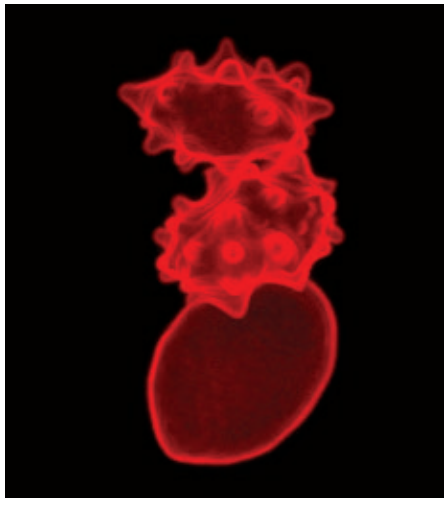
As human red blood cells mature, they lose all of their DNA and organelles. The cells' stripped-down structure inspired the traditional view that "they are blah, inert bags of hemoglobin," able to do little beyond ferry oxygen, says pulmonary and critical care physician Nilam Mangalmurti of the University of Pennsylvania Perelman School of Medicine, whose team made the DNA-capturing discovery. Over time, however, scientists have uncovered other functions for the cells, such as managing blood levels of nitric oxide, the molecule that spurs blood vessels to dilate.

In principle, red blood cells would make good protectors of the human body—not only because of their numbers, but also because they penetrate the nooks and crannies of organs and tissues. Researchers have found that some vertebrates such as fishes and birds enlist red blood cells for defense—and even sic them on pathogens. But in humans and other mammals, evidence for an immune role remained inconclusive.

In a 2018 study, Mangalmurti's team found a clue. The researchers determined that red blood cells harbor a type of molecular sensor, known as toll-like receptor 9 (TLR9), that recognizes and sticks to DNA molecules containing pairs of the nucleotide bases cytosine and guanine. Damaged human cells release such DNA, and the DNA of bacteria and other pathogens is also rich with these cytosine-guanine duos, or CpG motifs. They incite a strong immune system reaction that researchers ascribed primarily to white blood cells, which also sport TLR9.

But Mangalmurti and colleagues' latest research, published this week in *Science Translational Medicine*, suggests red cells take part as well. In the test tube, the researchers traced what happens when human red cells latch onto CpG-containing DNA. Small amounts of the DNA didn't appear to affect the cells, but larger amounts caused them to scrunch up, suggesting they were responding to the stimulus.

That red blood cell response could alert the immune system to foreign DNA or tissue damage, the researchers propose. It could help drive a less beneficial immune response as well. Mangalmurti and colleagues found that giving mice red blood cells that had captured CpG-containing DNA triggered bodywide inflammation. In humans, such inflammation is a hallmark of sepsis and can erupt because of injuries or other illnesses, including COVID-19.



When red blood cells latch onto DNA drifting in the circulatory system, they pucker.

IMAGE: LAM ET AL./SCIENCE TRANSLATIONAL MEDICINE

Red cells' sentinel role could also help explain a common complication for hospitalized patients. "Almost all critical care patients are anemic by their third day in the ICU [intensive care unit]," Mangalmurti says. As blood passes through the spleen, immune cells called macrophages ordinarily devour old and damaged red blood cells. Healthy red blood cells can avoid being culled if they display a particular "don't eat me" protein on their surface. But in the test tube, the researchers found, an encounter with CpG-carrying DNA causes the cells to conceal the portion of this protein that rebuffs macrophages.

The team observed the consequences of this change in mice. When the researchers transfused red blood cells into the animals, macrophages in their spleens gobbled cells that had been exposed to CpG-laden DNA, but ate fewer red blood cells that hadn't encountered the DNA. The researchers also found red blood cells from patients with sepsis and anemia sported more of the DNA than did cells from patients with only sepsis. Patients with COVID-19 and anemia showed the same disparity compared to patients who only had COVID-19, suggesting high levels of the DNA provoke destruction of the cells.

"We can now say that red blood cells have an immune function" in humans and other mammals, Mangalmurti says. Most of the time, she suggests, the cells serve as janitors for normal cleanup duties, sweeping up potentially harmful DNA that leaks into the circulation from the many body cells that die every day. But during an infection or after an injury, this DNA may flood the bloodstream. The red blood cells then sacrifice themselves by encouraging macrophages to eat them, alert the immune system, and trigger inflammation. If macrophages consume too many red blood cells, however, anemia may result.

Drugs that prevent the wayward DNA from adhering to TLR9 on red blood cells could treat the anemia, Mangalmurti proposes. She and her colleagues are also investigating whether DNA stuck to red blood cells could help diagnose infections.

The researchers "should be commended for the rigor of their experimental approach," says clinical pathologist Robertson Davenport of the University of Michigan, Ann Arbor. But he adds that further data is necessary to confirm that red blood cells are essential for mammals' immune defense.

Davenport adds that anemia in patients with sepsis can result from several causes and there's little evidence that excessive macrophage consumption of red blood cells is responsible. Then again, he notes, "We haven't looked very carefully."

In time, Mangalmurti is convinced, biologists will look at these oxygen carriers in a whole new light. ■

Mongolian horses like these, as well as almost all other living horses, descend from steeds domesticated north of the Caucasus Mountains, according to a study of ancient genomes.

DOMESTICATION

Ancient DNA reveals long-sought homeland of modern horses

Bronze Age herders began riding the ancestor of modern horses in the grassy western Eurasian steppe

By **Ann Gibbons**

About 4200 years ago, a few herders on the western Eurasian steppe got a brand-new mount. They were experienced at herding wild horses for food, but their new steeds had a calmer disposition and a stronger back, making the horses easier to train and ride, perhaps for the first time. The new model galloped across Eurasia within a few centuries, triggering major shifts in Bronze Age human cultures. "Once we domesticated this new kind of horse, suddenly they were everywhere," says molecular archaeologist Ludovic Orlando of the French national research agency CNRS and Paul Sabatier University.

Orlando is lead author of a landmark study that pinpoints where the new horse first appeared and how it replaced earlier equids to become the ancestor of modern horses, from Shetland ponies to massive Clydesdales and sleek thoroughbreds. He and a giant interdisciplinary team analyzed 300 ancient horse genomes from 121 archaeological sites in Eurasia. In a paper in this week's issue of *Nature*, they conclude that the ancestor of all modern horses made its first appearance by 4200 years ago in the western Eurasian steppe.

"They have found the original homeland of modern horses," says molecular archaeologist Christina Warinner of Harvard University and the Max Planck Institute for the Science of Human History, who was not part of the study.

The first evidence of horse domestication comes earlier, from Kazakhstan, where herders of the Botai culture corralled mares for meat and perhaps milk about 5500 years ago. Researchers haven't proved the Botai horses, whose teeth show wear likely from bits, were actually ridden, but archaeologists assumed for years that they were ancestral to modern horses.

Then in 2018 Orlando and colleagues tested ancient DNA from the Botai horses and got a surprise: The horses were not the forerunners of modern horses. Instead, they were the ancestors of today's Przewalski's horses, endangered "wild" horses found only in Mongolia that escaped domestication long ago (*Science*, 6 April 2018, p. 111).

This put geneticists back at the starting gate in their quest to find the ancestor of modern horses, which they call DOM2 for "second domestication" horse, although they can't be sure whether the genetic change happened in wild or previously domesticated horses. "We asked, where were they? The where was huge, because it was all of Eurasia," Orlando says.

He traveled from Kazakhstan to Siberia and Mongolia to gather every sample he could get of bone from horses that lived 10,000 to 2000 years ago. After the genomes were sequenced, he and postdoc Pablo Librado sorted them into a family tree, to see how they were related to each other and to modern horses. The equids that lived more than 4200 years ago showed a rich genetic diversity. "The horses living in Anatolia, Europe, Central Asia, and Siberia used to be genetically quite distinct," Orlando says.

But starting by about 4200 years ago, that diversity began to disappear. The genetic profile of one type of horse, closely related to modern horses, began to spread rapidly across Eurasia, replacing the others. By 4000 years ago, the new horse dominated a region from central Anatolia to central Russia, where people of the Sintashta culture buried horses with the earliest spoked wheels and chariots in mounds called kurgans. By 3000 years ago, the replacement was complete. The new horse became an icon of late Bronze Age art, carved in the bone hilts of daggers and cast into bronze figurines, says archaeologist Pavel Kuznetsov of the Russian Academy of Sciences (RAS).

To learn what was so special about this new horse, researchers sought gene variants present in all DOM2 horses but not in earlier equids. They found one in the *ZFPM1* gene that helps control anxiety and aggression, and one in the *GSDMC* gene that's associated with chronic back pain in humans. The researchers suggest the new horse was calmer and less likely to have back problems, traits that would have made it easier to train and ride into battle.

Riding the steppe

Close cousins of the ancestor of all of today's horses—the early DOM2 horses—first appeared in the western Eurasian steppe. Once the DOM2 horse itself arose, it spread rapidly throughout Eurasia.

CREDITS: (PHOTO) LUDOVIC ORLANDO; (MAP) N. DESAI/*SCIENCE*; (DATA) SABINE REINHOLD; P. LIBRADO *ET AL.*, *NATURE*, 20 OCTOBER, "THE ORIGINS AND SPREAD OF DOMESTIC HORSES FROM THE WESTERN EURASIAN STEPPES"

To find out where in the vast Eurasian steppe herders first began to favor this valuable new horse, Orlando put out a call for more horse bones from archaeological sites. The DNA samples he got back solved the mystery. Although the team didn't find the very first DOM2 horse, they identified several closely related ancestors that lived from 5400 to 4600 years ago.

Those early cousins came from three sites in a center of horse domestication near the lower Volga and Don rivers in the northern Caucasus, which is now southern Russia (see map, p. 384). Here, pastoralist cultures of the time herded goats, sheep, cattle, and perhaps horses in the dry, open steppe, says archaeologist Sabine Reinhold of the German Archaeological Institute (DAI), a co-author. The horses of the region didn't yet have the two winning genetic adaptations, but they were part of the wild or managed herds that gave rise to that coveted horse, says DAI archaeologist Svend Hansen, another co-author.

One of the oldest dates, about 5400 years ago, comes from a horse's leg bone buried with a small child in a kurgan built by the Maykop culture at Aygurskiy, in southern Russia. Another early relative of the DOM2 horse was found at Repin Khutor, a site on the Don River that was home to the famed Yamnaya herders who make up a significant part of Europeans' genetic ancestry. Some researchers had thought the Yamnaya rode into Europe 5000 years ago on the ancestors of modern horses. But samples from the new study show that once they got to Europe, the Yamnaya used earlier European horses rather than the DOM2 variety.

The new horse's spread and takeover between 4200 and 3000 years ago coincide with a "cultural genesis" during the late Bronze Age in the Volga-Don region, Kuznetsov says. First came the DOM2 horse, followed by the spoked wheel and chariots. Together, those innovations gave people the ability to travel off well-trodden paths, allowing them to find new sources of tin, gold, and other metals; set up long-distance trade networks; and herd other livestock farther to new pastures, says Natalia Roslyakova of RAS.

Climbing on top of the DOM2 horse was a "complete game changer," says University of Oxford archaeologist Greger Larson, who wasn't part of this study. "You now have an advantage over anyone else who wants to get around quickly." This paper also shows, he says, that data from animals can reveal major cultural changes in the past. "I love the part that it's animals that can be used to test what happens to people … to understand their culture." ■

PRIMATOLOGY

Misuse of funds leads to shutdown of top primate center

Kyoto University's move follows dismissal of director

By **Dennis Normile**

Kyoto University plans to disband its celebrated Primate Research Institute (PRI) because of misuse of government funds under former Director Tetsuro Matsuzawa. By March 2022, the university will move PRI's research groups to other institutes and university departments. The institute's 12 chimpanzees and dozens of monkeys will continue to be housed and will still be used for research, but PRI's operations may be gradually scaled back.

The decision was first reported by the *Sankei Shimbun* newspaper on 14 October, although a shake-up had been rumored for several weeks. "This is a sad and shocking outcome," John Mitani, a primate behavioral ecologist at the University of Michigan, Ann Arbor, wrote in an email to *Science*. "PRI is arguably the leading center for the study of primates in the world."

"This outcome … is bound to be devastating for Japanese primatology."
Frans de Waal, Emory University

A spokesperson for Kyoto University says the school has no comment on the decision, which has not been finalized. But sources close to PRI confirmed the impending closure. Matsuzawa, who was dismissed from his professorship at the Kyoto University Institute for Advanced Study in November 2020 and is now a visiting researcher in psychology at the California Institute of Technology, told *Science* he has "no comments" at this time.

Kyoto University also has a Wildlife Research Center housing chimps and bonobos released from biomedical research in a sanctuary in Kumamoto, on the southern island of Kyushu. It is unclear what will happen to that facility, though sources expect it to be scaled back as well.

PRI researchers won international recognition for their surprising observations of Japanese macaques washing dirt off sweet potatoes before eating them and passing this cultural practice on to descendants in the wild. The institute's scientists were also in the forefront of documenting tool use by chimpanzees in Africa and showing that younger animals picked up this know-how by watching their seniors. Work with captive chimpanzees revealed unexpected cognitive abilities, such as connecting colors with their written kanji characters, and extraordinary memory. "Kyoto researchers have had an enormous impact on international primate studies," says Nobuyuki Kutsukake, a specialist in animal behavior at Graduate University for Advanced Studies in Hayama, near Tokyo. He adds that PRI also trained primatologists from throughout Asia.

To provide a more natural environment, Matsuzawa won $10 million in funding to construct large, interconnected cages on PRI's campus in Inuyama, near Nagoya, and at the Wildlife Research Center. Investigations by the university and the Board of Audit of Japan found mismanagement of construction contracts involving 1.1 billion yen ($9.6 million), according to a university report released last year. Questionable practices included working with favored contractors and not specifying building details until after contracts were awarded. Matsuzawa was never accused of taking money for personal use or diverting it to other projects.

In a statement posted on his website in November 2020, he acknowledged mistakes were made, but wrote that the conclusions of the investigative committee were based on improper fact finding. In May, he sued the university seeking reinstatement, according to local press reports.

The *Sankei Shimbun* says Kyoto University has already returned about 900 million yen to the Japan Society for the Promotion of Science, a government funding agency, and the Ministry of Education, Culture, Sports, Science and Technology may ask for additional reimbursements. The university may rely on savings from disbanding PRI and reducing staff to raise the money.

Primatologist Frans de Waal of Emory University says he can't comment on whether the disciplinary action is warranted. "But [I] do wish to express my dismay at this outcome, which is bound to be devastating for Japanese primatology," de Waal wrote in an email to *Science*. PRI's facilities "are first-rate and we can only hope that the animals will be taken care of into the foreseeable future." ■

Faced with unidentified human remains, forensic anthropologists can sometimes determine their race.

PHOTO: ISTOCK.COM/D-KEINE

FORENSICS

Anthropologists grapple with identifying remains by race

Professional group is developing new standards for linking human remains to specific populations instead of races

By **Lizzie Wade**

When an unidentified body arrives in the laboratory of Allysha Winburn, a forensic anthropologist at the University of West Florida, it's her job to study the bones to help figure out who the person was when they were alive—to give the biological remains a social identity. "We have this vast population of possible missing persons the [remains] could match, and we need to narrow down that universe," she says.

She measures the length of the limb bones to estimate height and examines the bones' development to estimate age at death. She studies the shape of the pelvis for clues to the person's likely sex. And, until recently, Winburn measured features of the skull, such as its length and the width of the nasal opening, to do what anthropologists call ancestry estimation. By comparing the measurements with those from skulls with known identities, she could predict the continental ancestry—and the commonly used racial categories that may correspond to it—that a person likely identified as when alive. In other words, she could predict whether they identified as Black, white, Hispanic, Asian, or Native American.

But Winburn, who is white, is now questioning whether she should continue to do so. And she's not the only one: Over the past year, debate about ancestry estimation has exploded in U.S. forensic anthropology, with a flurry of papers examining its accuracy, interrogating its methods, and questioning its assumptions. A committee of the American Academy of Forensic Sciences is now hammering out a new standard that would, if adopted, direct professionals away from racial categories and toward more specific social and biological populations, such as Japanese or Hmong instead of Asian.

The standard will likely come to a vote by the end of the year, with potential implications for the hundreds of U.S. forensic anthropologists, as well as the more than 20,000 missing and nearly 14,000 unidentified people in the United States, and their loved ones. "This debate has been percolating for decades," says Shanna Williams, a forensic anthropologist and anatomist at the University of South Carolina School of Medicine, Greenville. Now, "There's more of a willingness to challenge the status quo."

Two biological and forensic anthropologists ignited the latest round of debate in the summer of 2020, in a letter to the *Journal of Forensic Sciences*. Jonathan Bethard of the University of South Florida and Elizabeth DiGangi of Binghamton University argued that ancestry estimation remains dangerously tangled up in the "race science" of the 19th and 20th centuries. Many of today's anthropologists were trained to identify race using the same techniques earlier generations of scientists employed to argue for biological differences and hierarchies among races. Today's anthropologists know those scientists were wrong both biologically and ethically. But translating skeletal data into race, a socially determined category, still reifies the erroneous notion that race is biological, Bethard and DiGangi argue. "Ancestry estimation is race science, pure and simple," says DiGangi, who is Black and biracial.

The pair also wrote that the criminal justice system might give less attention to remains when they are classified as members of marginalized groups. ["We can't] assume that our work is not harmful," says Bethard, who is white. The pair urged forensic anthropologists to stop performing ancestry estimates and study whether the practice results in discrimination.

The letter triggered an explosion of debate. "It took a lot of courage for them to put that letter out," says Kyra Stull, a forensic anthropologist at the University of Nevada, Reno. "It's been a catalyst for a lot of positive change."

But Stull disagrees with Bethard and DiGangi, and she co-authored a response to the article that argues ancestry estimates are important tools. "Right now, the reality is that social race is part of how people identify in the United States, and that follows them [in death]," says Stull, who is white. Race is included in missing persons reports, police case files, and in almost every other description of a person. "That has to change for [ancestry estimation] not to be useful."

"Many of the unknown individuals that come to us are from disenfranchised populations," adds Williams, who is Black. "When you take off the table a parameter that could help somebody get home to their family … then it's not the greater good."

The fact that ancestry estimation sometimes works "does not in any way, shape, or form mean that [races] are biological categories," stresses Agustín Fuentes, an anthropologist at Princeton University who is Hispanic and white. There's no checklist of skeletal, physical, or genetic traits shared by all people of a certain race; in fact, there's far more variation within racial categories than between them.

At the same time, "We're all a product of our environment, evolution, and history," says Ann Ross, a forensic anthropologist at North Carolina State University. People who share both deep evolutionary history and more recent social contexts, such as an

industrial lifestyle or a history of discrimination, tend to also share some biological traits, including similar cranial measurements. But Ross and some other anthropologists wonder whether current ancestry estimates are accurate enough to be helpful.

Winburn and a colleague did a study to try to find out. Among about 250 resolved cases in which forensic anthropologists offered an ancestry estimate, they correctly identified a person's social race about 90% of the time, the team reported in April in the *Journal of Forensic Sciences*. But when anthropologists identified someone's ancestry as "mixed" or "other," they were wrong 80% of the time. In the United States, 13% of unidentified people are listed as likely belonging to "multiple races"; another 21% are classified as "uncertain." Based on her research, most of them are likely to be people of color, Winburn says. "Who are we serving and who are we failing with these continental ancestry estimates?" she asks.

Some forensic anthropologists who want to jettison the five commonly used racial categories say skeletal features could let them make more specific, useful distinctions. For example, cranial measurements can distinguish between Maya groups from Guatemala and Mexico and differentiate each from non-Maya people, according to a February study in the *American Journal of Physical Anthropology*. Such specificity could help identify people who die crossing the U.S.-Mexico border, the study argued.

The United States's traditional racial categories don't reveal such potentially informative variation—they obscure it, Ross says. People from any population in Mexico, Guatemala, and the rest of Latin America—as well as many populations in the United States—are all classified as "Hispanic," a label Ross, who is from Panama, calls "biologically meaningless." The proposed standard, which Ross is helping draft, will reflect that concern, promoting the use of "population affinity estimates" that refer to more specific social and biological groups.

Winburn supports the change to population affinity, but says more research is needed to ensure it lives up to its potential. For example, many reference samples are themselves categorized by social race or continental ancestry and will have to be refined for population affinity studies. "We're going to have to get creative," she says.

In the meantime, Winburn won't provide ancestry estimates until and unless she can take a population approach that doesn't focus on reductive racial categories. She no longer wants to risk using a tool that she thinks may do more harm than good. ■

PHOTO: HORACIO VILLALOBOS/CORBIS/GETTY IMAGES

COVID-19

What tore apart *The Lancet*'s COVID-19 origin probe?

Former members cite confidentiality dispute and challenge concerns over conflicts of interest that derailed effort

By **Jon Cohen**

An effort to probe the origin of the COVID-19 pandemic that was intended to sidestep politics has foundered amid accusations of conflicts of interest and bias. The head of *The Lancet* COVID-19 Commission, an interdisciplinary initiative set up by the prestigious medical journal to improve the world's response to the pandemic, last month quietly and abruptly dissolved its origin task force after 10 months of work. That news, first reported by *The Wall Street Journal*, was preceded by a dispute over academic freedom, *Science* has learned, and also reflects bitter and deepening divisions over whether a natural spillover of a virus was the trigger or laboratory studies might have played some role.

Jeffrey Sachs last month abruptly ended a task force on COVID-19's origin.

The commission's chair, Columbia University economist Jeffrey Sachs, says he was concerned that five of the 12 task force members, including the original and current chairs, had potential conflicts of interest because of their direct and indirect ties to a lab in Wuhan, China, at the center of the origin debate. Gerald Keusch, the current chair, charges that Sachs didn't trust that the task force members would give the lab-origin hypothesis a fair evaluation and attempted to influence how they conducted their work, including demanding access to interviews with experts on both sides of the debate who had been promised confidentiality.

"Our mistake was to think that we were appointed as an independent autonomous group of experts without the supervision, intervention, or the micromanagement by the sponsor, the COVID-19 commission," says Keusch, who is associate director of Boston University's National Emerging Infectious Diseases Laboratory Institute.

The demise of the task force, one of 12 under the commission's umbrella, comes after the World Health Organization's (WHO's) high-profile effort to probe the charged issue also stumbled. Critics complained it downplayed the lab-origin possibility to appease China, and WHO last week rebooted the effort with a new group of scientists (see p. 380).

The odds looked better for the effort organized by *The Lancet*, which does not answer to governments and has a long history of forming commissions that largely consist of leading academic scientists who are independent. To chair the task force Sachs initially chose zoologist Peter Daszak, one of the world's most experienced researchers of bat coronaviruses. Daszak runs a research nonprofit, the EcoHealth Alliance, that has long collaborated with the Wuhan Institute of Virology (WIV), a focus of suspicion for lab-origin advocates. They have suggested SARS-CoV-2 traces back to bat samples WIV researchers either collected in the wild or created, sparking the pandemic when it somehow escaped. Daszak has called such suggestions a "conspiracy theory."

In a draft of a 10 December 2020 letter EcoHealth provided to *Science*, Sachs boasted about the quality and objectivity of the origins task force. "This a very highly regarded and experienced global group of scientists," Sachs wrote to Representative Bill Posey (R–FL), who had asked about the task force. "The Commission and the task force have no preconceived conclusions in the outcome of this work. We are interested in ascertaining the truth."

Former President Donald Trump and many others pilloried EcoHealth—and, by extension, Daszak—for collaborating with WIV, which led the National Institutes of Health (NIH) to defund a grant to the nonprofit. At first, Sachs strongly defended his choice of Daszak. He told Posey, for example, that Daszak was open to "all theories" and said his ties to WIV would be an asset,

noting he "is in a position to query senior China's officials [sic] at WIV and elsewhere regarding Covid-19 and its antecedents."

But as new details of EcoHealth's work and its collaborators were revealed, Sachs says, his doubts about Daszak grew. "It was not well known to me how potentially conflicted he was."

Sachs adds that in the spring he requested to see EcoHealth's grants and proposals with NIH, but Daszak refused. "He told me those were confidential," Sachs says. "I said there was no way he could be task force chair or on the commission if he didn't share those." The other 11 people on the task force refused to remove Daszak from their ranks, but agreed to make Keusch their chair instead.

Daszak declined an interview request, but a spokesperson for EcoHealth noted that Sachs "was fully aware of our work with the Wuhan Institute of Virology." Richard Roberts of New England Biolabs, a Nobel Prize winner who organized a letter with 76 other laureates castigating NIH for canceling EcoHealth's NIH grant at Trump's urging, guffaws at the accusation that Daszak failed to disclose conflicts of interest. "This is just ridiculous—he made no effort to pretend that he was not a part of the EcoHealth Alliance and that they had contacts with the Wuhan institute," Roberts says.

Last month, Sachs says, his concerns about conflicts broadened beyond Daszak to other task force members. On 10 September, he learned details of another NIH grant to EcoHealth, revealed following Freedom of Information Act requests from *The Intercept*. Keusch and three other task force members are listed as co-investigators. "None of them reported this involvement with the EcoHealth Alliance grant, though they had been asked to do so," Sachs says. "In these circumstances, I ended the task force."

Filippa Lentzos, a sociologist at King's College London who specializes in biosecurity issues, believes Sachs was "absolutely right" to dissolve the task force. "At this point in the pandemic, building trust in how its origins are investigated is absolute key," Lentzos says.

Keusch, however, calls Sachs's conflict-of-interest concern "disingenuous to the extreme." Keusch, who learned on 15 September about Sachs's decision to dissolve the group, adds, "The grant was an application under review for a networking agreement and it involved no funding—only the potential for NIH support."

Instead, Keusch asserts, Sachs's decision reflected his own biases. "Anybody who had a connection to EcoHealth became persona non grata," Keusch says. "I had a long email to Jeff, which said you're conflating expertise, collaborations, or connections with conflict of interest."

Sachs's concerns also baffle the University of Melbourne virologist Danielle Anderson, a task force member who was on the EcoHealth grant. "Instead of being clearly defined by specific criteria, conflict of interest became a personal, shifting opinion that I believe was politically motivated," says Anderson, who had long collaborated with WIV.

Keusch charges Sachs with interfering with the task force's work. Keusch, Daszak, and other members planned to speak with about 50 experts holding a range of views about the pandemic's origin, promising them confidentiality so they could speak freely. But in the spring, Sachs told the task force he wanted access to the interview transcripts. "We immediately started an ongoing set of discussions with him about the task force, its integrity, its autonomy," Keusch says.

> **"There were ways that he could have handled this that would have been less problematic."**
>
> **Stanley Perlman**, University of Iowa

Keusch and other task force members say Sachs also pushed them to investigate biosecurity at WIV and add new members who had publicly promoted the lab-leak hypothesis.

Sachs acknowledges the dispute. "I didn't approve of secret, confidential discussions," he says, because the approach "lacked transparency." He also says it's "an open question" how SARS-CoV-2 first infected humans, and he thinks the task force was too conflicted to explore it objectively.

That suggestion offends University of Malaya virologist Lam Sai Kit, a task force member who was not on the grant revealed by *The Intercept*. Lam, who helped discover the Nipah virus in 1999, says, "I take great umbrage to this [dismissal] as a scientist who has worked in the field of emerging infectious diseases, especially those caused by viruses with a zoonotic origin. Personally, I feel that injustice has been done to us by *The Lancet* commission."

The showdown with Sachs need not have happened, says task force member Stanley Perlman, a coronavirus researcher at the University of Iowa. "There were ways that he could have handled this that would have been less problematic," Perlman says.

Several members of the task force say they hope the group can reincarnate itself under another organization and continue its work. ■

CLIMATE CHANGE

Hunt begins for ancient Antarctic ice

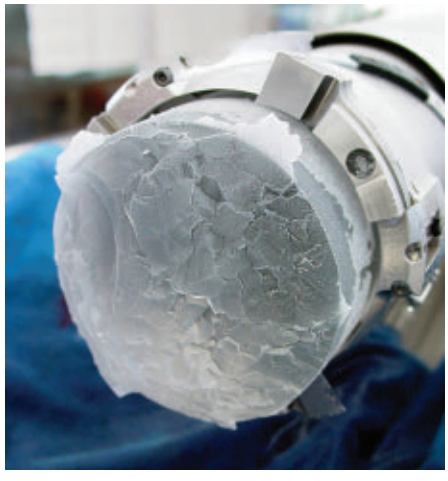
Six teams plan to drill cores that could shed light on warm climates 1.5 million years ago

By **Paul Voosen**

In East Antarctica, ice has piled up, year after year, to more than 3 kilometers thick. It preserves many secrets: meteorites from outer space, dust from ancient volcanic eruptions, and mysterious subglacial lakes. But the most prized scientific treasures from this realm are the tiniest: trapped air bubbles that record shifts in Earth's climate. By drilling deep cores from the ice, scientists have read this record as far back as 800,000 years.

Now, six international drilling teams are attempting to double that history—reaching back 1.5 million years. "It's kind of become a trend," says Dorthe Dahl-Jensen, a glaciologist at the University of Copenhagen. "Everyone has to drill for the oldest ice." It's not just for bragging rights: The goal is to probe a time when Earth was warmer and ice ages were shorter than today, in search of clues to how Earth will react as the climate warms again because of greenhouse gas emissions.

One leading effort, an €11 million European-funded project, will begin drilling



This Antarctic ice core contained air bubbles tracing 800,000 years of climate history.

PHOTO: REUTERS/L. AUGUSTIN/LABORATOIRE DE GLACIOLOGIE ET GEOPHYSIQUE DE L'ENVIRONMENT (LGGE)/HANDOUT

this year. The scientists flew to New Zealand this month, where they will isolate to avoid carrying COVID-19 on their journey south to their target drill site, Little Dome C, high on the Antarctic plateau. They will have company there: About 6 kilometers away, scientists from Australia will also begin drilling this year.

China began its own drilling in 2012, at Kunlun Station on Dome A, near the highest point on the Antarctic Ice Sheet. But progress has been slow, because reaching the remote location leaves only a few weeks to drill each year. Japan and Russia plan to begin to drill similar cores later this decade at Dome Fuji and Dome B. And last month, the United States funded a $25 million Center for Oldest Ice Exploration (COLDEX), led by Oregon State University (OSU), to hunt for its own drill site.

The "domes" the teams are targeting are stable regions in East Antarctica where the ice stands high. Snow falls slowly but steadily at those sites, trapping air bubbles that survive as the snow is compacted into ice.

The teams hope to extract a perfect layer-cake record of climate changes stretching back 1.5 million years. But success is far from certain. Shifting ice can fold young ice layers beneath older ones. Unexpected melting from subsurface heat can create gaps. "The oldest ice project is inherently high risk," says Kenji Kawamura, a glaciologist at Japan's National Institute of Polar Research, who is leading the Japanese effort. The redundant drilling efforts, he says, better the odds that at least a couple will succeed, allowing the cores to corroborate each other.

Geopolitical ambitions are also driving the multiple drilling efforts. With its $33 million campaign at Little Dome C, Australia is hoping to reestablish its ability to travel deep into East Antarctica from its oceanside bases. Similarly, Russia will erect a new base at its historic Vostok drilling site so it can attempt to drill for old ice 300 kilometers away at Dome B, says Vladimir Lipenkov, an ice core scientist at Russia's Arctic and Antarctic Research Institute.

Once the precious cores are returned to laboratories, geochemists plan to analyze the ancient air inside the bubbles. They can directly measure ancient carbon dioxide (CO_2) levels, and from the mix of carbon isotopes can infer how much of the gas came from biological sources such as decaying plants, versus geological ones like volcanoes. Meanwhile, the ratio of oxygen isotopes from the melted ice can point to temperatures and storm frequency. Together, those data can shed light on what Dahl-Jensen, a leader of the European Beyond EPICA drilling project, calls a "powerful argument in our climate debate": how much of a role CO_2 played in past temperature shifts.

Dahl-Jensen's project is a successor to the European Project for Ice Coring in Antarctica (EPICA), which in 2004 finished drilling an 800,000-year-old core, the current record holder, at Dome C. But the ancient climate it revealed was much the same as the modern climate, at least before human-driven global warming set in: a comparatively cold period when the world went in and out of deep 100,000-year-long ice ages, prodded by tiny swings in Earth's orbit.

The new cores would open a window on a warmer climate, more than 1 million years ago, when ice ages occurred every 40,000 years. They might also reveal whether changes in atmospheric carbon were behind Earth's jump to the 100,000-year ice age cycle. In one theory, carbon had little role: It holds that until 1 million years ago, the northern ice sheets formed on top of a thick layer of soil that, like grease, kept them unstable and relatively short-lived. Once the ice sheets scraped away the soil and hit bedrock, however, they grew taller and more persistent, perhaps leading to enhanced reflection of sunlight and other feedbacks that drove cooling and extended their life span (*Science*, 24 August 2018, p. 739). In another view, however, the trigger for longer ice ages was a small drop in CO_2 levels, perhaps driven by dust-fertilized plankton in the Southern Ocean that absorbed the gas from the atmosphere.

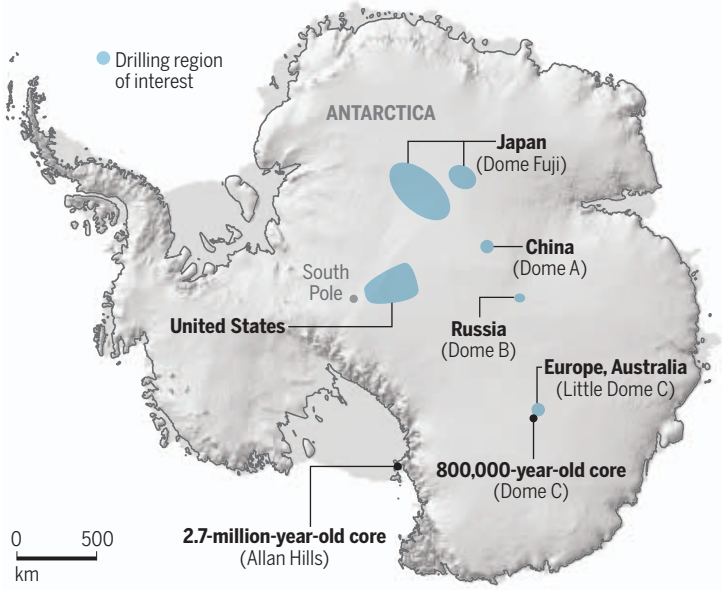
In the past few years, ice even older than the drillers will reach has offered tantalizing clues. In shallow "blue ice" regions in the Allan Hills of Antarctica, ice flows across buried rocky ridges, driving layers as old as 2.7 million years upward while harsh winds strip away younger ice (*Science*, 18 August 2017, p. 630). Although blue ice samples lack the easy dating and high resolution of the continuous ice cores, they revealed that CO_2 levels in the earlier, warmer period were on average quite similar to levels seen in the past million years, a hint that a CO_2 drop might not have triggered the climate shift. "We feel like we've just scratched the surface of these blue ice regions," says Ed Brook, COLDEX's principal investigator at OSU Corvallis.

The COLDEX team will continue to explore the blue ice even as the researchers hunt for a continuous core at a deep drill site somewhere between the South Pole and Dome A. They will scout cautiously, deploying a heated probe that melts its way through the ice all the way to the bedrock, measuring dust as it descends. Dust levels can be compared with known records as a quick and dirty way to gauge the age of the ice. "If we don't get older than 800,000 years, we'll say OK, that's not a good place," Brook says.

Wherever the U.S. researchers end up, they will be able to build on the experience of the European and Australian teams, which hope to hit bedrock by the middle of this decade. The teams have already discussed how best to make use of their cozy proximity, by Antarctic standards, and are even talking about getting together for an outdoor soccer match, says Joel Pedro, a paleoclimatologist at the Australian Antarctic Division who leads the project. "Though we'd probably struggle" against the Europeans, he says. "It might have to be a cricket match—so we'll win." ■

Frozen in time

In search of clues to Earth's ancient climate, six teams will drill cores containing ice as old as 1.5 million years. The new cores would surpass the age of one from Dome C, and could rival ancient but discontinuous samples from Allan Hills.



MAP: C. BICKEL AND N. DESAI/*SCIENCE*

A GOLDEN MENACE

An invasive mussel is devastating ecosystems as it spreads through South American rivers, threatening the Amazon basin

By **Sofia Moutinho**

On a stuffy summer day, Maria Cristina Dreher Mansur peers into the brownish water of Lake Guaíba in the southern Brazilian city of Porto Alegre. While locals watch the sunset from the shore, the 76-year-old retired biologist wades up to her neck, carefully dragging her bare feet on the muddy floor. She feels with her toes for hard shells and retrieves them with a dip net.

Mansur, an expert in bivalve mollusks, has been coming to this spot with students for more than 30 years to monitor local species. In the 1990s, the shore was covered with sedge and the lake bottom was strewn with the rounded shells of native mollusks. Now, the shore is barren and the aquatic ecosystem unrecognizable. "I could fill a bucket with native species in a few minutes," Mansur remembers. "Now, it can take me a whole morning to find half a dozen."

Behind the transformation are tiny mollusks with arrow-shaped, caramel-colored shells, native to the Yangtze River in China: golden mussels (*Limnoperna fortunei*). When Mansur first noticed them on the shore in 1998, she recognized them at first sight based on warnings from colleagues who were already tracking them in Argentina. The species had arrived in the Americas a few years earlier in the ballast water of ships coming from Asia.

Months after Mansur first spotted them, the animals had taken over Lake Guaíba, plugging the pipes that supplied the city with water. Masses of dark, thumb-size adults mixed with golden pinkie nail–size newborns encrusted bedrock, boats, piers, and bridges, forming dense reeflike structures with more than 200,000 individuals per square meter. Without any local predator to control them, they choked and rotted the roots of plants along the shore, and even grew on top of other animals such as native mollusk species and crabs, suffocating them. "In 2 years, the golden mussel transformed the lake's sandy beaches and vegetated margins into piles of dark and stinky shells," Mansur recalls.

Similar scenes have played out in many rivers, lakes, fisheries, and hydroelectric dams as the mussels have advanced north through Brazil at a speed of about 240 kilometers per year. Now listed as one of the three invasive species of most concern in the country, they are far from the first aquatic invader to gain a foothold in Brazil. But their speedy reproductive cycle and unique ability to cling to surfaces have made their economic and ecological impacts more severe—similar to those caused by the zebra mussels infesting the Great Lakes in the United States, but much more widespread.

An even bigger catastrophe looms: the invasion of the Amazon and its tributaries, part of the largest drainage basin in South America, which spans eight countries and is one of the richest hot spots for biodiversity on the planet. Golden mussels have been documented in the Pantanal wetlands just 150 kilometers from the Téles Pires River, which

PHOTOS: (LEFT TO RIGHT) GOVERNOR JOSÉ RICHA HYDROELECTRIC POWER PLANT/COPEL; CARLOS EDUARDO BELZ

Golden mussels (shown as newborns on top of an adult, right) grow in reeflike structures that cover underwater surfaces and clog pipes in hydroelectric power plants (left).

flows into the Amazon basin and connects to the Tapajós River, a tributary of the Amazon.

"It only takes one boat encrusted with the mussel to cross the wetlands for the invader to make a new home in an Amazon river," says biologist Marcia Divina at the state-owned Brazilian Agricultural Research Corporation. If the invader spreads in the Amazon, it could wipe out native species that scientists have not even studied yet, she adds. "We can't even calculate the size of the impact."

Despite a late and anemic government response, researchers funded largely by affected hydroelectric companies have developed new tools to track the mussels' relentless advance. And some are looking to an aggressive form of genetic engineering to eradicate it. That untested strategy is still likely years from being ready to deploy, but scientists see few other ways to slow an invader so easily spread by human transit and trade.

IN 1991, Argentinian biologist Gustavo Darrigran of the National University of La Plata was collecting snails for a class when he spotted the first golden mussels in South America, on rocks at the La Plata River. He knew the animals didn't belong there; the continent has no native freshwater mussel species. He counted five mussels per square meter. Four years later, there were 150,000.

Golden mussels attach to surfaces by releasing superadhesive, almost invisible filaments known as byssal threads. The threads allow them to take over areas unexplored by native mollusks—none of which has this sticking power. Their reproductive powers are also formidable. If water is warmer than 15°C (as most Brazilian rivers are year-round), they spew thousands of eggs and sperm. These unite to form microscopic, free-floating larvae that attach to a new surface and grow. At 1 year old, the sesame seed–size mollusk starts to reproduce.

The population Darrigran observed in the La Plata eventually stabilized at 85,000 per square meter in 2019, the last count before the pandemic. But by then, everything around it looked different. Golden mussels are "ecosystem engineers," Darrigran explains, "altering the interactions between the local fauna and flora forever." As mussels build up on a riverbed, they impede water flow, which allows organic matter to accumulate, nourishing a new set of insect larvae, snails, and small gastropods. On average, the mussels attract 10 to 15 new species when they take over, while crowding out previous occupants. "You always find the same species in invaded areas," Darrigran says, "and the local diversity is lost."

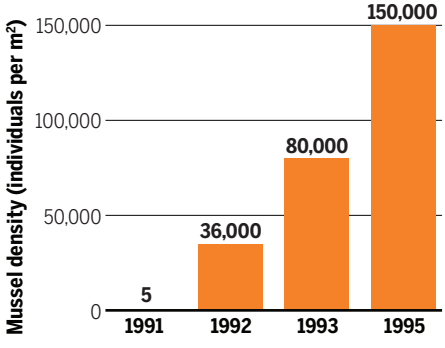
The mussels eat by filtering plankton, monopolizing nutrients that would otherwise feed other river-dwelling species or flow into the ocean. A single adult filters, on average, half a liter of water per hour—about 10 times more than the zebra mussels infesting the U.S. Great Lakes. The water gradually becomes clearer and lets in more light, which encourages blooms of toxic cyanobacteria. A 2008 assessment of Southern Brazil's Alto Paraná River found the mussels had completely wiped out colonies of freshwater sponges—species endangered in Brazil that act as nurseries for larvae of several types of aquatic insects.

Similar damage now threatens the Amazon. More than 2500 fish species unique to the basin depend on the nutrient-rich Amazon water and its resident plants and algae. The Amazon accounts for 15% of the world's river discharge into the oceans; depleting nutrients that support sea life in the Atlantic Ocean "will affect this balance on a still-unknown scale," Darrigran says.

Humans are aiding the mussels' advance. In 2009, veterinarian Carlos Belz, who was doing his Ph.D. on golden mussels at the Federal University of Paraná, Curitiba, set out to understand how these tiny creatures that can't swim have spread north—against river currents—so quickly. He and a colleague stood at a highway patrol post near the banks of the already-infested Paraná and, with help from the police, stopped 32 trucks ferrying

Exponential invasion

After arriving in Argentina, golden mussels quickly multiplied on rocks at the La Plata River. The population there is now stable at 85,000 per square meter.



CREDITS: (GRAPHIC) K. FRANKLIN/*SCIENCE*; (DATA) GUSTAVO DARRIGRAN/NATIONAL UNIVERSITY OF LA PLATA

sand to artificial beaches along its tributary, the Iguaçu. Sifting through the sand, they found golden mussel shells and two living mussels. The species can survive for up to 10 days out of water, Belz says.

In other rivers, Belz found mussels and larvae in water retained in fishing boats and in the stomachs of fish, which eat the mussels and defecate them without digesting their hard shells. "To control the spread of the golden mussel, we would need to attack on so many fronts … it's almost impossible," Belz says. "And that is why no one is really doing anything about it."

MANSUR HAD hoped the Brazilian government would spring into action when she first reported the golden mussels in Lake Guaíba in 1998. She urged measures such as requiring fishers and riverboat operators to bleach boat hulls, bilges, and wells after passing through an affected area. She also recommended a ban on throwing the mussels back into the water after cleaning them off dams and other structures. But when Mansur called local and federal environmental agencies, she got no response.

Only in 2003 did the Ministry of the Environment create a task force to propose control measures. Implementation of such proposals "should have started 10 years ago," Mansur says. In contrast, North America appears to have kept golden mussels out so far, likely thanks in part to interagency plans for early detection of invasive species put in place after the U.S. zebra mussel invasion, and requirements that ships treat ballast water and empty it far off the coast before docking. Such ballast water disposal has been mandatory in Brazil since 2005, but oversight has been lacking, Belz says.

The golden mussel invasion in Brazil did raise alarms, however, for the hydroelectric power industry, whose dams are being overrun. By now the mussel has invaded about 40% of Brazil's hydropower plants—with costly consequences. The Governor José Richa Hydroelectric Plant on the Iguaçu, which generates energy for 4 million people, spends about 1 million reais (about $200,000) per year pumping toxic chemicals that curb mussel proliferation into the water tubes that surround and cool its turbines. If the tubes clog, water stops running and the turbines can overheat and stop. Sharp mussel shells coat the stairs and handrails in the dam's chambers, making it hard for crews to do routine maintenance.

A preprint published in 2018 in *PeerJ* estimates the mussel plague costs the Brazilian electrical sector $120 million per year. And available control methods can only keep mussel populations stable enough to keep plants running—not eradicate the species or prevent its ecological damage. Because there is no regulation on how to dispose of the mussels, most plants throw them back into the river by the ton, fueling the spread.

The threat to their business makes electric utilities the biggest supporter of research on golden mussels. With funding from power plants, a startup company called ATGC, incubated at the Federal University of Paraná, has developed a simple way to detect and measure infestations using traces of genetic material the mussels leave in rivers.

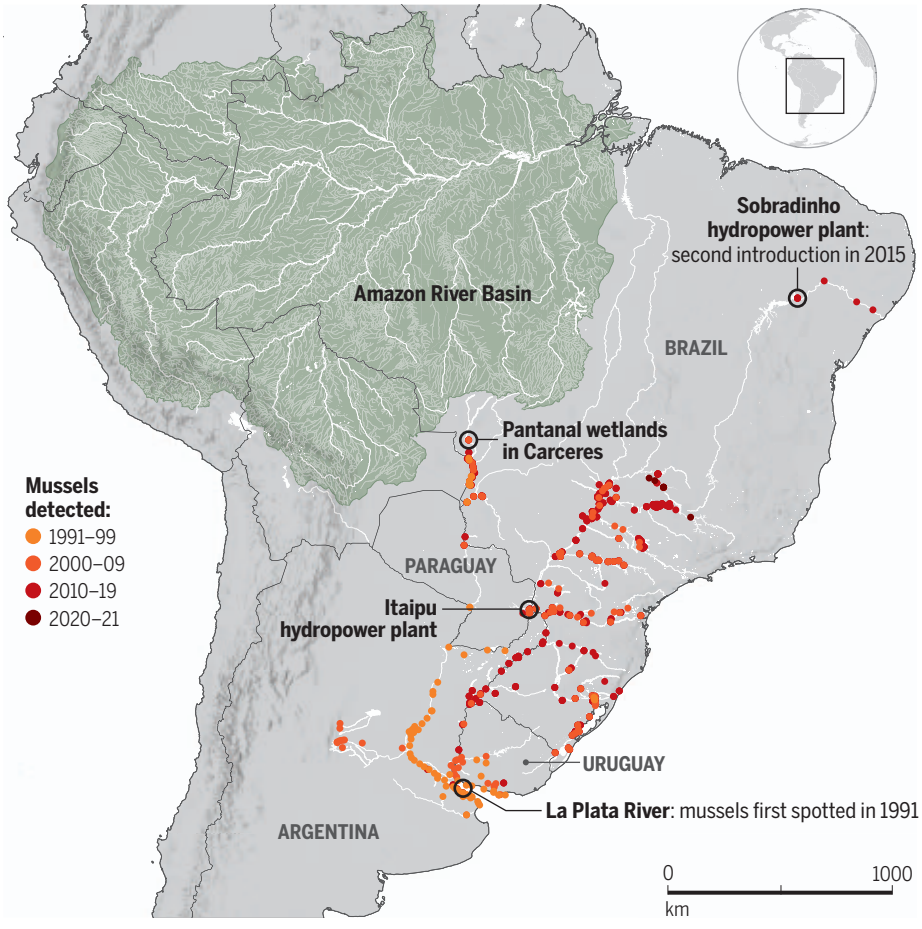
ATGC geneticist Marcio Pie has identified two genetic sequences unique to the golden mussel and developed a polymerase chain reaction (PCR) test, like the ones used for COVID-19, that detects these sequences in water. In lab experiments, the test could detect a single mussel in 950,000 liters of water. The researchers also use a fluorescent labeling technique to count particles of mussel DNA, which allows them to track changes in the density of infestations. "Now we can monitor the golden mussel in real time," Pie says.

Each month, the company receives plastic bottles of river water from 15 hydroelectric dams already infested or worried about an infestation. Pie hopes the strategy can be expanded into a national monitoring network. For now, information on the mussels' spread comes from sparse network of research teams and hydropower plants that voluntarily report new invasions on a collaborative website. "The biggest challenge will be to coordinate actions in a country as huge as Brazil," Pie says.

In 2020, Brazil's Institute of the Environment and Renewable Natural Resources (IBAMA) finally launched a national plan to monitor and control the golden mussel. The plan, collectively created by authorities, academics, and industry, established goals for the next 25 years to mitigate the invader's impact and prevent its arrival in the Amazon basin. The measures include fostering research, informing the public about the invasive species, and creating legislation that would, for example, require hydropower plants to check for the mussels in their dams.

A path of destruction

Since the 1990s, golden mussels have been advancing north from Argentina at a speed of 240 kilometers per year. In 2015, a second invasion began at the Sobradinho hydropower plant in northeastern Brazil. The mussels are now within 150 kilometers of the Amazon River Basin, one of the richest biodiversity hot spots in the world.



CREDITS: (MAP) K. FRANKLIN/*SCIENCE*; (DATA) HTTPS://BASE.CBEIH.ORG

But the pandemic and federal budget cuts have delayed the planned actions, says biologist Rogerio Santos at the Federal University of Mato Grosso, Cuiabá, who is part of the national task force for controlling the mussel's spread. Underfunded control efforts have focused on local measures that protect individual dams and private companies, he says. Meanwhile, scientists know more about the invader than the local fauna it is destroying, which makes it hard to measure the mussels' toll. He believes the arrival of the golden mussel in the Amazon basin is inevitable and worries that stopping it is still not a priority of the federal government.

Requests for comment from IBAMA on its response to the golden mussel went unanswered.

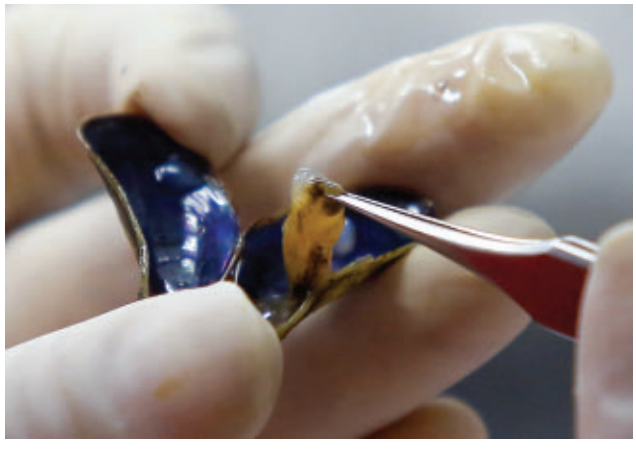
Studies suggest Amazonian rivers' water temperature, salinity, and nutrient content are perfect for the golden mussel. In 2010, a mathematical model created by Fabiano Alcísio Silva at the Bioengineering Centre of Invasive Species and his team predicted a high risk of invasion in the Amazon basin by 2050. Those calculations may have been conservative. They ignore, for example, the fact that new Brazilian hydropower plants are legally required to breed fish in their reservoirs to compensate for damage to local fisheries, which often means importing exotic species. Golden mussels can hitch a ride in the guts of these fish or in the water used to transport them. According to Brazil's Ministry of Mines and Energy, 79 new Amazonian hydroelectric dams are in the planning stages—each one a potential site of a new mussel introduction.

In July, Silva found a few golden mussels in a water sample from a power plant in an Amazonian river. But two subsequent PCR tests did not detect the animal's DNA in samples from the same dam. He is now testing more water to determine whether that first sample signals the start of an invasion. "Not all invasions are necessarily successful," he points out. "Usually, an invader has to try many times."

ON THE DOOR of a small laboratory at the Federal University of Rio de Janeiro's main campus, next to a sign with a biohazard symbol, is a drawing of a saber-toothed tiger, the logo of the '80s cartoon series *ThunderCats*. The show follows a group of heroic humanoid cats who defend their planet from invaders. The lab's main investigator, biologist Mauro Rebelo, says he draws inspiration from the show and its motto, "Sight beyond sight." "Here, we all have sight beyond sight, working with molecular biology and things you can't see with the naked eye," Rebelo says.

On the other side of the door, Rebelo's team is developing a mutant mussel that could mate with wild, invasive mussels and produce infertile offspring, wiping out the population. Rebelo, whose work is supported by electric companies, is not only picking a fight with the golden mussel, but with inheritance itself. Normally, when two animals mate, their offspring inherit half of their genes from each parent, and the next generation ends up with only 25% of the genes from their grandparents. But a genome-altering strategy called gene drive breaks those rules, allowing a modified gene to spread through a population much faster than natural inheritance allows.

Rebelo's group aims to spread genetic modifications that interfere with fertility. The technique relies on the gene-editing tool CRISPR-Cas9, which uses molecular "scissors" to make precise changes to DNA. The researchers will use these scissors to cut DNA in golden mussel sperm. They'll insert both a sequence that disrupts fertility genes and a sequence encoding the scissors themselves. When mutant sperm fertilize eggs in the lab, the offspring will have the fertility-blocking mutations—and the genetic instructions for the scissors—in their germ cells, which give rise to gametes (sperm and eggs).



One golden mussel filters half a liter of water per hour to obtain nutrients.

That's where the gene drive steps in to turn these offspring into superspreaders of infertility: When they reach adulthood and produce gametes, the scissors, which are designed to only be active in germ cells, will cut out and replace the normal fertility genes inherited from the mother so that all the gametes carry the altered genes. When these mutant mussels breed with wild mussels, the offspring will still carry normal fertility genes from their wild parent, so they won't be infertile. But all of their gametes will carry the mutations. As the population of modified mussels grows, the chance of two mutants breeding—creating fully infertile offspring—increases. And as mutants slowly replace the original population, it collapses entirely.

Rebelo's team estimates that infertile mussels would make up 50% of the population 2 years after the release of the modified mussels, and 90% after 10 years. "One modified mussel can release millions of modified sperm," Rebelo says. "My modified mussel factory is the mussel itself." The researchers expect to have the first mutant mussel next year and will then test its power to crash populations in tanks.

But they are still far from releasing the animals in nature. Although several teams are developing gene drives to control disease-carrying mosquitoes and other pests, the strategy has never been tried outside of the lab. Brazilian law doesn't regulate gene drives specifically, but it requires time-consuming studies to assess the environmental impact of genetically modified animals before approving their release.

Other researchers fear that once released, a gene drive will be hard to control. Sandra Ludwig, a geneticist studying golden mussels at the Federal University of Espírito Santo, Goiabeiras, worries the fertility-blocking modification might disappear over time—or that mussels could even develop and spread mutations that increase fertility, making the population "worse than the golden mussel we have today."

After more than 20 years of following the golden mussel's conquest, Mansur would like to see the genetically modified mussel working. But she sees the species advancing at a faster pace than the sparse scientific efforts to stop it. "I am not seeing much hope," she confesses. At best, she says, the mussel invasion serves as a warning to anticipate and stop the next threat.

In a drawer in her home office, she keeps her recent nightmare: the palm-size, brown striped Chinese pond mussel (*Sinanodonta woodiana*). Producers in Europe and Costa Rica imported it to cultivate pearls in artificial ponds. But it has escaped into the wild and is now spreading, threatening native biodiversity and affecting water quality in more than 21 countries. Mansur has alerted the Brazilian Ministry of the Environment about the threat, and there's still time to restrict its import to Brazil. "If this leaves a tank and enters the environment here … oh God!" she says. "We already have the golden mussel, we cannot let this animal enter, too." ■

Sofia Moutinho is a journalist based in Rio de Janeiro.

PHOTO: AP PHOTO/LEO CORREA

EVOLUTIONARY BIOLOGY

Of war, tusks, and genes

Societal conflict leaves an evolutionary signature in wildlife

By **Chris T. Darimont**[1,2] *and* **Fanie Pelletier**[3]

Elephants have long been dragged into war. Referred to as "elephantry," military units rode into battle atop these giants over millennia. On page 483 of this issue, Campbell-Staton *et al.* (*1*) describe the evolutionary aftermath of a different type of wartime elephant use. Seeking ivory to finance a civil war in Mozambique, poachers relentlessly targeted specific African elephants (*Loxodonta africana*), individuals with tusks, sending the population—and the frequency of this important trait—into decline. The study reveals the consequence of this intense selective killing on the persistence of tusks and the genes associated, as well as for population dynamics. Campbell-Staton *et al.* also identify the simple genetic architecture underlying the presence of tusks. The findings bring new evidence to inform debates on the roles of environmental and selective forces underlying trait variation in populations subject to harvest.

Over and above the environmental variation to which populations are subject, selective killing of specific phenotypes can influence traits such as body or ornament (for example, horn, antler, or tusk) size. Through multiple analyses, Campbell-Staton *et al.* report how intensive selective killing of African elephants caused the rapid evolution of increased tusklessness in females. Field data revealed that the proportion of tuskless females increased by more than 30% as the population declined over 28 years, which included 15 years of civil war. The frequency of tuskless phenotypes among adult females born after the war was also higher than before the conflict, suggesting an evolutionary response. Simulations showed that the observed increase in tusklessness is extremely unlikely to have occurred without selective killing of tusked animals. Model outputs

PHOTO: ELEPHANTVOICES

Elephants in Gorongosa National Park, Mozambique, declined by 90% during the 20-year civil war because of ivory poaching. This spurred the rapid evolution of tuskless females, such as this female *Loxodonta africana*.

estimated that the survival of tuskless individuals was five times higher than that of tusked individuals.

Campbell-Staton *et al*. also investigated whether these phenotypic changes were accompanied by a genetic signature. Analyses of whole-genome sequences from individuals with and without tusks supported the hypothesis of a more severe population decline among tusked compared with tuskless individuals. The authors then looked for a pattern of inheritance that could explain the variation in tusk morphology observed in the field. Using data on mother-offspring phenotypic associations, they found that for ~9 out of 10 offspring, the phenotype of the offspring was consistent with a single-locus X-linked dominant model of inheritance. The observed sex bias in the offspring produced by tuskless mothers suggest that tuskless male offspring were nonviable. Using a candidate gene approach, they then identified two genes involved in tusk presence that explain a large amount of variation: *AMELX* (*X-linked isoform of amelogenin*) and *MEP1a (meprin A subunit alpha*), which are known to have functional associations with the development of mammalian teeth. Physical linkage between *AMELX* and male-lethal loci nearby on the X chromosome may explain lethality among males inheriting the trait.

Campbell-Staton *et al*.'s elegant approach is among the rare studies to document a genetic response to harvest selection, informing debate about the potential for selective harvests to lead to evolutionary responses. Work on trophy hunting of bighorn sheep (*Ovis canadensis*) in Canada, for example, has shown that size-selective regulations with no harvest quotas can lead to the evolution of smaller horns (*2*), although the genetic mechanism remains to be elucidated. The lack of evidence linking observed phenotypic trends to changes in the genetic composition of harvested populations more broadly, however, has been used by wildlife and fisheries managers and scientists to argue against the likelihood or importance of potential evolutionary changes in harvest systems (*3*). The comprehensive work by Campbell-Staton *et al*. has clearly satisfied the burden of evidence, showing that selective killing can indeed leave a strong evolutionary signature. Restoration of the trait and its associated ecosystem function might therefore require longer time scales than those for phenotypic changes not associated with genetic changes, an important implication relevant to other systems.

Generalization of the findings on African elephants is constrained by the relatively simple genetic basis underlying tusklessness. Although other studies have shown that ornament traits can be influenced by a single gene [such as horn size in Soay sheep, *Ovis aries* (*4*)], it is more commonly observed that traits involved in response to environmental changes, including selective mortality, are affected by both many small-effect genes and in some cases one or a few genes of major effect (*5*). Accordingly, a quantitative understanding of whether and how much the phenotypic changes observed in myriad animal populations subject to harvesting (*6*) are associated with genetic changes remains a complex challenge (*7*).

More broadly, even perfect knowledge of underlying genetic contributions do not address social-evolutionary processes that influence nonhuman life in today's world. An extreme social event (a war, in this case) that triggered intense, selective exploitation of elephants crisply illustrates the pronounced coupling between human societies and evolutionary processes in other life forms. Through humanity's cultures, economies, medicines, built environments, and more, societies have set in motion selective landscapes never before experienced by the world's biota (*8*). Moreover, and often related to ecological changes imposed by humans, societies must commonly respond to the evolution of other organisms, a reality brought into painful relief during this extended COVID-19 pandemic. Recently, such complex relationships among society, ecology, and evolution have been well examined in perhaps the most radically changed of all landscapes, cities (*9*). The conceptual advances gained in urban systems can inform work in other contexts in which humanity's hand in, and response to, evolution will likely also be observed. Future work could also draw upon the well-developed theoretical basis for understanding sustainability in what has been referred to as "social-ecological systems" (*10*), which could be adapted to consider evolutionary interactions and outcomes.

Progress in understanding these complex relationships will require more interdisciplinary research. Until recently, natural and social sciences have largely been independent enterprises. In the context of harvest selection, evolutionary ecologists could team up with social scientists. They may ask, for example, how hunters and fishers would trade-off, personally and economically, the ability to target the largest phenotypes now with the specter of losing those phenotypes in the future. As noted in the context of ecology (*10*), past and current societies and the institutions that govern them might have considered similar trade-offs and sought solutions from which researchers and managers can potentially learn. ■

[1]Department of Geography, University of Victoria, Victoria, BC, Canada. [2]Raincoast Conservation Foundation, Bella Bella, BC, Canada. [3]Département de Biologie, Université de Sherbrooke, Sherbrooke, QC, Canada. Email: darimont@uvic.ca

REFERENCES AND NOTES

1. S. C. Campbell-Staton *et al*., *Science* **374**, 483 (2021).
2. G. Pigeon, M. Festa-Bianchet, D. W. Coltman, F. Pelletier, *Evol. Apps* **9**, 521 (2016).
3. A. Kuparinen, M. Festa-Bianchet, *Philos. Trans. R. Soc. Lond. B Biol. Sci*. **372**, 20160036 (2017).
4. S. E. Johnston *et al*., *Nature* **502**, 93 (2013).
5. R. A. Bay *et al*., *Am. Nat*. **189**, 463 (2017).
6. C. T. Darimont *et al*., *Proc. Natl. Acad. Sci. U.S.A*. **106**, 952 (2009).
7. F. Pelletier, D. W. Coltman, *BMC Biol*. **16**, 7 (2018).
8. S. R. Palumbi, *Science* **293**, 1786 (2001).
9. S. Des Roches *et al*., *Evol. Apps* **14**, 248 (2020).
10. E. Ostrom, *Science* **325**, 419 (2009).

ACKNOWLEDGMENTS

We thank M. Festa-Bianchet and A. Hendry for insights.

10.1126/science.abm2980

ORGANIC CHEMISTRY

Iron-catalyzed coupling of three reactants

Vinyl boronates couple under mild conditions to two partners through carbon-carbon bonds

By **Guillaume Lefèvre**

New methods that efficiently form carbon-carbon bonds are welcomed by organic chemists if they run with high yield in few steps, introduce structural diversity, and if catalysts and co-reagents are required, are as inexpensive and nontoxic as possible. Multicomponent reactions (MCR) that allow reaction of three or more species in a single, convergent process introduce structural diversity in one step. Efficient MCR-based processes can provide an easy access to the synthesis of molecule libraries, which is a valuable asset for the fabrication of families of drug-like compounds. On page 432 of this issue, Liu *et al.* (*1*) describe an efficient three-component reaction that difunctionalizes vinyl boronates by means of iron coordinated with a bulky phosphine ligand. Grignard reagents (ArMgX, where Ar is an aryl group and X is a halide) and alkyl halides (R′X) are partners in these transformations that occur under very mild conditions (see the figure, top).

Iron-catalyzed cross-couplings have been extensively exploited in carbon-carbon bond formation processes developed during the past few decades [for a complete review, see (*2*)]. Those methods involve the creation of a covalent bond between two reagents through the reaction of an electrophile (typically an electron-poor derivative, such as an organic halide or pseudo-halide such as tosylate) with a nucleophile (typically an organometallic reagent bearing an electron-rich metalated carbon atom—for example, an organomagnesium, an organozinc, or less often, an organoboron compound). The low cost and the environmentally friendly and biocompatible nature of iron make this metal particularly suitable for synthetic applications in the development of products of pharmaceutical or biological interest (*3*). Many iron-catalyzed two-component cross-couplings have been reported, and some of them are now used on large preparative scales. However, the state of the art for iron-mediated multicomponent transformations that involve organoboron derivatives has been particularly limited so far.

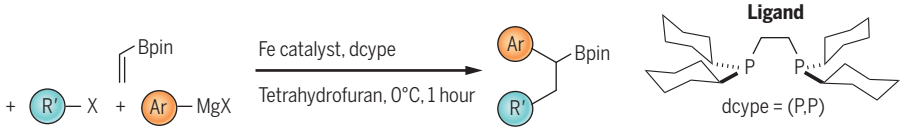
Vinyl boronates are particularly useful platforms in radical processes because of ease of generation of α-boryl radicals that have been used as MCR partners in nickel-photoredox catalysis, given their intrinsic enhanced stability (*4*). The methodology discussed by Liu *et al.* relies on the successful association of a simple iron salt with an electron-rich bulky diphosphino [or (P,P)] ligand—in this case, 1,2-bis-dicyclohexylphosphinoethane (dcype)—and encompasses a large scope of reagents. This method also enables the use of fluoroalkyl halides as radical precursors, which leads to the incorporation of fluoroalkyl chains in the final product as well as the development of annulation-arylation sequences. These capabilities help address two outstanding challenges of iron-coupling catalysis. Such a versatility opens the way for a large number of synthetic combinations and makes a broad range of organic scaffolds accessible in a single step.

Liu *et al.* performed thorough mechanistic studies using iron-57 Mössbauer spectroscopy as well as isolation and characterization of key intermediates by use of x-ray diffraction. These studies demonstrate the crucial role of the number of hydrocarbyl groups transferred by the Grignard reagent ArMgX onto the starting iron(II) complex [for a recent report on the benefits of physical-inorganic techniques in the elucidation of mechanistic facets of iron-mediated processes, see (*5*)]. Depending on their coordination number, organoiron(II) species can display very distinct reactivities, and their oxidoreduction properties also strongly vary (*6*–*9*). Thus, mastering and exploiting the reactivity panel offered by the diverse coordination possibilities of organoiron species is often the key of a successful iron-catalyzed coupling sequence.

The catalytic process was initiated by a single-electron transfer between in situ generated bis-arylated complex $(P,P)Fe^{II}(Ar)_2$ and the alkyl halide. As is often observed in iron-catalyzed two-component cross-coupling chemistry, this step ensures the formation of a free carbon-centered radical. The latter does not undergo a fast recombination with a transient organoiron, which is the usual final step in a classic two-component cross-coupling process. Instead, this radical is added to the carbon-carbon double bond of the vinyl boronate (see the figure, bottom). The newly formed radical then forms the expected difunctionalized product after recombination with a

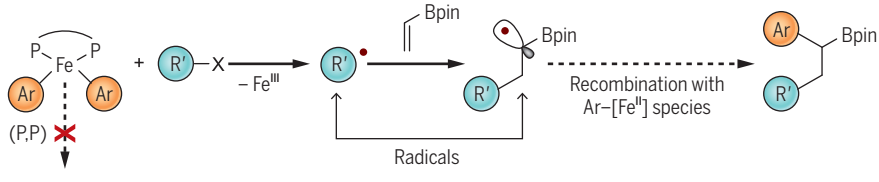
Iron mediates a three-component reaction

Structural diversity can be achieved more rapidly if three reactants can form a single product in one step. Liu *et al*. report the difunctionalization of a vinylboron mediated by an iron catalyst under mild conditions.



Double addition to vinyl boronates

The one-step reaction between a vinyl boronate (Bpin, vinylboronic acid pinacol ester), a Grignard reagent (where Ar is aryl), and an organic halide (where R′ is an alkyl group and X is a halide) is catalyzed by a bulky diphosphine-ligated iron complex (the structure of the dcype ligand is shown on the right).



Mechanistic steps

Key steps of the transformation are shown. The bulky dcype ligand prevents binding of a second ligand that would form an inert complex. The activated alkyl radical reacts with the vinyl boronate to create a second radical, which undergoes a final redox step to complete the catalytic cycle.

Chimie ParisTech, PSL University, CNRS, Institute of Chemistry for Life and Health Sciences, CSB2D, 75005 Paris, France. Email: guillaume.lefevre@chimieparistech.psl.eu

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

mono-aryliron(II) intermediate, and a final redox comproportionation step completes the catalytic cycle.

Elucidation of the structure of the active species also showed that control of the chelation of the (P,P) ligand (dcype in this case) to iron(II) is critical. This ligand helps selectively form unsaturated intermediates bearing only one dcype, unlike other (P,P) ligands with similar electronic properties, which formed saturated octahedral $(P,P)_2Fe^{II}(Ar)_2$ species. Those species are unable to undergo a single-electron transfer with the organic halide, so no catalytic activity is observed.

The selective formation of species ligated by one dcype ligand originates in the high steric pressure brought by the cyclohexyl substituents of dcype, which hampers the inclusion of a second dcype in the iron coordination sphere. Thus, the judicious choice of this substituent limits the distribution of iron-containing intermediates to mono-dcype species and allows the crucial formation of the unsaturated bis-arylated $(P,P)Fe^{II}(Ar)_2$ after transmetallation with the Grignard reagent. Both steric and electronic effects of the dcype ligand contribute to the feasibility of this step.

The combination of a cutting-edge design of the interlocking of coordination properties with stereoelectronic ligand effects governing the reactivity of this bulky iron-diphosphine platform allowed Liu *et al.* to develop a new and robust synthetic strategy. It can potentially benefit various fields—such as synthesis of pharmaceuticals, materials, and pesticides—in the near future. Beyond this breakthrough in synthetic applications, this work also illustrates the importance of accurate mechanistic investigations in the field of iron catalysis. The distinctive and complex reactivity displayed by iron, involving both one- and two-electron processes, as well as a broad range of oxidation and spin states should spur the development of new synthetic methods that will require a detailed understanding of the mechanistic aspect of these transformations (*10*). ■

REFERENCES AND NOTES

1. L. Liu *et al.*, *Science* **374**, 432 (2021).
2. I. Bauer, H.-J. Knölker, *Chem. Rev.* **115**, 3170 (2015).
3. R. B. Bedford, P. B. Brenner, in *Iron Catalysis II*, E. Bauer, Ed. (Springer, 2015).
4. M. W. Campbell, J. S. Compton, C. B. Kelly, G. A. Molander, *J. Am. Chem. Soc.* **141**, 20069 (2019).
5. M. L. Neidig *et al.*, *Acc. Chem. Res.* **52**, 140 (2019).
6. R. B. Bedford *et al.*, *Angew. Chem. Int. Ed.* **53**, 1804 (2014).
7. S. L. Daifuku, J. L. Kneebone, B. E. R. Snyder, M. L. Neidig, *J. Am. Chem. Soc.* **137**, 11432 (2015).
8. Z. Mo, L. Deng, *Coord. Chem. Rev.* **350**, 285 (2017).
9. L. Rousseau, N. Touati, L. Binet, P. Thuéry, G. Lefèvre, *Inorg. Chem.* **60**, 7991 (2021).
10. A. Fürstner, *ACS Cent. Sci.* **2**, 778 (2016).

10.1126/science.abl4762

NEUROSCIENCE

Toward collective animal neuroscience

Social intelligence requires the study of both groups of brains and the individual brain

By **Julia Sliwa**

From humans to fish to ants, social animals coordinate their behavior when in groups, and their neural system might do so as well. Yet the individual brain, studied in isolation, has so far been the fundamental unit of analysis of most neuroscience research. Animal wisdom has been searched for in how neurons interact within individual brains. But what if embedding the brain in a group changed how it works? And even more boldly, what if the laws of wisdom were to be found in how brains interact in a group, rather than in how neurons interact in a single brain? On pages 422 and 421 of this issue, Rose *et al.* (*1*) and Báez-Mendoza *et al.* (*2*) show that brains of bats and rhesus macaques track group interactions by, respectively, coordinating their activity and representing the group in single-neuron activity.

Group living enables species to come up with solutions that each individual would not have accomplished alone. These include reducing predation or accomplishing substantial tasks, such as ants' nests or human cities. These exceptional endeavors emerge from individuals' social interactions, which are sometimes as simple as copying local neighbors and sometimes more elaborate, involving memory of others and signaling (*3*–*5*). The study of social interactions in neuroscience has been carried out by recording neural activity in the brains of two individuals simultaneously and led to the discovery that brains in a group of two can couple their activity (*6*). Brain-to-brain coupling has been reported in various social interactions, including communication, cooperation, competition, and learning. It has been shown to affect neural activity at different spatiotemporal scales (including neuronal firing, neural oscillations, and hemodynamic responses) and in different brain regions. Further, the temporally coherent link between the activity of two brains was either synchronous or asynchronous and directed from one brain to the other. Brain-to-brain coupling has been demonstrated with increasingly sophisticated analysis methods and in different species, including bats, mice, and monkeys (*7*–*9*).

Yet it is only when multiple social interactions are coupled together that the power of a collective dynamic emerges. The change of scale from studying one social interaction between two individuals to studying collective social interactions within a group of individuals must go hand in hand with innovative methods. In human studies, the strategy has been to choose devices that are easy to use in a real environment, such as a classroom or a music studio, while being well coordinated to simultaneously record brain activity within the group in real time (*10*–*12*).

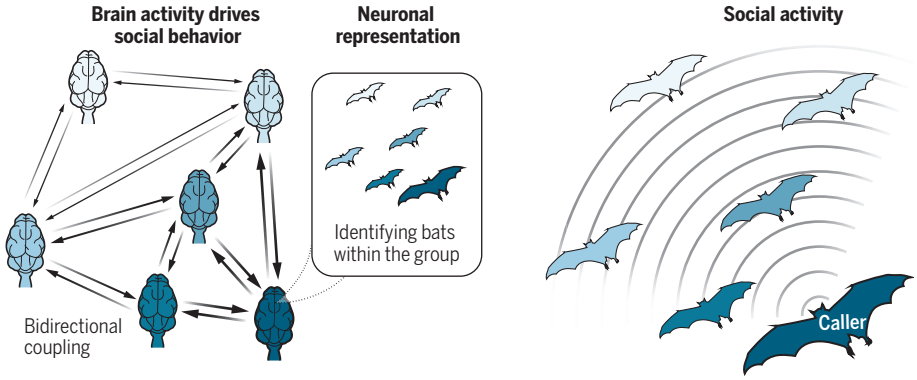
In their investigations with Egyptian fruit bats and rhesus macaques, Rose *et al.* and Báez-Mendoza *et al.*, respectively, took another direction. They did not want to trade off the resolution of single-neuron activity. They focused on the frontal cortex, which is involved in making and pondering decisions; the medial region is a hub of social cognition. They recorded from close, but not homologous, brain areas in the two species, allowing them to identify the activity of neurons in the frontal cortex that is sensitive to the identity of the individuals that the bats and monkeys interacted with. This means that interactions were tailored to the specific interacting partner (see the figure). In bats, this neuronal sensitivity was even more precise for close partners within the group than for distant partners, which were less well differentiated by neurons. In monkeys, dorsomedial prefrontal cortex neurons were active only during interactions; looking at the partner did not activate these neurons. When the interacting partner was a monkey plush puppet or a bat playback sound, these neurons stopped being active altogether. This contrasts with neurons representing peers in the nonhuman primate temporal cortex that are active when their peers' faces are presented as pictures (*13*). In the frontal cortex, these neurons appear to be tuned for possible action from peers, representing them as communication partners, competitors, and collaborators.

When socially interacting, the brains of two bats will couple their activity by syn-

Sorbonne Université, Paris Brain Institute–Institut du Cerveau, Inserm, CNRS, APHP, Hopital Pitié-Salpêtrière, Paris, France. Email: julia.sliwa@icm-institute.org

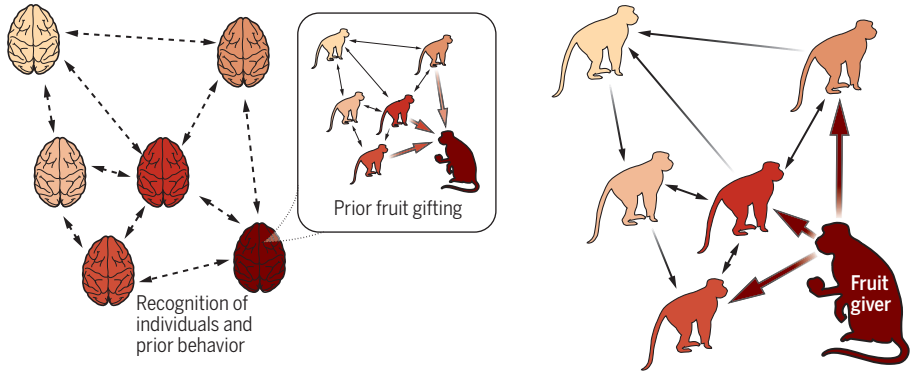
Animal group social interactions

Groups of interacting bats and monkeys exhibit neural activation in the frontal cortex that drives social behavior.



Bats: Synchronized neural activity in the frontal cortex

When socially interacting, such as a bat calling out to the group, bidirectional coupling of brain activity occurs within group members listening to the call. The strength of interbrain coupling depends on the socially central position of the caller.



Rhesus macaques: Reciprocating gifts according to previous interactions

A monkey can choose to give a piece of apple to members of the group. But their choice depends on the history of each monkey in giving gifts of apple when it was their turn. This reciprocation behavior was determined by the activity of single neurons in the dorsomedial prefrontal cortex.

chronizing their frontal cortex neural oscillations as well as neuronal spiking activity (*7*). However, Rose *et al.* found that when a bat emits a call, interbrain coupling applies to all listening bats. Calling not only widely distributes information within the group but also induces collective brain coupling. By providing shared auditory stimulation to all brains at once, it could be that a communication call is not more specific than a gong or a bell ring in inducing collective brain coupling. However, communicative calls were found to be very specific in how they affected collective brain coupling within the bat group. The authors trained bats to emit vocalization to receive food from humans in the presence of other bats, but these trained calls did not elicit the same collective brain activity coupling as real communication. The caller's brain was no longer influenced by activity from the listener's brain when emitting trained calls. Conversely, real communication relied on bidirectional interbrain coupling, where the caller's brain accounts for the listener's brain before they send a signal.

Should research now focus on groups? In social insects, such as ants or bees, individuals are unable to represent the collective outcome of their group. Conversely, individual humans and most nonhuman primates can recapitulate part of the knowledge of the group, including the structure of their society and its culture, and use it for their own purpose (*14*). This makes it interesting to study how a single brain accomplishes this advanced cognitive ability. Báez-Mendoza *et al.* show how a group of rhesus macaques derive economic laws from their trading interactions and represent this knowledge in their brain to guide future decisions. The authors used a simple, yet highly controlled, task that taps into the foundation of economic trade in humans: giving, receiving, and reciprocating (*15*). Monkeys sitting at a turntable were sequentially distributing pieces of apple to each other. Monkeys chose strategically to reciprocate gifts of apple to monkeys that gave them apple and to retaliate against those who did not give to them. In their decisions, monkeys not only considered the last gift received but also the history of giving and even the reputation of individuals to give to guide their own behavior. Such behavior was absent when monkeys could reward plush monkey totems.

During the economic trade, Báez-Mendoza *et al.* found that single-neuron activity from the dorsomedial prefrontal cortex reflects reward history in the group, monitoring the social outcomes of other monkeys' choices. Based on neuron activity, they were able to decode past interactions. Future reciprocation and retaliation were also predicted, even before the monkey decided to give a piece of apple. The neuronal activity that they discovered not only reflected monkeys' behavior but also was necessary for performing reciprocation behavior. When neuronal activity of the medial prefrontal cortex was temporarily disrupted—through reversible microstimulation—just before a monkey made their decision, reciprocation was diminished.

The findings of Rose *et al.* and Báez-Mendoza *et al.* are major steps in identifying the neural mechanisms for maneuvering in complex social structures, including those for interacting with others and for knowing the specific identity of the interactors and the history of those interactions. It remains unclear how neurons from the same area can be involved in these different processes at the same time. It will be important to understand which neurons' activity contributes to synchronization versus encoding social knowledge. Are there separate populations of neurons, or are the same neurons doing both tasks? The resolution with which brain activity is probed will be key to answering this question. These studies also emphasize that in the future, it will be necessary to navigate between two levels of investigation to fully grasp social intelligence: that of the group of brains and that of the individual brain. Further research should assess additional dynamical properties of animal and human collectives, including changes that may result from virtual collectives on the internet. ■

REFERENCES AND NOTES

1. M. C. Rose *et al.*, *Science* **374**, eaba9584 (2021).
2. R. Báez-Mendoza *et al.*, *Science* **374**, eabb4149 (2021).
3. I. Couzin, *Nature* **445**, 715 (2007).
4. A. Pentland, *Adapt. Behav.* **15**, 189 (2007).
5. J. Krause *et al.*, *Trends Ecol. Evol.* **25**, 28 (2010).
6. G. Dumas *et al.*, *PLOS ONE* **5**, e12166 (2010).
7. W. Zhang, M. M. Yartsev, *Cell* **178**, 413 (2019).
8. L. Kingsbury *et al.*, *Cell* **178**, 429 (2019).
9. K. M. Gilbert *et al.*, *bioRxiv* 10.1101/2021.02.08.430294 (2021).
10. S. Dikker *et al.*, *Curr. Biol.* **27**, 1375 (2017).
11. V. Müller *et al.*, *Ann. N. Y. Acad. Sci.* **1423**, 198 (2018).
12. J. Yang *et al.*, *Nat. Neurosci.* **23**, 754 (2020).
13. S. M. Landi *et al.*, *Science* **373**, 581 (2021).
14. D. Cheney *et al.*, *Science* **234**, 1361 (1986).
15. M. Mauss, *The Gift: Forms and Functions of Exchange in Archaic Societies* (Cohen & West, 1966).

ACKNOWLEDGMENTS

Funding is from the Agence Nationale de la Recherche (ANR-21-CE37-0016-02, ANR-19-CE37-0002-01, and ANR-10-IAIHU-06) to J.S.

10.1126/science.abm3060

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

MICROBIOLOGY

Bacterial defense islands limit viral attack

Mobile phage-defense genes underpin rapid evolution of resistance in nature

By **Sean Meaden**[1,2] and **Peter C. Fineran**[2,3,4,5]

In 1927, Felix d'Herelle attempted to treat cholera epidemics in India using viruses. Bacteriophages are viruses that infect bacteria, and d'Herelle was trying to use them as a natural weapon against the bacterial pathogen *Vibrio cholerae*—a strategy called phage therapy. Although d'Herelle had some success in treating cholera patients, he noted that the bacteria he studied could acquire resistance to the phages (*1*). Now, researchers are unraveling the complexity of the relationship between bacteria and phages using similar systems. Two studies, by Hussain *et al.* (*2*) on page 488 of this issue and by LeGault *et al.* (*3*), combine large ecological datasets, genomic analyses, and molecular genetics to clarify how phages and phage defenses coexist in the environment. Understanding these dynamics is important when considering the therapeutic use of phages and their major role in global biogeochemical cycles.

Both studies show that such defenses reside on accessory mobile genetic elements (MGEs). MGEs may be transferred between organisms by horizontal gene transfer (HGT), rather than through vertical parent-offspring inheritance. Such transfer through HGT decouples the defenses from the core genome and likely facilitates more rapid spread through bacterial populations. Phage defenses are frequently clustered together in bacterial genomes in "defense islands" (*4*), indicating that this genetic information has recently arrived in the genome through HGT and providing a useful resource for the discovery of phage defense systems (*5*, *6*). However, little is known about the drivers of phage resistance in natural ecosystems, leading instead to a reliance on in vitro systems that may not fully represent the diversity of resistance mechanisms present in the environment.

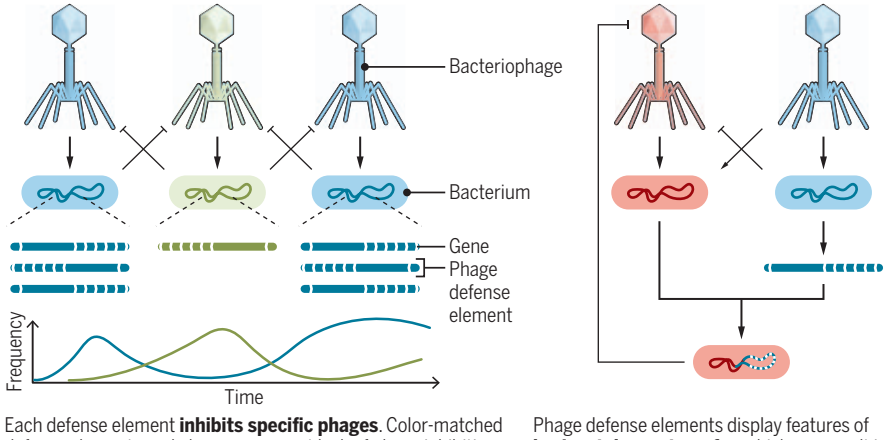
Both studies co-isolate vibrios and their phages during longitudinal sampling and use genomic analyses to tease apart the evolutionary dynamics between phages and phage defenses. LeGault *et al.* compare vibrios and phages in samples from cholera patients in Bangladesh over a 20-year period to measure coevolution. Notably, they report that infectivity fluctuates between years, and that these fluctuations are driven by MGEs called SXT integrative conjugative elements (ICEs), which carry genes that confer resistance against phage infection. SXT ICEs are self-transmissible MGEs that are key vectors of antibiotic resistance in *V. cholerae* (*7*). By contrast, Hussain *et al.* use a genome-resolved cross-infection network of marine vibrios, isolated from coastal waters in Massachusetts, USA, over 3 months to identify the genetic determinants of phage resistance. Similar to LeGault *et al.*, they discover resistance encoded on potential MGEs that show fast evolutionary turnover, which they call phage defense elements (PDEs) and are akin to genomic defense islands (*4*). Unexpectedly, they find that such PDEs constitute >90% of the flexible genome, which represents genes that are sometimes dispensable and vary across a bacterial population. This is consistent both in vibrios and other bacterial taxa, reinforcing that phages are a major evolutionary driver of bacterial accessory gene content (see the figure).

To prove that these islands are sufficient to prevent phage infection, LeGault *et al.* demonstrate that introduction of defense-carrying SXT ICEs into a strain naturally lacking similar elements leads to resistance. Notably, although LeGault *et al.* find that resistance fluctuates over time, Hussain *et al.* find a different pattern: Resistance is cumulative. They demonstrate that deletion of a single PDE in a vibrio does little to reestablish phage infectivity, but subsequent deletion of multiple PDEs restores the ability of phages to infect. These results show the hallmarks of certain types of coevolution (*8*). Cycling of resistance profiles over time is indicative of a fluctuating selection dynamic, where there is a benefit to being rare—either for the host bacterium or the phage. By contrast, escalating defense and counterdefense is typical of an arms-race dynamic (*8*). The results from these studies hint at which underlying dynamics may be taking place in each environment, with the patterns observed by LeGault *et al.* indicative of fluctuating selection, and the cumulative resistance observed by Hussain *et al.* more consistent with an arms-race dynamic.

These studies go further than existing work by combining experimental and molecular data with long-term field sampling that yields considerable mechanistic insight. Although both studies demonstrate that re-

Phage defense elements

The accessory genome of *Vibrio* bacteria was found to contain phage defense elements, which contain groups of genes that prevent bacteriophage infection.



Each defense element **inhibits specific phages**. Color-matched defense elements and phages represent lack of phage inhibition. The frequencies of these elements oscillate through time.

Phage defense elements display features of **horizontal gene transfer**, which can result in the transfer of phage resistance between bacteria.

GRAPHIC: N. DESAI/*SCIENCE*

[1]Biosciences, University of Exeter, Penryn, UK. [2]Department of Microbiology and Immunology, University of Otago, Dunedin, New Zealand. [3]Bioprotection Aotearoa, University of Otago, Dunedin, New Zealand. [4]Genetics Otago, University of Otago, Dunedin, New Zealand. [5]Maurice Wilkins Centre for Molecular Biodiscovery, University of Otago, Dunedin, New Zealand. Email: s.meaden@exeter.ac.uk; peter.fineran@otago.ac.nz

sistance to phages can be decoupled from the core bacterial genome, there are some key differences that could explain the potentially varied coevolutionary signals observed. LeGault *et al.* use isolates from a much longer 20-year period, whereas Hussain *et al.* investigate resistance over a single season. There may also be key differences due to the locations where the isolates were collected, population sizes, and their respective lifestyles: pathogenic *V. cholerae* versus free-living *V. lentus*. For example, pathogenic *V. cholerae* populations will be subject to frequent bottlenecks upon human infection, in turn reducing genetic diversity. Together, the ecology of these organisms may determine their exposure to phages and therefore explain the varying coevolutionary dynamics.

It may also be no coincidence that both studies investigate vibrios, a genus that is genetically tractable and easily isolated from environmental samples. In *V. cholerae*, restriction modification (RM) was playing a key role in resistance, whereby non-self DNA that does not carry the required DNA modifications is targeted for degradation. The recently discovered BREX (bacteriophage exclusion) system was also identified (*9*), which restricts phages through an unknown mechanism that relies on epigenetic modification of specific DNA sequence motifs (*10*). LeGault *et al.* also reveal how phages may respond to such defenses, identifying a candidate phage gene that overcomes restriction from the BREX system—similar to how phages can encode anti-CRISPR proteins and other antidefense systems (*11*). Notably, they find that stochastic epigenetic escape is important, where RM systems occasionally fail to impede phage infection, which results in the release of phages that have modified genomes and thereby enables evasion from host nucleases in subsequent infections. By identifying the genes responsible for defense and counterdefense, LeGault *et al.* again demonstrate a coevolutionary pattern in which phage genotype frequencies fluctuate alongside SXT ICE frequencies. Similarly, Hussain *et al.* identify RM systems residing on bacterial MGEs and also discover other genes involved in phage resistance that remain to be characterized. Thus, a wealth of defenses remain to be found on these PDEs and other defense islands (*5, 6*).

Much in vitro work has identified phage-receptor mutations as key drivers of bacterial resistance to phages (*8*), yet the studies of LeGault *et al.* and Hussain *et al.* identify intracellular defenses as the principal form of resistance. Moreover, Hussain *et al.* identify the receptors for two phages and find much lower genetic diversity at these sites than in the rest of the genome across all isolates, suggesting that they are under strong purifying selection (*12*). Overall, these results suggest that the routes to resistance that typically evolve in the lab are not fully representative of those that emerge in more complex environments. In part, this may be because natural microbial communities have a greater pool of genetic material to call upon for phage defense—the so-called pan-immune system (*13*)—whereas most experimental evolution studies are conducted in monoculture, which fails to address the importance of HGT.

These studies also offer a cautionary tale for the use of phages as therapeutic agents. Specifically, the rapid acquisition of resistance being transmitted through MGEs offers a parallel to the spread of antibiotic resistance on plasmids in bacteria. LeGault *et al.* also highlight the competitive interactions that can occur between MGEs. They find that acquisition of a key toxin-carrying phage (CTXϕ) is reduced when SXT ICEs are present, underscoring the trade-off between virulence factor acquisition and defense against phages that *V. cholerae* may face. In addition, if phage therapy is more widely adopted, it is important to avoid the spread of resistance through MGEs. Although no known antibiotic resistance genes were found on the PDEs described in *V. lentus*, the *V. cholerae* SXT ICEs are notorious for conferring antibiotic resistance. Perhaps most concerning, LeGault *et al.* report that phage infection itself increases the spread of SXT ICEs through conjugation. Previous research has shown that phages can both promote HGT of phage defenses through transduction (*14*) and can act as "superspreaders," releasing substantial amounts of intact plasmid DNA for transformation of bacteria following cell lysis (*15*). Although d'Herelle originally noted the ubiquity and importance of phages, the complexity of the relationship between phages and their hosts is still being unraveled. ■

REFERENCES AND NOTES

1. F. d'Herelle, T. L. Rakieten, *J. Infect. Dis.* **54**, 313 (1934).
2. F. A. Hussain *et al.*, *Science* **374**, 488 (2021).
3. K. N. LeGault *et al.*, *Science* **373**, eabg2166 (2021).
4. K. S. Makarova *et al.*, *J. Bacteriol.* **193**, 6039 (2011).
5. S. Doron *et al.*, *Science* **359**, eaar4120 (2018).
6. L. Gao *et al.*, *Science* **369**, 1077 (2020).
7. M. Spagnoletti *et al.*, *mBio* **5**, e01356 (2014).
8. M. A. Brockhurst *et al.*, in *Bacteriophages: Biology, Technology, Therapy*. (2021), pp. 231–251.
9. T. Goldfarb *et al.*, *EMBO J.* **34**, 169 (2015).
10. J. Gordeeva *et al.*, *Nucleic Acids Res.* **47**, 253 (2019).
11. L. M. Malone *et al.*, *Curr. Opin. Biotechnol.* **68**, 30 (2021).
12. A. R. Burmeister, P. E. Turner, *Curr. Biol.* **30**, R1120 (2020).
13. A. Bernheim, R. Sorek, *Nat. Rev. Microbiol.* **18**, 113 (2020).
14. B. N. J. Watson *et al.*, *mBio* **9**, e02406 (2018).
15. E. C. Keen *et al.*, *mBio* **8**, e02115 (2017).

ACKNOWLEDGMENTS

S.M. is funded by a Marie Skłodowska-Curie grant (EU Horizon 2020 no. 842656). P.C.F. is supported by the Marsden Fund, Royal Society of New Zealand, and Bioprotection Aotearoa.

10.1126/science.abm2444

MATERIALS SCIENCE

Step aside, aluminum honeycomb

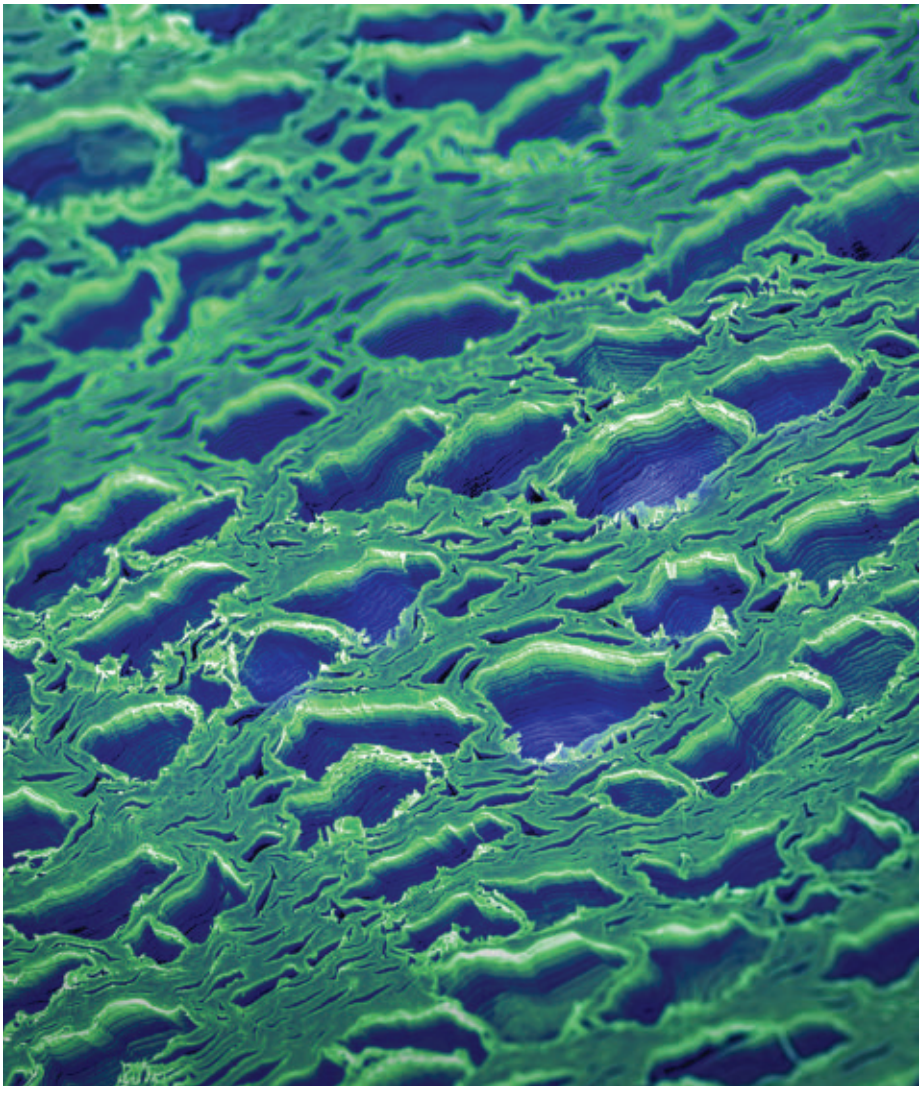
A two-step process can turn hardwood into a strong and flexible moldable material

By **Mehdi Tajvidi**[1,2] *and* **Douglas J. Gardner**[1,2]

Wood is a highly sophisticated material with a multilevel hierarchical structure that has evolved over eons to support trees that are hundreds of feet tall. The composite structure of wood functions both as the mechanical support and the means for water and nutrient transport within the living tree. Composed of long strands of fiber, wood tends to be strong and flexible along the grain but weaker and more brittle in the transverse direction. Because of this anisotropy in the mechanical properties of wood, certain processing and manufacturing techniques cannot be used. For example, three-dimensional compression molding is often off-limits for wood because of the tendency for compression failure to occur in its cellular structure. On page 465 of this issue, Xiao *et al.* (*1*) report on a workable modification and molding process for low-density, diffuse-porous hardwoods, opening the door for the design and construction of wood-based structures with complex geometries.

Pound for pound, wood is among the strongest building materials on the planet, especially when its anisotropic mechanical properties are used to its advantage. The outstanding strength of wood can be attributed to the hierarchical structure of the cell walls, the strength and stiffness of the cellulose nanofibrils and microfibrils, and the arrangements of the fibrils within the cell walls. A great number of research efforts have been made to improve the versatility of this natural material. For instance, by extracting nanocellulose—the structural building block of wood—and chemically modifying it to be used together with other materials (*2, 3*) and by controlling the orientation of the nanocellulose to take advan-

[1]School of Forest Resources, University of Maine, Orono, ME 04469, USA. [2]Advanced Structures and Composites Center, University of Maine, Orono, ME 04469, USA. Email: mehdi.tajvidi@maine.edu



Shown is a microscopic view of the wrinkled structure of the moldable wood. The larger holes are the wood vessels, also known as pores, after being selectively reopened during the "water shock" process. The walls of the fiber cells, however, remained collapsed and contain small lumens, sometimes known as cavities, which leads to the densification of the wood.

PHOTO: JIAQI DAI

tage of its direction-dependent properties (*4*), researchers have been able to produce strong and stiff products than can outperform some of the strongest human-made fibers and products.

Instead of focusing on nanocellulose, Xiao *et al.* focused on harnessing the micro- and nanoscale strength and stiffness of cellulose fibrils within the wood cell wall structure, as reported in the group's earlier work (*5*). Their manufacturing approach is a combination of existing modification techniques of delignification and densification (*6*).

Generally speaking, wood modification techniques can be categorized into either active or passive modification (*7*). Whereas active modification such as acetylation, cross-linking, or cell wall deterioration can alter the chemical structure of the cell wall, passive modification tends to retain the chemistry of the cell wall but not its physical structure. As an active modification process, the delignification process is widely used in papermaking, where the wood pulp is produced by removing lignin, followed by a bleaching process (*8*). Conversely, wood densification is normally conducted to improve surface hardness and strength and is a well-known passive wood modification technique, usually applied to the bulk wood or to the wood surface.

Xiao *et al.* combine these two modification processes into a new approach toward reengineering wood cell walls for better functionality and processability. By first partially removing the lignin—the glue that bonds wood cell walls together—and then drying and applying a facile "water shock," the internal structure of hardwood cells is altered in such a manner that it renders the resulting material easily bendable and foldable both along and across the grain. This flexibility is enabled by the selective reopening of some vessels (the pores) while leaving the densified fiber cells in their compressed state. Upon partial delignification and further processing, both cell lumens (the cavities) and vessels (the pores) collapse, leading to wood densification upon drying. The delignification process exposes more hydroxyl groups on the cellulose molecules, making them available for hydrogen bonding. Hydrogen bonding plays an important role in the strength and stiffness of the cellulose macromolecule, as well as wood adhesion, and has been recently reported to be stronger than previously thought (*9*). The subsequent exposure to water—the "water shock"—swells the cell walls, but because lignin is partially removed, the formation of hydrogen bonds keeps the cell walls together as the vessels regain their original shape. This process leads to the formation of a special wavy structure in the wood cell walls that enhances their flexibility for molding processes.

Potential applications for this wood modification technique are numerous. Xiao *et al.* provide examples of corrugated or honeycomb core composite structures that can be made using this technique. These cores can have wood or metal faces to form sandwich panels with outstanding mechanical performance while being lightweight and having good dimensional stability and durability compared with normal wood. The building, automotive, and packaging industries may also benefit from this new class of renewable and sustainable materials based on wood when addressing global challenges of sustainability. The life-cycle analysis of the new material is promising when compared with aluminum, providing the possibility of lightweight structures that are metal-free. Like many innovative manufacturing processes developed and demonstrated in a laboratory setting, the challenge will be to transition this wood manufacturing process to one that is commercially viable. ■

REFERENCES AND NOTES

1. S. Xiao *et al.*, *Science* **374**, 465 (2021).
2. R. J. Moon, A. Martini, J. Nairn, J. Simonsen, J. Youngblood, *Chem. Soc. Rev.* **40**, 3941 (2011).
3. Y. Habibi, *Chem. Soc. Rev.* **43**, 1519 (2014).
4. K. Li *et al.*, *ACS Nano* **15**, 3646 (2021).
5. C. Chen *et al.*, *Nat. Rev. Mater.* **5**, 642 (2020).
6. Forest Products Laboratory, "Wood handbook—wood as an engineering material" (General Technical Report FPL-GTR-282, US Department of Agriculture, Forest Service, Forest Products Laboratory, 2021).
7. D. Sandberg, A. Kutnar, G. Mantanis, *IForest* **10**, 895 (2017).
8. E. Sjöström, *Wood Chemistry: Fundamentals and Applications* (Gulf Professional Publishing, 1993).
9. D. J. Gardner, M. Tajvidi, *Wood Fiber Sci.* **48**, 234 (2016).

10.1126/science.abm3882

WATER CAPTURE

An improved water-harvesting cycle

Tuning metal-organic frameworks at the atomic level helps release trapped water molecules

By **Lars Öhrström** *and* **Francoise M. Amombo Noa**

The importance of the phase transformations of water cannot be overstated. The steam engine was a leap in engineering as well as a major inspiration for thermodynamics. Cloud seeding using silver iodide is arguably one starting point of crystal engineering. On page 454 of this issue, Hanikel *et al.* (*1*) tackle the goal of harvesting water from desert air using water-catching channels inside porous metal-organic frameworks (MOFs). This technique may become important as fresh, potable water is turning scarce in many regions. The authors' design was guided by single-crystal x-ray diffraction studies of water assembly, molecule by molecule, inside a prototype MOF. This MOF was then tuned at the atomic level to achieve better water-harvesting properties.

Over 2 billion people live in water-stressed countries (*2*), and almost two-thirds of the global population experience water scarcity every year (*3*). Air contains a lot of water, even over deserts and semi-deserts. Plucking the water out of the air is relatively easy, but the great challenge is inventing a system that can also release water without a high energy cost.

Although zeolites, aluminosilicate minerals, can sorb water easily, liberating the water molecules may require temperatures of 250°C. Molecular-level interactions are the key to unlocking the water, with two areas of importance. One occurs between the water molecules and the channel surface of the material, and the other between the water molecules inside the channels.

Coordination bonds exist where the water binds to a metal ion with one of its free electron pairs. Hydrogen bonds between free electron pairs on nitrogen (N) or oxygen (O) atoms in the framework and water hydrogens. The reverse is also true with interactions between framework N-H and O-H groups and water oxygens. In addition, there are hydrogen bonds between the water molecules themselves.

To tune these with atomic-level control in zeolites, micro-, and mesoporous carbon and silica, for example, is difficult. MOFs allow this sort of control because of the ability to change and mix linkers and metal ions (*4*). As the coordination bonds are relatively strong, sorption is easy, but this also may hinder the release. These bonds are engineered by creating, or blocking, free sites on the metal ions. The hydrogen bonds to the framework are controlled by the linkers and to some extent also the metal-ion nodes.

The interactions between the water molecules define water and any polymorphs of ice, and the closer the sorbed water structure gets to the local environment in water or ice, the tighter they will be bound. These interactions are controlled by the framework topology and the linker size.

Hanikel *et al.* started with the observation of a notch in the water sorption data of MOF-303. Some water molecules were sitting rather tight even at low relative water pressures, which the authors call "step S", thus impeding the full water-harvesting potential. This prompted an atomic-level study of the sequential water filling using 23 single crystal structure determinations. The authors determined all water positions and hydrogen bond interactions from the fully loaded material with 31 weight percent water to the empty MOF (see the figure) and identified "step S" as two waters strongly hydrogen bound by N-H groups in two neighboring linkers. The authors then used the multivariate approach, mixing different, but size- and shape-congruent linkers, by introducing a linker with N-H replaced by C-H in 50% of the sites to eliminate "step S."

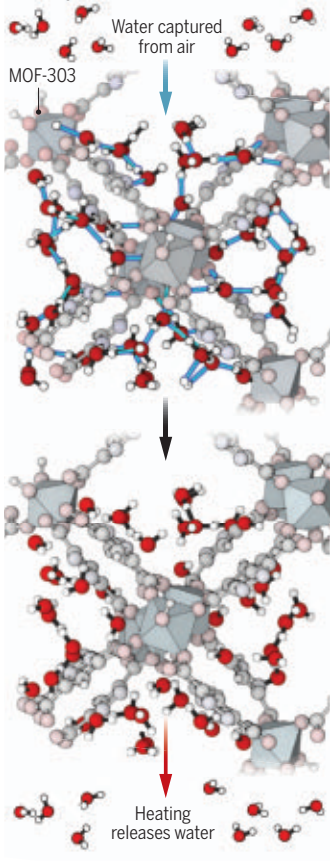
Not only the structures but also the energies are important, as revealed by quantum chemistry and thermodynamics. This method has implications for other devices that use water phase transformations. Air conditioning, for example, accounts for nearly 20% of the electricity used in buildings around the world (*5*). MOF-based water-adsorption–driven chillers could potentially improve the efficiency (*6, 7*).

Other applications are removal of trace amounts of water from natural gas or from organic solvents. Again, a lower heat of desorption makes the regeneration less energy demanding compared to commercial zeolites (*8*). The relative humidity for a healthy environment falls between 45 and 65%, and a different type of MOF may be designed for such uses (*9, 10*). Another idea is adsorption desalination to make potable water (*11*).

The detailed structures of the sorbed waters increase the possibilities of designing specific MOFs. However, knowing the structure alone is only a part of the puzzle. Determining the mass and heat transfer is also important for improving MOFs for applications, and these are less obvious properties to derive directly from the structure. ■

Catch and release

The molecular evolution of the water capture in metal-organic framework (MOF)-303 was determined on an atomic level using single-crystal diffraction. After molecular fine-tuning of the cavities was carried out, a material that more easily desorbs water was created.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

Department of Chemistry and Chemical Engineering, Chalmers University of Technology, Gothenburg, Sweden.
Email: ohrstrom@chalmers.se; mystere@chalmers.se

REFERENCES AND NOTES

1. N. Hanikel *et al.*, *Science* **374**, 454 (2021).
2. U. N. Water, Summary Progress Update 2021: SDG 6, United Nations (2021).
3. M. M. Mekonnen, A. Y. Hoekstra, *Sci. Adv.* **2**, e1500323 (2016).
4. L. Öhrström, F. M. Amombo Noa, *Metal-Organic Frameworks* (American Chemical Society, 2021).
5. IEA, "The future of cooling," Paris (2018); www.iea.org/reports/the-future-of-cooling.
6. S. Cui *et al.*, *Sci. Rep.* **8**, 15284 (2018).
7. K. H. Cho *et al.*, *Nat. Commun.* **11**, 5112 (2020).
8. X. Liu, X. Wang, F. Kapteijn, *Chem. Rev.* **120**, 8303 (2020).
9. R. G. AbdulHalim *et al.*, *J. Am. Chem. Soc.* **139**, 10715 (2017).
10. P. K. Tsobnang *et al.*, *Cryst. Growth Des.* **19**, 2869 (2019).
11. E. Elsayed, R. AL-Dadah, S. Mahmoud, P. Anderson, A. Elsayed, *Desalination* **475**, 114170 (2020).

10.1126/science.abm1854

VIEWPOINT: COVID-19

Tackling the pandemic with (biased) data

Data are crucial for understanding and addressing the pandemic, but there are pitfalls

By **Christina Pagel**[1] *and* **Christian A. Yates**[2]

Accurate and near real-time data about the trajectory of the COVID-19 pandemic have been crucial in informing mitigation policies. Because choosing the right mitigation policies relies on an accurate assessment of the current state of the local epidemic, the potential ramifications of misinterpreting data are serious. Each data source has inherent biases and pitfalls in interpretation. The more data sources that are interpreted in combination, the easier it is to detect genuine changes in an epidemic. Recently, in many countries, this has involved disentangling the varying impact of rising but heterogeneous vaccination rates, relaxation of mitigations, and the emergence of new variants such as Delta.

The exact data collected and their accuracy will vary by country. Typical data common to many countries are numbers of tests, confirmed cases, hospital and intensive care unit (ICU) admissions and occupancy, deaths, and vaccinations (*1*). Many countries additionally sequence a proportion of new positive tests to identify and track emerging variants. Some countries also now collect and publish data on infections, hospitalizations, and deaths by vaccination status (e.g., Israel and the UK). Stratifying all available data by different demographic factors (e.g., age, location, measures of deprivation, and ethnicity) is crucial for understanding patterns of spread, potential impact of policies, and efficacy of vaccines (age, timing of breakthrough infections, and prevalent variants).

It is also necessary to be aware of what data are not being collected. For example, persistent symptoms of COVID-19 (Long Covid) were recognized as a long-term adverse outcome by the autumn of 2020. However, no simple diagnostic test has been associated with the up to 200 different reported symptoms (*2*). Counting Long Covid relies on a clinical diagnosis, based on a history of having had COVID-19 and a failure to fully recover, with development of some characteristic symptoms and with no obvious alternative cause (*3*). These features make it very difficult to measure routinely, and so it rarely is. As a result, Long Covid is often neglected in decision-making. Failure to account for the disease load associated with Long Covid may lead to an unnecessary long-term societal health burden.

The feedback between different types of outcomes, different severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants, different mitigation policies (including vaccination), and individual risks (a combination of exposure and clinical risk) is complex and must be factored into both interpretation of data and the development of policy. Using all available data to quantify transmission is crucial to ensuring rapid and effective responses to early phases of renewed exponential growth and to evaluating mitigation measures. Relying too much on a single data source, or without disaggregating data, risks fundamentally misunderstanding the state of the epidemic.

The inherent biases and lags in data are particularly important to understand from the point of view of policy-makers. Because of the natural time scales of COVID-19 disease progression (see the figure), policy changes can take several weeks to show up in the data. Purely reactive policy-making is likely to be ineffective. When cases are rising, increases in hospital admissions and deaths will follow. When a new variant is outcompeting existing strains, it is likely to become dominant without action to suppress. The precautionary principle suggests acting early and emphatically. Conversely, when releasing restrictions, governments must wait long enough to assess them before continuing with re-opening.

The most up-to-date indicator of the state of the epidemic is typically the number of confirmed cases, as ascertained through testing of both symptomatic individuals and those tested frequently regardless of symptoms. Symptom-based testing is likely to pick up more adults and fewer younger individuals (*4*). Infections in children are harder to detect: children are more likely to be asymptomatic than adults, are harder to administer tests to (particularly young children), are often exposed to other viruses with similar symptoms, and can present with symptoms that are atypical in adults (e.g., abdominal pain or nausea). Children under 12 are not routinely offered the COVID-19 vaccination, and their mixing in schools provides ongoing opportunities for the virus to circulate, so it will be important for countries to track infections in children as accurately as possible. Other testing biases include accessibility, reporting lags, and the ability to act lawfully upon receiving a positive result. Substantial changes in the number of people seeking tests may further confound case figures (*5*). Case positivity rates may provide a more accurate reflection of the state of the epidemic (*6*) but are dependent on the mix of symptomatic and asymptomatic people being tested.

SARS-CoV-2 variants have been an important driver of local epidemics in 2021. The four main SARS-CoV-2 variants of concern, to date, are B.1.1.7 (Alpha), B.1.351 (Beta), P.1 (Gamma), and B.1.617.2 (Delta). Some have been more transmissible (Alpha), some have substantial resistance to previous infection or vaccines (Beta), and some have elements of both (Gamma and Delta) (*7*). Currently, the high transmissibility of Delta combined with some immune evasion has made it the world's dominant variant. Determining which variants pose a substantial threat is difficult and takes time, particularly when many variants cocirculate. This is especially true for situations in which a dominant variant is declining, and a new one growing. While the declining variant remains dominant, its decrease masks increases in the new variant because case numbers remain unchanged or fall overall. Only when a new variant becomes dominant does its growth become apparent in aggregated case data, by which time it is, by definition, too late to contain its spread. This dynamic has been observed across the world with Delta over the latter half of 2021.

With multiple variants circulating, there are, effectively, multiple epidemics occurring in parallel, and they must be tracked separately. This typically requires the availability of sequencing data, which is unfortunately limited in most countries. Sequencing takes time and so is typically a few weeks out of date. These lags, and the uncertainty in sampling, can lead to hesitancy in acting. The rapid path to dominance of the Delta variant in the UK highlights the need for action when a quickly growing variant represents a few percent (or less) of overall cases.

Hospital admissions or occupancy data do not suffer the same biases associated with testing behaviors and provide unequivocal evidence of widespread transmission, its geography, and demographics. However, hospital admissions lag infections more than reported cases do, rendering these data

[1]University College London, London, UK. [2]University of Bath, Bath, UK. Email: c.pagel@ucl.ac.uk; c.yates@bath.ac.uk

less useful for proactive decision-making. Hospital data are also biased toward older people, who are more likely to suffer severe COVID-19, and now, unvaccinated populations. ICU occupancy data show a younger age profile than admissions because younger patients have a better chance of benefitting from the invasive treatment procedures (*8*).

Deaths are the most lagged indicator, typically occurring 3 or more weeks after infection and with an additional lag in registration and reporting. Death data should never be used to inform real-time policy decisions. Instead, death figures can act as an eventual measure of the success of a country's epidemic strategy and implementation. The age distribution of those who eventually die from COVID-19 is different from other metrics of the epidemic—skewed furthest toward older age groups (*9*). Those with clinical risk factors (such as immunodeficiency, obesity, or existing lung conditions), high exposure (health care workers and low-income workers), and the unvaccinated are overrepresented in COVID-19 deaths.

In countries with high vaccination rates, vaccination has had a substantial impact—reducing COVID-19 cases, hospitalizations, and deaths. However, when looking at the raw numbers in highly vaccinated populations, it can be the case that more fully vaccinated people are dying of COVID-19 than unvaccinated. If these raw statistics are misinterpreted—or worse, deliberately misused—they can damage vaccine confidence. More vaccinated people may die than unvaccinated because such a high proportion of people are vaccinated (*10*). This does not mean vaccines are not effective at preventing death. Looking at the rates of death in vaccinated and unvaccinated individuals separately within age groups demonstrates that vaccines provide considerable protection against severe disease and death. This example illustrates how important it is to curate and manage the way in which data are presented.

Each country has established its own vaccination priority lists and dosing schedules to best achieve its goals (*11*, *12*). Each of these strategies will manifest differently in the data. Additionally, many countries are using multiple vaccines in tandem and administer them differently for different demographics. Some countries are vaccinating adolescents, and others are not or not offering them the full approved dose. Most vaccines require two doses, spaced between 3 and 12 weeks apart, except for the Johnson & Johnson single-dose vaccine. This matters, particularly as variants spread, because different vaccines have different effectiveness after one and two doses, different timelines to full effectiveness, and different effectiveness against variants (*13*).

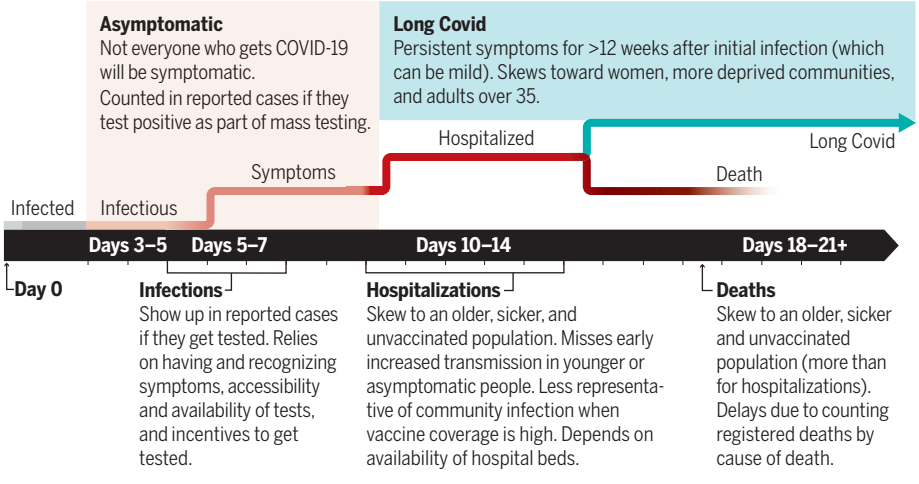
Data published on the vaccination delivery itself must thus go beyond the raw numbers of people vaccinated. Vaccine uptake must be reported by whether fully or partially (one-dose in a two-dose regimen) vaccinated and using the whole population as a denominator. It is vital to disaggregate vaccine data by age, gender, and ethnicity as well as location so that it is possible, for example, to understand the impact of deprivation on vaccine coverage or vaccine hesitancy in particular demographics. When interpreting vaccination data, it is important to remember that there is also a lag between delivery and the build-up of immunity.

Data on reinfection and post-vaccination (breakthrough) infection are also important to determine the relative benefits of infection-mediated and vaccine-mediated immunity and the length of protection offered. Studies that show those who were immunized earlier were acquiring COVID-19 with higher rates than those vaccinated more recently may suggest waning vaccine protection (*14*). Such studies have already prompted vaccine booster programs in some countries. However, any study that suggests waning immunity must be extremely careful to ensure that the "early" and "recent" subgroups are properly controlled. Differences in prior exposure, affluence, education level, age, and other demographic factors between these cohorts may be enough to explain the disparities in SARS-CoV-2 infection rates, even in the absence of waning immunity. Waning immunity must also be reported separately for different outcomes; for example, there might be waning in terms of preventing symptomatic infection but far less or none in preventing death (*15*). Additionally, there are ethical concerns about mass booster programs in high-income countries while many lower-income countries have been unable to procure vaccines.

Moving into the vaccination era, reported cases, hospitalizations, and deaths should also be disaggregated by vaccination status (and by which vaccine), which will be easier in countries where national linked datasets exist. Additionally, incorporating Long Covid into routine reporting and policy-making is crucial. Consistent diagnostic criteria and well-controlled studies will be vital to this effort. These elusive data will be of critical importance to navigate our way successfully out of the pandemic. ■

COVID-19 progression

An approximate timeline from infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) to various outcomes. When current infections show up in different data sources depends on this timeline. Collecting data for Long Covid, asymptomatic infection, and vaccine history will improve understanding of the pandemic.



GRAPHIC: N. CARY/*SCIENCE*

REFERENCES AND NOTES

1. M. Roser *et al.*, *Our World in Data* (2021); https://bit.ly/3kepLgw.
2. H. E. Davis *et al.*, *E. Clin. Med.* **38**, 101019 (2021).
3. M. Sivan, S. Taylor, *BMJ* **371**, m4938 (2020).
4. S. M. Moghadas *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 17513 (2020).
5. J. Wise, *BMJ* **370**, m3678 (2020).
6. D. Dowdy, G. D'Souza, COVID-19 Testing: Understanding the "Percent Positive" (2020); https://bit.ly/3CeN8wl.
7. C. E. Gómez *et al.*, *Vaccines* (Basel) **9**, 243 (2021).
8. A. B. Docherty *et al.*, *BMJ* **369**, 1985 (2020).
9. Office for National Statistics, Deaths registered weekly in England and Wales by age and sex: covid-19 (2021); https://bit.ly/3Ci2obS.
10. C. Yates, *BMJ* (2021); https://bit.ly/3kfqIoH.
11. Centers for Disease Control and Prevention, COVID-19 Vaccines for specific groups of people (2021); https://bit.ly/39aijwp.
12. Joint Committee on Vaccination and Immunisation, Priority groups for coronavirus (COVID-19) vaccination: advice from the JCVI, 30 December 2020 (2020); https://bit.ly/2VIhfwC.
13. J. P. Moore, *JAMA* **325**, 1251 (2021).
14. Y. Goldberg *et al.*, medRxiv 10.1101/2021.08.24.21262423 (2021).
15. Public Health England, Duration of protection of COVID-19 vaccines against clinical disease (2021); https://bit.ly/3CCoVAq.

ACKNOWLEDGMENTS

C.P. and C.A.Y. are both members of Independent SAGE: www.independentsage.org.

10.1126/science.abi6602

POLICY FORUM

CLIMATE CHANGE

How ambitious are oil and gas companies' climate goals?

Despite many claims, the sector is not on track

By **Simon Dietz[1,2], Dan Gardiner[1], Valentin Jahn[1], Jolien Noels[3]**

The oil and gas (O&G) industry faces an existential threat from the transition to a low-carbon economy. Companies are increasingly responding by setting greenhouse gas (GHG) emissions targets, which are presented as being compatible with this transition. Many stakeholders, including investors that own O&G companies, want to understand how ambitious these targets are. In this paper, we present a forward-looking method of estimating the life-cycle carbon emissions intensity of O&G producers based on their public disclosures, and we use it to compare companies' targets with international climate goals. The sector is not on track. Recent trends in emissions intensity have been mostly flat. Nearly half the companies we assess have yet to set emissions targets or provide sufficient clarity on them. Of those that have set targets, most are either too shallow or too narrow. Two companies have set targets that would bring their GHG intensity below international climate goals by mid-century.

In 2019, the O&G sector supplied 55% of global primary energy (*1*) and was responsible for 56% of all energy-related carbon dioxide (CO_2) emissions (40% of total GHG emissions) (*2*). To meet international climate goals, the world will need to transition away from burning O&G, and the O&G sector itself will need to curb its operational emissions.

One way to understand the positioning and performance of O&G companies on climate change is to use environmental, social, and governance (ESG) ratings. These ratings have a traditional focus on companies' managerial practices (e.g., whether companies have established a climate change policy). Yet, although measures of company inputs and effort are useful, it is GHG emissions that ultimately matter for climate change and companies' transition risk. Thus, many organizations rely on estimates of companies' current carbon footprint, in particular GHG intensity (GHG emissions normalized by a measure of activity). However, typical carbon footprinting methods have several limitations. First, they are often limited in scope to companies' operational emissions, whereas substantial emissions can occur either further up the value chain or, as in O&G, further down (e.g., when burned in automobile engines). Second, emissions are often normalized by company revenue, but revenue is both volatile and difficult to project into the long run, compared with measures of physical production. Third and most important, by being restricted to historic emissions, they provide limited information about companies' future emissions and therefore about companies' preparedness to undergo a low-carbon transition.

In this paper, we present a method of estimating forward-looking, life-cycle GHG intensities of public O&G producers and comparing them with scenarios that limit global warming to different levels, e.g., 1.5°C, with the aim of understanding how companies' climate targets compare with international goals, notably those of the 2015 UN Paris Agreement on climate change. We provide an independent means of evaluating the growing number of company claims to be either "net zero" or "Paris-aligned." We also analyze covariates of companies' GHG intensities, paving the way to assess the potential of strategies such as reducing operational emissions and shifting to gas.

Our method is based on GHG intensities so that we can control for the role of company size in determining emissions. This creates a challenge, however, as keeping temperatures to 2°C or below requires staying within an absolute CO_2 emissions budget (*3*). We follow the Sectoral Decarbonization Approach (SDA) in reconciling an intensity approach with an absolute emissions budget (*4*). An energy model is run within an absolute emissions budget to obtain projections of emissions and activity/production by sector. Dividing the former by the latter gives sectoral GHG intensity pathways, which are used as benchmarks against which to compare companies' own transition pathways.

Our measure of companies' GHG intensity includes not only their operational (Scope 1 and 2) emissions, but also emissions down the value chain from use of sold products (Scope 3, category 11)—i.e., burning O&G for energy in buildings, electricity, industry, and transport (the three Scopes are from the widely used GHG Protocol Corporate Accounting and Reporting Standard). Emissions are normalized by energy sales and compared with global primary energy supply, including coal, O&G, nuclear, renewable electricity, and biofuels. We present a method of sourcing emissions and energy sales conveniently yet consistently using companies' public disclosures.

We focus on the world's top 50 public O&G producers by free-float market capitalization. The focus on large, publicly listed companies is motivated by the importance of O&G companies' transition risks to investors, and the growing importance of investors as actors trying to positively drive a low-carbon transition. The companies belong to the Industry Classification Benchmark "Oil and Gas Producers" sector, including upstream exploration and production (E&P) companies, integrated O&G companies involved not only in E&P but also in downstream refining, distribution, and retail, as well as a few specialist downstream companies. Since 2017, the list of companies in the top 50 has been subject to some fluctuation. We include new entrants to the top 50 since 2017, as well as retaining companies that have fallen outside the top 50, resulting in a total of 52 companies included in our analysis. Our assessment includes company emissions disclosures up to September 2020 and emissions reduction targets announced by 1 January 2021. Details on all data and methods can be found in the supplementary materials.

THE STATE OF COMPANY EMISSIONS DISCLOSURE AND CURRENT EMISSIONS INTENSITY

The vast majority of public O&G producers (45 out of 52) now publish data on their Scope 1 and 2 emissions (table S1). However, only 23 companies disclose an estimate of their Scope 3 use of sold product emissions, and these estimates are based on different and often incomparable methods, because, in a largely voluntary disclosure regime, companies enjoy considerable latitude in choosing how to measure them (e.g., what organizational boundary to choose and which products to include).

As a result, we developed a bottom-up method of calculating Scope 3 use of sold product emissions on a consistent basis using companies' energy sales and production data. The method is similar to that employed in (*5*). Scope 3 use of sold product emis-

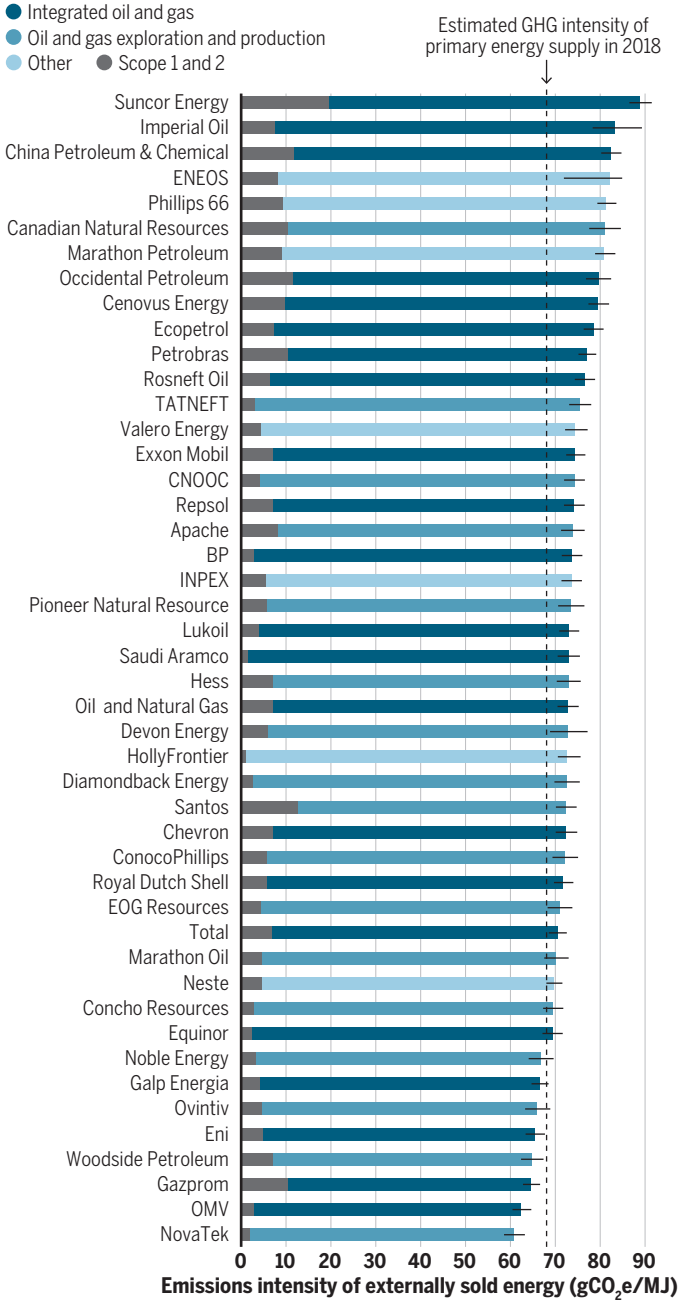
[1]Grantham Research Institute on Climate Change and the Environment, London School of Economics and Political Science, London, UK. [2]Department of Geography and Environment, London School of Economics and Political Science, London, UK. [3]Organisation for Economic Co-operation and Development (OECD), Paris, France. Email: s.dietz@lse.ac.uk

sions are estimated by applying product-specific emissions factors to externally sold volumes of a wide range of energy products. Similarly, the energy value of companies' sales is estimated by applying net calorific values to those sales. Electricity is converted to primary energy using the physical energy content method, applying average thermal power plant efficiencies. This method is applied to all companies, i.e., we do not use companies' own disclosed Scope 3 use of sold product emissions.

A plot of company GHG intensities in 2018 establishes companies' current position [2018 being the most recent year for which sufficient data are widely available (46 of 52 companies)] and compares that with an estimate of the GHG intensity of global primary energy supply (see the first figure). The average GHG intensity in 2018 was 73.3 grams of CO_2 equivalent (gCO_2e)/MJ on an unweighted basis (72.9 gCO_2e/MJ weighted by energy sales). This implies 424 $kgCO_2e$ per barrel of oil equivalent, assuming that an average barrel of oil contains 5779 MJ of energy. Disaggregating results by subsector, we estimate an overall emissions intensity of 71.1 gCO2e/MJ for O&G E&P (N =17), compared with 74.0 gCO2e/MJ for integrated O&G (N = 22). This accords with expectations, because integrated O&G companies have energy-intensive refining and distribution businesses to add to their E&P businesses, though the sample sizes are small and the difference is not statistically significant. We estimate an overall emissions intensity of 76.3 gCO2e/MJ for "Other O&G" producers, most of which are specialist refining and marketing companies (N = 7). Companies' GHG intensities are clearly dominated by Scope 3 use of sold product emissions. Scope 1 and 2 emissions account for only 8.6% of companies' overall intensity on average.

Emissions intensity of public oil and gas producers

Life-cycle intensities of the world's top public producers in 2018 are shown. All non–Scope 1 and 2 emissions are from Scope 3 use of sold products. Error bars represent uncertainty about Scope 3 emissions from use of sold products, using upper- and lower-bound effective carbon emissions factors. The dotted line is based on data from IEA/EDGAR. Bars are colored according to GICS subindustry.



EDGAR, Emissions Database for Global Atmospheric Research; GHG, greenhouse gas; GICS, Global Industry Classification Standard; IEA, International Energy Agency.

Most O&G producers' GHG intensities are above the global energy sector average, although eight are below. The global energy sector average includes coal, which raises the average, but more notably, it includes biofuels, nuclear energy, and renewables, which reduce the average. O&G companies currently sell very little of these forms of energy. Compared with data on GHG intensity at the field level (*6*), public O&G producers' operational GHG intensity shows less variation, possibly owing to diversification of production across regions.

Although emissions from fossil fuel combustion are generally well understood, we perform sensitivity analysis using upper- and lower-bound effective carbon emissions factors (*7*). The error bars in the first figure show the results of this analysis. Uncertainty stemming from Scope 3 use of sold product emissions is relatively small and is highly correlated across companies because it stems from uncertainty about carbon embodied in generic fuel types. It is largest for companies selling relatively large quantities of bitumen and/or coal. There is additional uncertainty about companies' Scope 1 and 2 emissions, which does not tend to be quantified in company disclosures.

COMPARING COMPANIES' TARGETS WITH THE PARIS AGREEMENT GOALS

Companies' future emissions intensity is a product of (i) their current emissions intensity and (ii) their emissions targets. As of 1 January 2021, 28 out of 52 companies (54%) had disclosed a quantitative emissions target and accompanying emissions/energy data sufficient to project their transition pathways into the future. These targets are disclosed on various bases, and we convert them to a common intensity metric. Some other companies have disclosed targets, but not in a form that can be independently assessed in terms of company-wide emissions (table S1). For example, these companies do not define the scope of emissions or organizational boundary to which the target applies, or targets are expressed relative to an unquantifiable baseline. Henceforth all emissions reductions are measured relative to the 2018 level.

On an unweighted average basis, companies are targeting reducing their GHG intensity by 16.6% by their end target year (which varies). On the same basis, the end target year for achieving this is 2037, i.e., most end targets are relatively long-term. The median target equates to just a 6.4% reduction in GHG intensity by the end target year, however. This shows that most targets are relatively unambitious. The unweighted average is heavily skewed by the ambitious targets of six companies: Occidental Petroleum (a 100%

GRAPHIC: K. FRANKLIN/*SCIENCE*

reduction in GHG intensity by 2050, i.e., net zero emissions by our measure), Royal Dutch Shell (–65% by 2050), Total (–58% by 2050), Eni (–55% by 2050), Repsol (–52% by 2050), and BP (–20% by 2050).

Companies' transition pathways are compared with three decarbonization scenarios for the energy sector (see the second figure), which are derived from the IPCC/IAMC 1.5°C Scenario Explorer: 2°C, below 2°C, and 1.5°C benchmarks. Only one company plans to reduce its GHG intensity below the 1.5°C benchmark by 2050: Occidental Petroleum. One further company plans to bring its GHG intensity under the 2°C benchmark by 2050: Royal Dutch Shell. No other company has set a target ambitious enough to beat the 2°C benchmark by 2050 (although Eni, Repsol, and Total come close), and most are far from that. Sixteen companies also have intermediate targets, including most of those with end targets after 2030.

Companies' targets fail to align with 2°C or below for one of two reasons: Either the targeted reduction is insufficient, or the targeted reduction does not apply to all company emissions. Particularly important is that most targets only cover Scope 1 and (mostly) Scope 2 emissions, and do not address the dominant emissions from Scope 3 use of sold products (compare table S1). Some of these targets limited to Scope 1 and 2 are described by the companies themselves as "net zero" targets. Put another way, these companies plan to decrease the GHG intensity of their operations, but not to change their core business model and diversify into low-carbon forms of energy. The companies with ambitious targets all include Scope 3 use of sold products in their targets.

Another way to look at these data is in terms of the compound annual growth/reduction rate (CAGR) of GHG intensity (fig. S1), which describes the average speed at which companies would need to reduce their GHG intensity to keep pace with the low-carbon transition. The CAGR required to bring the average company's GHG intensity down to the 2°C benchmark in 2050 is –3.0% (measured from the unweighted average company GHG intensity in 2018 of 73.3 gCO2e/MJ). The corresponding CAGR required to meet the below 2°C benchmark is –5.3%, and for 1.5°C it is –8.5%. These reduction rates are far in excess of what has been achieved recently. On an unweighted average basis, GHG intensities declined by just 0.1% per annum (p.a.) between 2014 and 2019 (notably, these data are for an unbalanced panel of companies, given varying degrees of historical emissions disclosure). European companies (excluding Russia) decreased their GHG intensity by –0.5% p.a. on average, whereas GHG intensities were approximately flat outside Europe (+0.02% p.a. on average). The benchmark CAGRs are also much faster than the CAGRs of the targets we assess. Excluding Occidental Petroleum, the unweighted average CAGR of company targets is just 0.7%. The absence of steep reductions in interim targets is also noteworthy. If companies do not meet the benchmarks until 2050, then unless their absolute energy sales are falling fast enough (e.g., because of falling market share), companies' cumulative absolute emissions between now and 2050 will be above the corresponding benchmark carbon budget.

FACTORS RELATED TO CURRENT GHG INTENSITY

Supplementary figures S2 and S3 analyze the relationship between companies' 2018 GHG intensity and several potential covariates that we were able to systematically collect.

Product mix

As companies' GHG intensities are dominated by Scope 3 use of sold product emissions, product mix matters. Life-cycle GHG intensity is strongly negatively correlated with the share of natural gas in companies' sales and production (Pearson's r=–0.76, p = 6.3E–10). Life-cycle GHG intensity is positively correlated with companies' GHG intensity of liquids (Pearson's r=0.36, p = 0.01), because liquids emit a wide range of CO_2 per unit of energy when burned (from 63.1 gCO_2/MJ for liquified petroleum gas to 80.7 gCO_2/MJ for bitumen). Companies' GHG intensity of liquids is therefore determined by the mix of liquids they sell. These results are consistent with the strategy of reducing emissions through shifting the product portfolio toward lower-carbon liquids and gas. However, the resulting emissions reductions would not be enough to meet climate goals in the long term.

Operational emissions

The second strongest association that we find is between life-cycle and operational (i.e., Scope 1 and 2) GHG intensity (Pearson's r = 0.63, p = 3.1E–06). This implies that O&G companies can reduce their life-cycle GHG intensity by reducing their operational GHG intensity, albeit the effect is more limited compared to the role of Scope 3. Operational GHG intensity does not appear to be related to product mix, at least as it affects emissions intensity. This is consistent with the analysis of the International Energy Agency, which suggests that the median operational GHG intensities of oil and gas are similar (*8*), although any correlation between operational GHG intensity and product mix may be masked by (i) aggregation to the company level and (ii) noise in the data.

Region and size

As shown in table S2, we find a significant association between companies' home country (i.e., the country in which the company is listed) and life-cycle GHG intensity using a Kruskal-Wallis rank sum test (p = 0.03). This could reflect the geographical nature of reserves (e.g., Canadian oil sands), markets, or regulation. It is consistent with evidence of considerable cross-country variation in the GHG intensity of oil production (*6*) and refining (*9*). We do not find any association between company size (as measured by both free float and non–free float market capitalization) or type and life-cycle GHG intensity.

FACTORS RELATED TO COMPANIES' GHG EMISSIONS TARGETS

Region and size

We do not find a significant association between whether a company has set a target (yes/no) and company region; however, target strength, measured in terms of the size of planned emissions reductions, is significantly associated with company region (Kruskal-Wallis rank sum test, p = 0.02). Most of the largest planned emissions reductions are by European companies (fig. S4). Whether a company has set a target (yes/no) is associated with company size as measured by free-float market cap (Kruskal-Wallis rank sum test, p = 0.01), but target strength is not significantly associated with size. Hence, bigger companies are more likely to set targets than smaller ones, but when it comes to target strength there is no significant difference.

Other factors

We do not find a significant association between target setting and the percentage of gas in the product mix, the emissions intensity of liquids sold, operational emissions intensity, Scope 3 use of sold product emissions intensity, or companies' life-cycle GHG intensity in 2018. One might have expected companies that are better positioned to transition to low-carbon energy by virtue of their business model and position in the sector to be more likely to set emissions targets, but our data do not show that.

LIMITATIONS AND FUTURE RESEARCH AGENDA

A number of caveats and limitations apply to our research.

First, although GHG intensity is a valid metric for a wide range of decarbonization strategies, decarbonization of energy sold is not the only path that O&G producers can take to cut emissions. For some companies, stopping investment in O&G production and returning cash to shareholders may be better. The effectiveness of this approach can only

be tracked using absolute emissions, but with absolute emissions one loses the capacity to compare companies' intensities. Even if a company does pursue a strategy to exit from energy, our approach can still help to separate out companies that plan to meaningfully reduce their emissions intensity from companies that do not, facilitating a better understanding of the company strategy. A two-part test may be appropriate, whereby companies can be aligned with climate goals either on the basis of their GHG intensity and decarbonization goals as set out in this paper, or their absolute GHG emissions and plans to wind down O&G production.

Second, we have been confronted with a series of decisions on the scope of our analysis. We aggregate O&G products because many companies that we assess produce both, and changing the product mix is a viable strategy for many. However, this can be contested given that O&G products serve rather different markets. We treat all oil producers as comparable, including upstream E&P companies and integrated companies. Again, this is contestable. However, our analysis indicates that there are only modest differences between E&P and integrated companies' GHG intensities (on various measures). We compare O&G producers to an emissions benchmark for all primary energy, including coal, nuclear, renewable electricity, and biofuels. This is challenging for O&G producers to achieve but reflects the fundamental shift in the primary energy mix in low-carbon scenarios. We include Scope 3 use of sold product emissions in our GHG intensity measure, because that is where the vast majority of life-cycle emissions come from.

Third, our method makes future projections based on companies' stated emissions targets. However, companies need to back up their targets with viable strategies, and thus it will be important to look at our data alongside complementary analyses, such as those of companies' governance and management practices (*10*) and their proved reserves and new capital expenditure (*11*). Particular attention needs to be paid to how much companies plan to rely on offsets, including nature-based solutions, and relatively unproven technology, such as carbon capture, utilization and storage, and direct air capture, to meet their targets.

Fourth, our method relies on the accuracy of company emissions disclosures. Although these disclosures are often independently audited, they have been questioned in various respects (*12, 13*). In O&G, a particular concern is possible underreporting of methane emissions (*14*). A systematic, like-for-like comparison of company disclosures with data from alternative sources such as asset-level data has yet to be undertaken and would be of value.

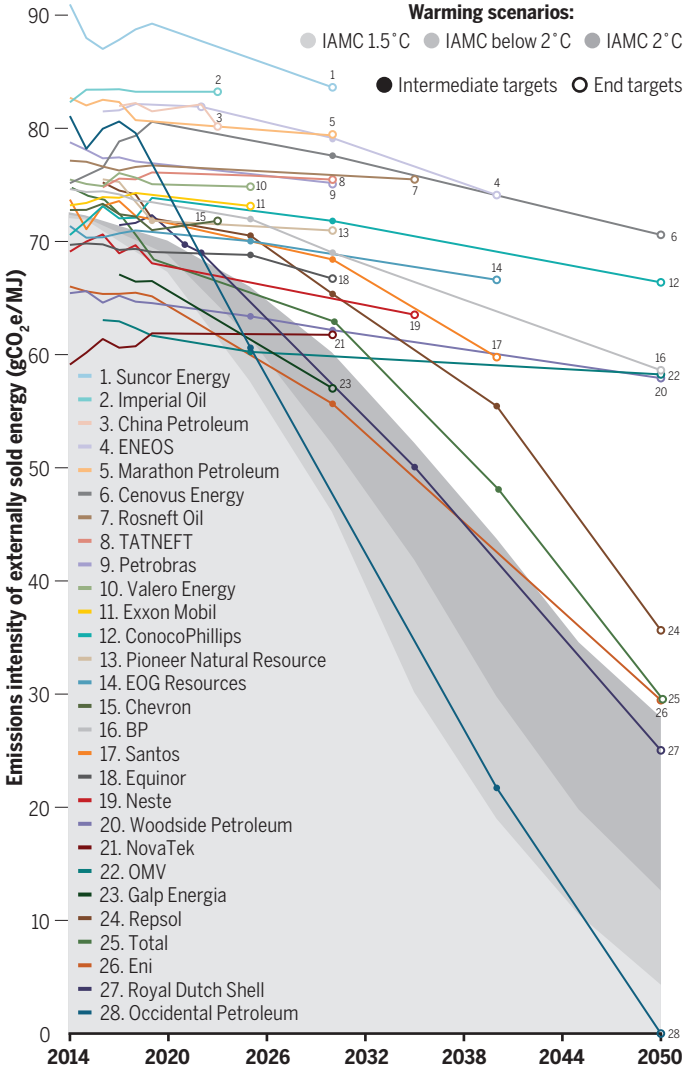
Fifth, disclosure frameworks and regimes currently allow companies to choose their organizational boundaries for emissions accounting (principally financial and operational control versus equity share). With a choice of consolidation approach and relatively small shifts in asset ownership, companies can influence both their emissions and emissions intensity.

FAR FROM BEING ALIGNED

Many companies are presenting their targets as compatible with international climate goals, either in a specific sense or in terms of a broad direction of travel, so that they can claim to still be good investments and more broadly still deserve a social license to operate. But as a group, the world's largest public O&G producers are far from being aligned with limiting global warming to 2°C or below. Some companies have yet to set emissions targets, and some others provide limited clarity on what they cover and how they will reduce company-wide emissions. Even for companies whose targets appear to be more ambitious, attention turns to the credibility of the strategies underpinning the targets. ■

Estimated transition pathways

Pathways are shown for 28 companies with emissions targets and sufficient emissions and energy data. Companies without targets are not shown. These are superimposed on three benchmark, temperature-constrained pathways for the greenhouse gas intensity of primary energy supply, using scenario data from the Intergovernmental Panel on Climate Change (IPCC)/Integrated Assessment Modeling Consortium (IAMC) 1.5°C Scenario Explorer.



GRAPHIC: K. FRANKLIN/*SCIENCE*

REFERENCES AND NOTES

1. International Energy Agency, "Global Energy Review 2019" (Paris, 2020).
2. International Energy Agency, "World Energy Outlook 2019" (Paris, 2019).
3. H. D. Matthews, N. P. Gillett, P. A. Stott, K. Zickfeld, *Nature* **459**, 829 (2009).
4. O. Krabbe *et al.*, *Nat. Clim. Chang.* **5**, 1057 (2015).
5. R. Heede, *Clim. Change* **122**, 229 (2014).
6. M. S. Masnadi *et al.*, *Science* **361**, 851 (2018).
7. IPCC, Working Group I Contribution to the IPCC Fifth Assessment Report, Climate Change 2013: The Physical Science Basis (Cambridge Univ. Press, 2013).
8. IEA, "The Oil and Gas Industry in Energy Transitions: Insights from IEA Analysis" (Paris, 582 2020).
9. L. Jing *et al.*, *Nat. Clim. Chang.* **10**, 526 (2020).
10. S. Dietz *et al.*, *Nat. Clim. Chang.* **8**, 1072 (2018).
11. Carbon Tracker Initiative, PRI, "2 Degrees of Separation: Transition Risk for Oil and Gas in a Low Carbon World" (2017).
12. E. H. Kim, T. P. Lyon, *J. Environ. Econ. Manage.* **61**, 311 (2011).
13. F. Depoers, T. Jeanjean, T. Jérôme, *J. Bus. Ethics* **134**, 445 (2016).
14. R. A. Alvarez *et al.*, *Science* **361**, 186 (2018).

ACKNOWLEDGMENTS

J.L. contributed to this work while working at the Grantham Research Institute on Climate Change and the Environment and prior to joining the OECD. This article should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed are those of the author.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abh0687

10.1126/science.abh0687

Ninth International Congress on Peer Review and Scientific Publication

Sept 8-10, 2022 | Swissôtel | Chicago, IL

Congress Advisory Board

Congress Directors
John P.A. Ioannidis, Meta-Research Innovation Center at Stanford (METRICS)

Michael Berkwits, *JAMA* and the JAMA Network

Congress Executive Director
Annette Flanagin, *JAMA* and the JAMA Network

European Director
Fiona Godlee, *BMJ*

European Coordinator
Theodora Bloom, *BMJ*

Associate Directors
Steve Goodman, Meta-Research Innovation Center at Stanford (METRICS)

Véronique Kiermer, *PLOS*

David Moher, Ottawa Hospital Research Institute

Director Emeritus
Drummond Rennie, University of California, San Francisco

Vivienne Bachelet, *Medwave,* Universidad de Santiago de Chile

Albert-László Barabási, Northeastern University

Howard Bauchner, Boston University School of Medicine

Lisa A. Bero, The University of Sydney

Patrick M. Bossuyt, University of Amsterdam

Lex Bouter, Vrije University

Isabelle Boutron, Université de Paris

An-Wen Chan, University of Toronto

Kay Dickersin, Johns Hopkins Bloomberg School of Public Health

Dan Evanko, American Association for Cancer Research

James Evans, The University of Chicago

R. Brian Haynes, McMaster University Faculty of Health Sciences

James Kigera, *Annals of African Surgery*

Sabine Kleinert, *The Lancet*

Christine Laine, *Annals of Internal Medicine*

José Florencio F. Lapeña Jr, University of the Philippines

Malcolm MacLeod, The University of Edinburgh

Emilie Marcus, University of California, Los Angeles

Ana Marušić, *Journal of Global Health*

Bahar Memani, Elsevier

Frank Miedema, University Medical Center Utrecht

Jigisha Patel, London, UK

Eric J. Rubin, *New England Journal of Medicine*

David Schriger, University of California, Los Angeles

Emily Sena, *BMJ Open Science*

Magdalena Skipper, *Nature* and Nature Research

Victoria Stodden, University of Illinois

Deborah Sweet, Cell Press

Sarah Tegen, American Chemical Society

Martin R. Tramèr, Geneva University Hospitals

Valda Vinson, *Science*

Erik Von Elm, Cochrane Switzerland

Eric-Jan Wagenmakers, University of Amsterdam

Jasmine Wallace, American Society for Microbiology

Rebecca Williams, National Library of Medicine

CALL FOR ABSTRACTS

Abstracts are due January 31, 2022

***JAMA, BMJ*, and METRICS invite abstracts for the Ninth International Congress on Peer Review and Scientific Publication.**

The Congress will provide a forum for the presentation and discussion of new research into the quality and credibility of peer review and scientific publication, to establish the evidence base on which scientists can improve the conduct, reporting, and dissemination of scientific research.

Abstracts reporting new research on any aspect of editorial and funding peer review, scientific publication, research and reporting practices, and information exchange will be considered.

Topics of Interest, include

- Bias
- Editorial and Peer Review Decision Making
- Research and Publication Ethics
- Improving the Quality of Reporting
- Models for Peer Review and Scientific Publication
- Dissemination of Scientific and Scholarly Information

More details, examples of Topics of Interest, Instructions for Preparing and Submitting Abstracts, and abstracts of previously presented research are available at **peerreviewcongress.org**

Abstracts may be submitted December 1, 2021 through January 31, 2022

Questions? jama-peer@jamanetwork.org

JAMA Network™ thebmj METRICS META-RESEARCH INNOVATION CENTER AT STANFORD

peerreviewcongress.org

HEALTH AND MEDICINE

The myth of healthy smoking

Two books examine timeworn tactics employed by the tobacco industry to attract new customers

By **Mark Parascandola**

A defining characteristic of the tobacco industry over the past century has been its ability to adapt to a changing world. When the link between cigarette smoking and lung cancer first gained public attention in the 1950s, for example, cigarette manufacturers responded by putting filters on their cigarettes. Although there was no evidence that filters provided any protection to the smoker, vigorous marketing from the tobacco industry drove smokers to adopt them.

Historian Keith Wailoo recalls when, in 1976, a 14-by-48-foot cigarette billboard went up at Woodhaven and Queens Boulevards near his family's LeFrak City apartment in Queens, New York. The billboard joined other colorful advertisements for tobacco products that had begun blanketing buses, subways, and city streets in major urban centers. The tobacco industry had recognized an opportunity to increase its customer base in predominantly Black, inner-city neighborhoods, and as Wailoo explains in *Pushing Cool*, the menthol cigarette was at the center of this push. It was not an accident or innate preference, Wailoo argues, that led Black smokers to take up menthol cigarettes but rather a highly organized, data-driven targeted marketing campaign.

Menthol cigarettes were first promoted to soothe the airways of "health conscious" smokers. Long used as an analgesic, menthol evokes a cooling sensation that masks the harshness of tobacco smoke. In the competition to capitalize on the growing menthol market, the industry's marketing experts "carved up, segmented, and fractionated" the population, exploiting psychology and social attitudes to shape product preferences.

Wailoo draws on collections of internal tobacco industry documents made public through litigation, along with other historical sources, to peer into the hidden world of tobacco marketing. "The success of menthol was not about the pull," he notes, "but about the shrewdness of the push."

In the 1960s, analysts advised tobacco companies that a growing African American market was "open to exploitation." When the Brown & Williamson Tobacco Corporation sought to revive its flagging Kool menthol brand in 1964, it targeted Black consumers through a national media campaign that featured images of an attractive Black couple sitting by a fountain surrounded by lush greenery with the tagline "feel extra coolness in your throat." Kool soon became the leading menthol brand.

Ads featuring Black smokers were initially viewed favorably as affirmations of Black influence.

Wailoo describes how the tobacco industry exploited broader social trends and attitudes about racial identity in its marketing efforts. The demonstrations and economic boycotts that defined the civil rights movement in the United States brought issues of race to the forefront, he notes, and Black consumers made up an increasingly influential market that companies were eager to court. By focusing on urban centers, cigarette companies were able to exploit a national demographic shift in which Black Americans became more and more concentrated in the inner cities. But this strategy required advertisers to tread a narrow line, as they remained wary, amid ongoing racial tensions, of alienating white consumers.

At the same time, the tobacco industry won support from some surprising quarters. *Ebony* magazine publisher John H. Johnson welcomed tobacco advertising featuring Black models as an affirmation of Black consumer influence. When tobacco control advocates campaigned against cigarette billboards, the National Association for the Advancement of Colored People came out in defense of commercial free speech, arguing that "it's racist to say black people are so malleable that they don't have the sense to decide whether to smoke or not." And when there was talk about banning menthol cigarettes, some influential Black leaders, such as the Reverend Al Sharpton, opposed a ban, arguing that it would result in an underground market and increased police action against Black smokers. Menthol smoking flourished, argues Wailoo, "precisely because marketers discovered how to exploit the vulnerabilities of Black life."

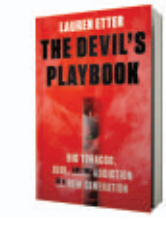
The reviewer is at the Center for Global Health, National Cancer Institute, Bethesda, MD 20892, USA. Email: paramark@mail.nih.gov

PHOTO: CHARLES 'TEENIE' HARRIS/CARNEGIE MUSEUM OF ART/GETTY IMAGES

The 2009 Family Smoking Prevention and Tobacco Control Act, which exempted menthol from the list of banned cigarette flavors, also left the door open, temporarily, for a new generation of electronic cigarettes. By 2014, there were nearly 500 e-cigarette brands and 8000 flavors being sold online, but in this crowded field, one product stood out. With its sleek design, which more closely resembled a USB stick than a cigarette, JUUL looked like a product of the future. In *The Devil's Playbook*, journalist Lauren Etter offers a front-row seat to the egos and forces behind the rise, and ultimate implosion, of JUUL. The book is based on extensive research and more than 100 interviews, including ones with anonymous industry insiders.

JUUL founders James Monsees and Adam Bowen first conceived of the device as part of their thesis project in the elite Stanford University product design program. The product they envisioned would allow the user to inhale nicotine without the smoke. Monsees and Bowen delved into industry documents to learn about cigarette manufacturers' research into nicotine delivery and early prototypes of low-emission cigarettes. But they positioned themselves as the anti-cigarette company.

Pushing Cool
Keith Wailoo
University of Chicago Press, 2021. 392 pp.



The Devil's Playbook
Lauren Etter
Crown, 2021. 496 pp.

After graduation, they pitched their idea to Silicon Valley investors, lambasting the cigarette as a relic of a bygone era. With financial backing, the company scaled up rapidly, adapting on the fly as it dealt with leaking pods and supply chain failures.

One of the features that distinguished JUUL from its competitors was its ability to deliver a far bigger nicotine hit. The nicotine content of one pod was roughly equivalent to that of a pack of Marlboros, and some teenagers were going through multiple pods a day. In 2017, JUUL went from 5% market share to almost 30% and growing. Suddenly, the device was everywhere, from Times Square billboards to school backpacks. Then, in August 2018, early data from the National Youth Tobacco Survey revealed an alarming 80% increase in e-cigarette use among high school students compared with the previous year. Teenage vaping was now officially an "epidemic," according to the US Food and Drug Administration (FDA).

As regulators closed in, JUUL Labs began to look and behave increasingly like a cigarette company, offering its own youth vaping prevention curriculum to schools, just as Philip Morris had done for youth cigarette smoking, and fighting local regulations on vaping products. In December 2018, JUUL employees learned that the company had reached a $12.8 billion deal that would give Altria, Philip Morris's parent company, a 35% ownership stake. K. C. Crosthwaite, a former Altria executive, took over as CEO of JUUL Labs and pledged to cooperate with policymakers and earn the public's trust—a tactic that, to critics, had the appearance of just another tobacco industry trick. Etter draws a parallel to the call, 20 years earlier, from Philip Morris executive Steven Parrish for the company to be "on the side of the angels."

In Etter's telling, Altria, unexpectedly, comes across as the adult in the room. As the company courted JUUL, Altria executives foresaw the disastrous consequences of marketing that appealed to teenagers, even though they were poorly positioned to do anything about it. "There's always been an art to running a tobacco company, a company that sells a product almost guaranteed to kill its customer," Etter writes. This subtlety was lost on the Silicon Valley set, who were accustomed to asking for forgiveness rather than permission. As JUUL and its investors rushed to scale up, marketing a highly addictive product like a fashion accessory, "they ignored that there might be consequences."

In 2019, vaping teenagers began showing up in emergency rooms with alarming respiratory symptoms. Laboratory data pointed to vitamin E acetate, an ingredient used in some tetrahydrocannabinol (THC)–laced vaping products, as the likely culprit. However, the experience added to growing concerns about youth e-cigarette use. Later that year, JUUL removed most of its flavored pods from store shelves. The company's valuation has since shrunk to less than $5 billion. But it is too soon to count the company out. On 12 October 2021, the FDA cautiously authorized the sale of an e-cigarette called Vuse, a move that could provide a path forward for other e-cigarette manufacturers.

Together, *Pushing Cool* and *The Devil's Playbook* demonstrate how tobacco companies have long designed and reimagined their products to attract new customers. The parallels in strategy and rhetoric employed by traditional tobacco companies and today's tech-infused disrupters are undeniable. In a dynamic world characterized by changing technology, shifting demographics, the growth of social media, and even the COVID-19 pandemic, the world of tobacco marketing has remained exceptionally, and disappointingly, constant. ■

10.1126/science.abl7101

ORNITHOLOGY

Contextualizing avian origins and evolution

Recent genomic advances have transformed our understanding of birds

By **Alan Feduccia**

The eminent 20th-century ornithologist Erwin Stresemann lamented in 1959 that despite 200 years of exhaustive study, all efforts to group birds into meaningful relationships had failed. But in 2014, after decades of phylogenetic speculation, whole-genome phylogenomics abruptly transformed the field (*1*). At long last, a believable avian family tree was born.

In his new book, *How Birds Evolve*, evolutionary biologist Douglas Futuyma, best known for his engaging text *Evolution* (*2*), skillfully weaves together salient themes of evolutionary biology from the living aviary, stressing the important role that birds played in advancing our understanding of evolution and taking advantage of the new systematics to provide a foundation for evolutionary explanations. Although casual readers might find the text somewhat advanced and laborious, the chapters are composed in well-written conversational prose, with expositions on multifarious evolutionary phenomena that are infused with scientific explanations.

Futuyma's discussion begins with a section that explains Charles Darwin's transformative ideas—from natural selection and fitness to brood parasitism to gene flow and genetic drift—and thematic chapters elaborate on the relationship between these ideas and bird lineages. The book describes complex evolutionary issues in understandable terms, but Futuyma's speculation about bird evolution surrounding the period of the Cretaceous–Paleogene (K–Pg) extinctions is based on indeterminately identified single fossils (e.g., *Vegavis* and *Asteriornis*), as well as conjecture on biogeography, which degrades into narrative "just-so stories."

The reviewer is S. K. Heninger Distinguished Professor Emeritus and former chair, Department of Biology, University of North Carolina, Chapel Hill, NC 27599, USA. Email: feduccia@bio.unc.edu

The plumage of the male ruff (*Calidris pugnax*) varies, depending on whether or not the bird defends a territory.

How Birds Evolve: What Science Reveals About Their Origin, Lives, and Diversity
Douglas J. Futuyma
Princeton University Press, 2021. 320 pp.

Among the intriguing topics that Futuyma discusses are the historically problematic ratites, considered cladistically an ancient flightless group of birds distributed by drifting continents. Early methodologies proved incapable of dealing with history, convergence, and heterochrony, but modern DNA analyses show that diverse ratite populations evolved flightlessness independently, their ancestors having flown to disparate homes after the K–Pg extinctions.

As biologist John Wiens has noted, "ontogeny discombobulates phylogeny" (*3*). Arrested development, which embryologist Gavin de Beer referred to as "Peter Pan evolution," can rapidly produce a race of giants with miniscule genetic change in developmental timing, leaving no signal of ancestry. Paleontologist Philip Gingerich articulated this problem in 1988, musing whether we expect too much from animals when we ask them to reveal their morphology as well as their genealogy (*4*).

Futuyma rightly notes that "snakes have [not] re-evolved legs, and no birds have re-evolved teeth." Flightless birds, likewise, have never, to our knowledge, re-evolved flight. But this begs the question: How did theropod dinosaurs, with their already foreshortened forelimbs and stubby hands, re-evolve highly extended wings with prolongated fingers?

Futuyma stresses the salient theropod features—obligate bipedality and an open hip joint—promoting the orthodoxy that "birds are living dinosaurs." In doing so, he circumvents the fact that this scheme launches a cascade of biological improbabilities, including a contentious, distinctive, and mysterious avian digital frameshift with no apparent selective advantage. Meanwhile, mid-Jurassic scansoriopterygids, the climbing and gliding maniraptoran dinosaurs most proximal to bird origins, are devoid of salient theropod characteristics, with a nearly closed acetabulum (*5*).

Futuyma advances Richard Prum's feather evolution model, which states that feathers evolved from ancestral filamentous protofeathers. This is problematic because the original "feathered dinosaur," *Sinosauropteryx*, had a mane of fibers beneath the skin that other investigators have argued consists of skin collagen fibers supporting a mid-dorsal frill (*6*). Conceiving of dinosaur filaments as feather precursors is flawed because it overvalues superficial similarities in appearance, makes little distinction between evolution and regeneration, and is ultimately reminiscent of Ernst Haeckel's long-disputed biogenetic law, which maintains that each embryo's developmental stage represents an adult form of an evolutionary ancestor. The corollary tenet of the new model "feathers before flight," which posits hypersophisticated aerodynamic flight feathers evolving in a nonflight context, likewise stretches biological credulity.

Nonetheless, I recommend this book because it provides a good account of many recent scientific advances and explains the consensus view of avian origins, albeit while evading the possibility that we do not fully understand them. Controversy, not consensus, is grist to the mill of good science, and the recent genomic renaissance has provided Futuyma with the needed tools to showcase the bright future for the study of birds, their origins, and their evolution. ■

PHOTO: WINFRIED WISNIEWSKI/MINDEN PICTURES

REFERENCES AND NOTES

1. E. D. Jarvis *et al.*, *Science* **346**, 1320 (2014).
2. D. J. Futuyma, M. Kirkpatrick, *Evolution* (Oxford Univ. Press, ed. 4, 2017).
3. J. J. Wiens, R. M. Bonett, P. T. Chippindale, *Syst. Biol.* **54**, 91 (2005).
4. P. D. Gingerich, *Nature* **336**, 628 (1988).
5. S. A. Czerkas, A. Feduccia, *J. Ornithol.* **155**, 841 (2014).
6. T. Lingham-Soliar, *The Vertebrate Integument, Volume 2* (Springer, 2015).

10.1126/science.abk1689

LETTERS

A girl takes a test to diagnose asthma. The design of a recent study exploring a possible asthma treatment in children has been the subject of debate.

PHOTO: IAN HOOTON/SCIENCE SOURCE

Edited by **Jennifer Sills**

Accurately assessing children's asthma study

In his News Feature "Failure to protect?" (13 August, p. 729), C. Piller misrepresents the Vitamin D Kids Asthma Study ("Vit-D-Kids" or "VDKA") (*1*). He reports concerns about the study's design, participant safety and selection, consent forms, and report transparency. These doubts are unfounded. VDKA ethically investigated a potentially important treatment for childhood asthma.

Piller reports pulmonary specialist Bruce Davidson's objections to using a placebo instead of a low-dose vitamin D control. However, as Piller acknowledges, screening for low vitamin D levels is not the standard of care in the United States, where most children with vitamin D levels lower than 20 ng/ml (including those with asthma) are untreated. Most remain asymptomatic (*2*).

Placebo-controlled trials such as VDKA are ethical when risks are low, and a control dose can cause falsely negative results (i.e., if both low and high doses affect the outcome, an effect can be obscured) (*3*). Piller failed to mention dietary counseling, which, along with other safety measures, made the risks of severe vitamin D deficiency or rickets (a condition associated with severe vitamin D deficiency) in VDKA participants minimal and no greater than that of millions of US children (*1*). Davidson's statement that children with vitamin D levels of 10 ng/ml remained in VDKA for 48 weeks is false. Levels were tested every 16 weeks, and only four participants had vitamin D levels <10 ng/ml at one visit (n = 2) or 10 to 13 ng/ml at two consecutive visits (n = 2); all were referred to endocrinologists (*1*).

Piller also expresses concern about overexposure to vitamin D, citing National Institutes of Health guidelines indicating that levels of vitamin D higher than 50 ng/ml are "potentially hazardous." However, vitamin D toxicity is defined as more than 150 ng/ml, and 100 ng/ml is the upper limit of normal vitamin D levels (*4*). Piller failed to mention any safety measures or the fact that vitamin D toxicity did not occur in any VDKA participant (*1*).

Piller alleges that the risk of bone fractures was not properly declared. However, an international consensus concluded that children with vitamin D deficiency without confirmed rickets are not at increased risk of fractures (*2*), and all participants were on low-dose inhaled corticosteroids, which have not been shown to cause fractures in children (*5*–*7*). Piller incorrectly notes that 14 participants in the placebo group took systemic steroids at least twice during the trial (*1*); the correct number is 8. There is no convincing evidence that one or two short courses of systemic steroids cause fractures.

Davidson misleadingly accuses the study authors of "covering up" fractures. Fractures were not listed as serious adverse events in the protocol and thus not reported as such (*1*). All fractures were reported to the data and safety monitoring board (DSMB). Davidson further alleges that the study could have been harmful, but the fracture number was similar in children receiving 4000 international units (IU)/day of vitamin D (five cases) and those in the placebo group (four cases). All fractures occurred in children with baseline vitamin D levels ≥20 ng/ml while playing contact sports or following accidents. Therefore, it is unreasonable to conclude that nonparticipation in VDKA or a control vitamin D dose would have prevented fractures. Fracture rates vary by age, sex, and ascertainment method. The overall VDKA fracture rate [9/192 (4.7%)] is not significantly different from plausible rates (*8*, *9*). This is true even when comparing a rate of 9/192 to one ~50% lower (5/192), as P = 0.42, indicating a 42% probability that such difference is due to chance alone.

Piller asserts that Black participants were overrepresented. Of VDKA children, 52% were Black and 31% were white, which adequately reflects the higher prevalence of both asthma (*10*) and vitamin D deficiency (*11*) in Black children relative to white children and the demographics of the study sites. Piller then quotes ethicist Harriet Washington, who alleges that

we aimed to attribute effects to genetics before social explanations, fitting "a racist pattern." We collected data on stress and socioeconomic status (*1*, *12*) and never suggested that we would prioritize genetics over socioeconomics.

Piller also quotes medical ethicist Ruth Macklin, who calls the institutional review board (IRB) decision to approve the study "surprising—if not appalling." Yet changes made to the consent form and participant eligibility cutoffs show that the investigators, the DSMB, and IRBs responded to safety concerns. For example, Piller acknowledges that the vitamin D level required for VDKA inclusion was amended to 14 ng/ml in 2018 (the change was approved in 2017) (*1*). Piller states that VDKA wasn't a success. By showing that vitamin D does not prevent severe asthma attacks, VDKA helped avoid unnecessary testing and treatment for millions of children. The VDKA trial was ethical and successful.

Juan C. Celedón
Department of Pediatrics and Departments of Medicine, Epidemiology, and Human Genetics, University of Pittsburgh, Pittsburgh, PA 15261, USA, and Division of Pulmonary Medicine, University of Pittsburgh Medical Center Children's Hospital of Pittsburgh, Pittsburgh, PA 15224, USA.
Email: juan.celedon@chp.edu

REFERENCES AND NOTES

1. E. Forno *et al.*, *JAMA* **324**, 752 (2020).
2. R. Munns *et al.*, *J. Clin. Endocrinol. Metab.* **101**, 394 (2016).
3. L. A. Frame, J. P. Fischer, G. Geller, L. J. Cheskin, *Nutrients* **10**, pii347 (2018).
4. M. F. Holick, *Mayo Clin. Proc.* **90**, 561 (2015).
5. R. G. Schlienger *et al.*, *Pediatrics* **114**, 469 (2004).
6. Y. Loke *et al.*, *BMJ Open* **5**, e008554 (2015).
7. C. V. Chalitsios, D. Shaw, T. McKeever, *Thorax* **76**, 21 (2021).
8. E. Degabriele *et al.*, *J. Pediatr. Child Health* **54**, 855 (2018).
9. R. A. Lyons *et al.*, *Inj. Prev.* **5**, 129 (1999).
10. US Centers for Disease Control and Prevention, Asthma, Child Prevalence (2018); www.cdc.gov/asthma/asthmadata/Child_Prevalence_Race.html.
11. L. E. Au, S. S. Harris, J. T. Dwyer, P. F. Jacques, J. M. Sacheck, *J. Pediatr. Endocrinol. Metab.* **27**, 1095 (2014).
12. F. Rosser *et al.*, *J. Allerg. Clin. Immunol.* 10.1016/j.jaci.2021.05.037 (2021).

COMPETING INTERESTS

J.C.C. received research materials from Pharmavite (vitamin D and placebo capsules) and GSK (inhaled steroids) to provide medications free of cost to participants in the Vitamin D Kids Asthma Study (VDKA). He has also received research materials from Merck to provide inhaled steroids to participants in an NIH-funded study unrelated to VDKA.

10.1126/science.abm1147

Misleading portrayal of children's asthma study

In his News Feature "Failure to protect?" (13 August, p. 729), C. Piller creates false impressions about the Vitamin D Kids Asthma Study ("Vit-D-Kids" or "VDKA") (*1*) and its investigators. The story questions the study's design, participant distribution, and reporting. The allegations in the story distort the study, which was carefully and ethically designed and appropriately diverse and reported.

Piller minimizes the rationale used to select the placebo-controlled trial design and suggests that there is agreement that such a design is unethical. The choice of a placebo control and protection of subjects are extensively discussed in the trial protocol (*2*), and ethicists have argued that such designs are ethical in appropriate circumstances, such as allowing participants access to vitamin D and supplementing participants if vitamin D levels dropped sufficiently low (*3*). Such conditions are met in this trial.

The News story notes that the majority of children in the trial were Black and states that this constitutes overrepresentation. Both asthma and low vitamin D levels were required to enroll in the trial, and Black children have rates of asthma 1.9 times as high as white children (*4*) and rates of low vitamin D levels 3 times as high as white children (*5*) [largely due to skin pigments blocking sun-induced vitamin D synthesis (*6*)]. Based on my simulations (*7*) using these factors and census-based demographics of the trial site locations, the study population (52% Black, 31% white) mirrors the affected population (46 to 68% Black, 18 to 30% white). Rather than being criticized, this trial should be commended for inclusion of appropriate trial participants (*8*).

Piller writes that participants were at increased risk of fractures and that the nine bone fractures experienced by study participants were more than anticipated, without specifying the magnitude of any increased risk or the anticipated number of fractures. However, there is no consensus that any increased risk exists, and the number of fractures does not differ significantly from the number expected (*7*). Moreover, more fractures occurred in the treatment than in the placebo group, as Piller eventually notes, providing no evidence of harm due to the use of the placebo control.

The News story refers to Vit-D-Kids as a "failed" trial with no advantage to society. A trial with clear data that refute a hypothesis is not a failure and provides considerable advantages by freeing affected individuals from trying unhelpful treatments and researchers to pursue other avenues. Deliberation about the many issues around clinical trials is important, but by portraying the Vit-D-Kids study as unethical and likely harmful, *Science* has not contributed positively to these discussions.

Jeremy M. Berg
Department of Computational and Systems Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA.
Email: jberg@pitt.edu

REFERENCES AND NOTES

1. E. Forno *et al.*, *JAMA* **324**, 752 (2020).
2. J. Celedón, "Vitamin D to prevent severe asthma exacerbations (Vit-D-Kids Asthma)" (2021); https://clinicaltrials.gov/ct2/show/NCT02687815.
3. L. A. Frame, J. P. Fischer, G. Geller, L. J. Cheskin, *Nutrients* **10**, 347 (2018).
4. US Centers for Disease Control and Prevention, Asthma, Child Prevalence (2018); www.cdc.gov/asthma/asthmadata/Child_Prevalence_Race.html.
5. US Centers for Disease Control and Prevention, NCHS Data Brief No. 59 (2011); www.cdc.gov/nchs/data/databriefs/db59.pdf.
6. L. Y. Matsuoka, J. Wortsman, J. G. Haddad, P. Kolm, B. W. Hollis, *Arch. Dermatol.* **127**, 536 (1991).
7. J. M. Berg, "The importance of careful use of comparative terms in quantitative arguments in science journalism" (2021); https://medium.com/@jeremymberg/the-importance-of-careful-use-of-comparative-terms-in-quantitative-arguments-in-science-journalism-4c95e75c7729.
8. S. E. Geller *et al.*, *Acad. Med.* **10**, 630 (2018).

10.1126/science.abm4776

Editor's note

The article by Charles Piller raises important questions about clinical trials that compare experimental treatments against a placebo instead of standard care by physicians—a topic of active debate in the scientific community. Letter authors Celedón and Berg criticize concerns expressed by Bruce Davidson or described by Piller about the ethics and design of the Vitamin D Kids Asthma Study ("Vit-D-Kids" or "VDKA"), but they do not note that multiple ethicists and experts in clinical trial design who reviewed key trial documents expressed the same concerns. Many studies and clinical guidelines written by authorities in child health and vitamin D also support possible health risks of the low vitamin D levels left untreated in children in the placebo group.

While *Science* appreciates the details offered in the Letter from Celedón, it is important to note that during the course of reporting the article, Piller requested comment on every issue that Celedón raises here. Notably, Piller asked for details on the bone fractures and why they were not publicly reported as adverse events. Celedón elected to respond to Piller's questions with a short general statement. On the day *Science* published the article, the fractures appeared on ClinicalTrials.gov as adverse events.

Piller also requested details on the demographics of the trial sites and the outreach to Black communities during trial enrollment. While the Letters from Celedón and Berg defend the racial breakdown of the trial as appropriate, they avoid discussing an ethical concern shared by experts quoted in the article, which is that any risks of the trial fell predominantly on Black children.

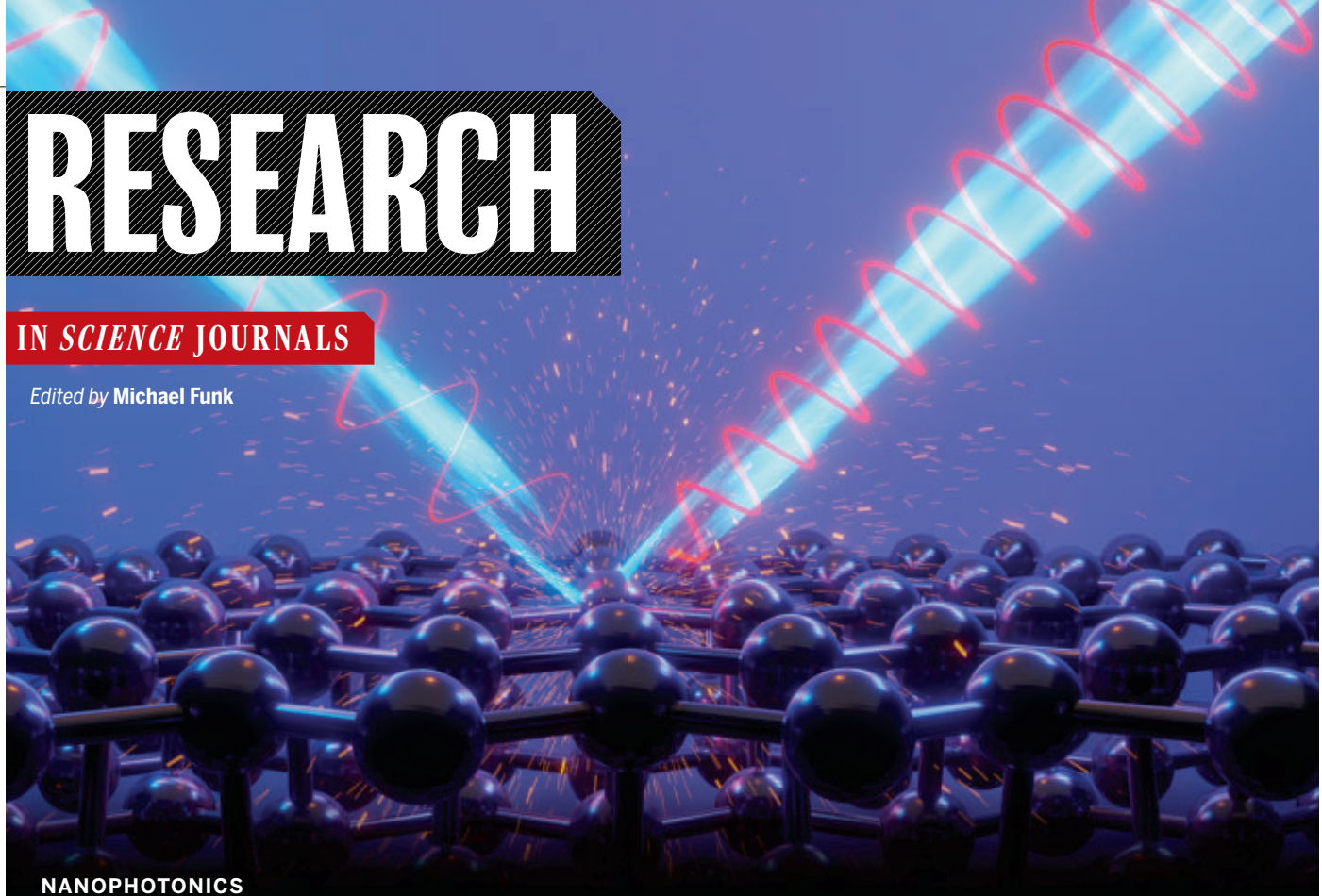
Tim Appenzeller
News editor

10.1126/science.abm7030

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**



NANOPHOTONICS

Metasurfaces tune polarization

In photonic systems, the polarization state of light is typically manipulated with bulky components based on dielectrics and liquid crystals. Active metasurfaces now provide the possibility of shrinking the size of these components to the nanoscale. Biswas *et al*. demonstrate an active nanophotonic structure capable of versatile electro-optic polarization conversion. Using trilayer black phosphorus sandwiched between electrical gates forming a cavity, they show that the polarization state of light across telecommunication wavelengths can be electrically switched. Such control can be useful for a number of applications in communications, imaging, and beam steering where the wavefront of the propagating light can be actively manipulated. —ISO

Science, abj7053, this issue p. 448

Artist's conception of a beam of light reflected off an atomically thin layer of black phosphorous, which can impart tunable variations in polarization

ILLUSTRATION: SOUVIK BISWAS/ATWATER LAB/CALTECH

CORONAVIRUS

Community of antibodies against COVID-19

The severe acute respiratory syndrome coronavirus 2 spike protein is the basis of many vaccines and is a primary target of neutralizing antibodies after COVID-19 infection. The Coronavirus Immunotherapeutic Consortium (CoVIC), comprising 56 partners across the world, has analyzed a panel of 269 monoclonal antibodies (mAbs) and, on the basis of competition profiles, sorted 186 mAbs that target the receptor binding domain into seven communities. Hastie *et al*. went on to structurally analyze representative antibody binding and used pseudovirus neutralization assays to study the effect of spike mutations on antibody function, including the combinations of mutations found in certain variants of concern. These results are important to guide both treatment and prevention efforts. —VV

Science, abh2315, this issue p. 472

METALLURGY

Fine-scale strengthening swirls

Creatively combining different alloys using additive manufacturing methods has the potential to produce materials with interesting properties. Zhang *et al*. used laser powder bed fusion to combine small amounts of 316L stainless steel into a Ti64 titanium alloy. This process creates an alloy with a distinctive microstructure that retains high strength while substantially improving ductility. The design strategy should be useful for improving mechanical properties in other alloy systems as well. —BG

Science, abj3770, this issue p. 478

GUT INFLAMMATION

Locking down access to the brain

Inflammatory bowel disease is best known for intestinal symptoms but can also cause a variety of extraintestinal manifestations in other organs. It can also be associated with cognitive and psychiatric effects, including anxiety and depression. Using mouse models of intestinal inflammation, Carloni *et al*. uncovered a potential pathogenic link between these aspects of inflammatory bowel disease. The inflammatory process causes the gut vascular barrier to become more permeable, resulting in the spread of inflammation beyond the intestine, while the vascular barrier in the choroid plexus shuts down, helping to protect the brain from inflammation but also potentially impairing

communication between organs and impairing some brain functions. —YN

Science, abc6108, this issue p. 439

CORONAVIRUS

SARS-CoV-2 across Africa

The impact of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic has been hard to track in African countries, largely because of patchy data. Wilkinson *et al*. curated viral genomes collected from several countries across the continent. Outbreaks during 2020 in each African country were initiated by imported cases, mostly from Europe. As the pandemic developed, case numbers in African countries were likely many times higher than reported, and subsequent waves of the pandemic appear to have been more severe. Consequently, high-transmission variants have emerged that have spread within the continent, and African countries must be included in global control efforts. —CA

Science, abj4336, this issue p. 423

NANOCATALYSTS

Sulfur-stabilized nanoparticles

Nanoparticles of intermetallic alloys of platinum could have enhanced electronic properties that improve their catalytic activity, but the high temperatures needed to ensure complete atomic diffusion often lead to the growth of larger nanoparticles—sintering—with low surface area and hence low overall activity. Yang *et al*. show that sulfur-doped carbon supports create strong platinum-sulfur bonds that stabilize small platinum alloy nanoparticles (<5 nanometers in diameter) to temperatures up to 1000°C. They screened libraries of platinum alloys and identified ones with high mass activity for the oxygen reduction reaction in hydrogen fuel cells. —PDS

Science, abj9980, this issue p. 459

RADIOCHEMISTRY

Protein captures unnatural elements

Short-lived radionuclides play an important role in the imaging and treatment of many different diseases. The actinium cation in particular has attracted considerable attention for targeted alpha therapy, but purification, extraction, and safe delivery have presented many challenges. Deblonde *et al.* show that a recently discovered natural protein, lanmodulin, has astonishing binding selectivity for actinium, which it certainly never encounters in nature. Binding to lanmodulin allows purification and recovery of actinium even in the presence of 10^{10} equivalents of competing ions. This protein could offer a new strategy for developing versatile chelating platforms for medical applications. —WSW

Sci. Adv. 10.1126/sciadv.abk0273 (2021).

NEUROIMMUNOLOGY

Boosting cleanup after stroke

After ischemic stroke, microglia and infiltrating myeloid cells limit inflammation and subsequent tissue damage by clearing dying cells through the process of efferocytosis. Using a mouse model of ischemic stroke, Nakahashi-Oda *et al.* examined the function of CD300a, a receptor for phosphatidylserine (PS) exposed on dying cells, in myeloid cell efferocytosis and neurological recovery after brain ischemia. CD300a inhibited signaling pathways leading to efferocytosis and competed for PS binding. Genetic deletion of CD300a in myeloid cells increased efferocytosis early after stroke, and treatment of mice with a CD300a-blocking antibody reduced brain inflammation and improved neurological recovery. These results highlight the potential of therapies that enhance efferocytosis during the acute phase of stroke to limit ischemia-induced pathology. —CO

Sci. Immunol. **6**, eabe7915 (2021).

IN OTHER JOURNALS

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

PHOTO: MANDEL NGAN/AFP/GETTY IMAGES

STEM WORKFORCE

Funding determines more than a stipend

One constant of graduate school is the need for funding. Grote *et al.* investigated to what extent different sources of doctoral funding influenced students' development of the four career-related skills of research, project management, mentoring, and communication. Using self-reported student survey data collected across the United States and a series of logistic regression analyses, the authors found significant differences in all four skills depending on the primary funding source for students. Research assistantships led to increased research and project management skills, whereas teaching assistantships led to increased mentoring skills. Students with external fellowships reported lower skill development across all four career-relevant skills. Results suggest that future STEM doctoral students should consider funding sources that foster the skills most closely aligned with their future career goals. —MMc

Int. J. STEM Educ. **8**, 50 (2021).

IMMUNOLOGY

Arms for autoantibody responses

Systemic lupus erythematosus (SLE) is a condition characterized by the activation of autoreactive B and T cells and the production of antibodies against nucleic acids and associated binding proteins. Germinal centers (GCs), areas within lymphoid organs where mature B cells expand and differentiate with T cell help, are thought to play an important role in this process. Chiang *et al*. found that when the costimulatory molecules CD80 and CD86 were specifically deleted from B cells, GC formation was inhibited in a B cell–driven mouse model of SLE. The disappearance of autoimmune GCs resulted in a loss of RNA-associated autoantibodies but not the anti-DNA and connective tissue autoantibody repertoire. Thus, separate

CONSERVATION

Not just removing mountains

Mountaintop coal mining in the Appalachian Mountains, a highly diverse, temperate region of the United States, is a destructive practice that scars the landscape. Moreover, alkaline runoff produced during the mining process damages aquatic organisms in receiving streams. In 2011, this impact led the US Environmental Protection Agency (EPA) to establish criteria for river pH by specific conductivity measurements. Simonin *et al*. assessed environmental DNA to survey species richness downstream of mining activities and found significant negative relationships between pH and species richness in algal, bacterial, and eukaryote species and shifts in community compositions. These changes occurred at pH values well below the EPA cutoff, indicating that current standards are not protecting these sensitive and globally important systems. —SNV *Ecol. Appl.* **31**, e02389 (2021).

The top of Kayford Mountain in West Virginia has been demolished to extract coal, a process still causing environmentally harmful alkaline runoff despite regulations.

IMAGE: THOMAS DEERINCK/NCMIR/SCIENCE SOURCE

extrafollicular and GC-derived pathways may control discrete arms of the autoantibody response in this disease. —STS

J. Immunol. 10.4049/jimmunol.2100548 (2021).

POLICING

Manipulation of policing metrics

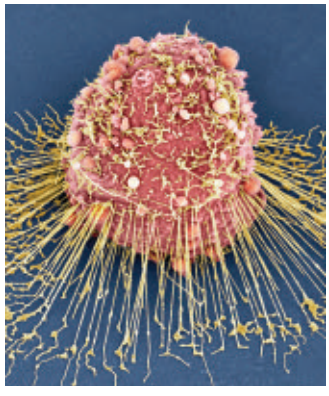
The use of metrics-focused management systems by police departments in 47 large US cities contributed to increased arrests for minor crimes without reducing serious crime. Using national crime reporting data spanning from 1990 to 2013, Eckhouse also shows that use of the CompStat system contributed to data manipulation in the form of downgrading of a serious crime, rape, which was often reclassified as "unfounded," suggesting that the crime did not occur. Although the use of data in public policy and management is critical, attention must be paid to how measurements and employee incentives are aligned so that attempts to manipulate the metrics can be mitigated. —BW

Am. J. Pol. Sci. 10.1111/ajps.12661 (2021).

NANOPARTICLES

Cell entry via microvilli

Therapeutic applications and potential harmful effects of nanoparticles depend on entry into cells, which is usually thought to occur through endocytosis or phagocytosis. Sommi *et al*. report that ceria and magnetite nanoparticles with a negatively charged coating (polyacrylic acid) enter mammalian cells by adhering to microvilli, the cell surface protrusions involved in processes such as nutrient adsorption and cell adhesion. The mechanism, which is not cell specific and was observed in HeLa cells and other human and primate cell types, may involve interaction of lipid rafts with the coating, because cells depleted in cholesterol showed decreased uptake. This mechanism could be exploited for drug delivery to gastric and intestinal epithelial cells, which have a high density of microvilli. —PDS



Some nanoparticles can enter mammalian cells, such as the HeLa cell pictured above, by adhering to their microvilli.

ACS Nano 10.1021/acsnano.1c03151 (2021).

MICROBIOLOGY

Coordinating pack hunting

Some bacterial species kill other microorganisms and use these prey cells as nutrients. The soil bacterium *Myxococcus* uses collective motility to seek and kill its prey by release of lytic enzymes, antibiotics, and outer membrane vesicles. Seef *et al.* studied how *Myxococcus* cells invade and grow over *Escherichia coli* prey cells. They found that the A-motility complex works in concert with another mechanism called the Kil complex. These systems coordinate to sense prey cells and, upon contact, pause motility and puncture the prey cell envelope. Kil is a type IV filamentous nanomachine related to the Tad systems involved in bacterial adhesion and secretion. The *Myxococcus* Kil complex recognizes other prey, including *Caulobacter crescentus* and *Bacillus subtilis*, and a Kil homolog mediates prey recognition and invasion in the intracellular predator *Bdellovibrio bacteriovorus*. Investigations on interspecies killing could reveal novel antimicrobial molecules and processes with clinical potential. —SMH

eLife **10**, e72409 (2021).

NEUROSCIENCE

To choose optimally

Goal-directed behaviors are needed for flexible responses to changing conditions. Some individuals may be more reliant on habitual responses because they have impaired abilities to detect causal relationships. The consequences can be disastrous. Candidate neural systems for mediating such action-outcome contingencies have been located to subregions of the prefrontal cortex and the subcortical caudate nucleus. Duan *et al.* investigated the anatomical details in primates and found that just prefrontal cortical area 24 and its target region in the caudate nucleus were necessary for detecting causal relationships. Anterior orbitofrontal area 11 regulated the sensitivity of the response. These findings elucidate goal-directed behavior and provide insights into psychiatric conditions such as obsessive-compulsive disorder. —PRS

Neuron **109**, 2485 (2021).

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

CORONAVIRUS

A smaller-dose jab does the job

Low-dose messenger RNA (mRNA) vaccines potentially allow health providers to administer more doses from a limited vaccine supply and can be less reactogenic. However, whether low-dose COVID-19 mRNA vaccines generate immune responses comparable to currently approved doses remains an open question. Mateus *et al*. report the results of a clinical trial comparing patients who received a 25-μg mRNA-1273 (Moderna) COVID-19 vaccine with 100-μg mRNA-1273 COVID-19 vaccinees and severe acute respiratory syndrome coronavirus 2–infected individuals. The low-dose Moderna vaccine generated long-lived T cell immunity that was equivalent between younger and older patients and could be enhanced by the presence of cross-reactive T cells. Moreover, antibody and T cell responses induced by the low-dose vaccine were comparable to natural infection and about half as strong as those seen with high-dose vaccination. —STS

Science, abj9853, this issue p. 420

SOCIAL NEUROSCIENCE

A mind's-eye view of others

Social interaction among groups of individuals is a complex proposition. Not only must an animal keep track of various vocalizations and direct interactions in the present but likely also their knowledge of every other individual and their history of interaction with that individual. Two studies begin to unravel the neuronal process by which such complexities are managed (see the Perspective by Sliwa). Báez-Mendoza *et al*. tracked the interactional dynamics among three Rhesus macaques and found that neurons in the dorsomedial prefrontal cortex represent details of the interaction, such as identity, context, and interaction history. Rose *et al*. remotely recorded from freely interacting Egyptian fruit bats and similarly found coordinated neural activity among individuals and a relationship between brain activity patterns and social preference. They also found that single neurons in the prefrontal cortex distinguished between the vocalizations of specific individuals. Together, these studies reveal clear evidence for neuronal encoding of social interaction and identity. —SNV

Science, abb4149, aba9584, this issue p. 421, p. 422; see also abm3060, p. 397

ORGANIC CHEMISTRY

Iron links a trio

Iron holds particular appeal as a catalytic metal: it is safe and abundant and is a mainstay of enzymatic reactivity. Nonetheless, in synthetic construction of carbon-carbon bonds, modern chemists have largely had to rely on rarer metals such as palladium. Liu *et al*. now report that coordination of iron by a bulky chelating phosphine ligand enables efficient mutual coupling of three different reactants, an alkyl halide, an aryl Grignard, and an olefin, to form two carbon-carbon bonds (see the Perspective by Lefèvre). A combination of Mössbauer spectroscopy, crystallography, and computational simulations illuminated the mechanism. —JSY

Science, abj6005, this issue p. 432; see also abl4762, p. 396

WATER CAPTURE

Designing water uptake

Although the locations of water molecules in some porous materials have been determined with diffraction techniques, determining the filling sequence of water sites has been challenging. Hanikel *et al*. used single-crystal x-ray diffraction to locate all of the water molecules in pores of the metal-organic framework MOF-303 at different water loadings (see the Perspective by Öhrström and Amombo Noa). They used this information on the water molecule adsorption sequence to modify the linkers of this MOF and control the water-harvesting properties from humid air for different temperature regimes. —PDS

Science, abj0890, this issue p. 454; see also abm1854, p. 402

STRUCTURAL MATERIALS

Turning wood into honeycombs

Wood is an attractive material for structural applications but it usually works best as boards or sheets. Xiao *et al*. have developed a process for engineering hardwood that allows these sheets to be manipulated into complex structures (see the Perspective by Tajvidi and Gardner). The key is to manipulate the cell wall structure by shrinking and blasting open the fibers and vessels by drying and "water-shocking" them. This process creates a window during which the wood can be manipulated without ripping or tearing. Honeycomb, corrugated, or other complex structures can be locked in once the wood dries. —BG

Science, abg9556, this issue p. 465; see also abm3882, p. 400

POACHING IMPACTS

Lose the tusks

Harvest and poaching of wildlife have increased as the human population and our technology have grown. These pressures now occur on such a scale that they can be considered selective drivers. Campbell-Staton *et al*. show that this phenomenon has occurred in African elephants, which are poached for their ivory, during the 20-year Mozambican civil war (see the Perspective by Darimont and Pelletier). In response to heavy poaching by armed forces, African elephant populations in Gorongosa National Park declined by 90%. As the population recovered after the war, a relatively large proportion of females were born tuskless. Further exploration revealed this trait to be sex linked and related to specific genes that generated a tuskless phenotype more likely to survive in the face of poaching. —SNV

Science, abe7389, this issue p. 483; see also abm2980, p. 394

MICROBIOLOGY

Wild phage evolution

Bacteria have a diversity of highly specific phage defense mechanisms that evolve rapidly and account for a large proportion of bacterial genomes. These dynamics must be understood if phage therapy is to be seriously considered for clinical use. Hussain *et al*. studied a set of nearly clonal wild-caught *Vibrio lentus* hosts and their phage and found that the hosts segregated into two groups infected by different viruses (see the Perspective by Meaden and Fineran). Puzzlingly, both host groups have the same surface phage receptors. This apparent paradox was resolved by sequencing of the hosts, which revealed distinct suites of endogenous but mobile phage defense elements (PDEs) in the two host phenotypes. Indeed, the PDEs constitute a large fraction of the flexible noncore genome of the bacteria. This means that the PDEs can evolve and transfer from cell to cell without interfering with the metabolic processes required to synthesize essential cell surface molecules. —CA

Science, abb1083, this issue p. 488; see also abm2444, p. 399

VACCINES

Fatal flaws unearthed

Two toddlers vaccinated with a formalin-inactivated vaccine against respiratory syncytial virus (FIRSV) died in 1967 from

enhanced RSV disease (ERD), and most FIRSV-immunized children required hospitalization upon RSV infection. Polack *et al*. characterized the clinical, immunopathological, and transcriptional profiles of ERD from the original autopsy reports and lung sections of the toddlers who died against age- and race-matched controls and identified respiratory eosinophils and a particular chemokine as markers that correlated with ERD. Type I interferon responses were suppressed, and pulmonary immune responses showed a shift toward type 2 polarization, thus validating observations in murine RSV models. These findings provide critical insights into ERD and how it can inform RSV vaccine development. —CNF

Sci. Transl. Med. **13**, eabj7843 (2021).

CANCER THERAPY

Tailoring cancer vaccines to patients

The immune response to cancer can be enhanced by molecules released from dying tumor cells, leading to proposals to use dead tumor cells as adjuvants in immunotherapeutics and cancer vaccines. However, Sriram *et al*. discovered that it is sometimes the remaining live tumor cells that are immunogenic. Treating cultured tumor cells with DNA-damaging chemotherapies stimulated cytotoxic T cell activation when the live, not the dead, cell fraction was injected back into tumors in mice. Combining these cells with immune checkpoint inhibitors resulted in tumor regression and antitumor immune memory. —LKF

Sci. Signal. **14**, eabc4764 (2021).

CORONAVIRUS

The importance of data

The severe acute respiratory syndrome coronavirus 2 pandemic has been tracked with numerous types of data, particularly numbers and rates of cases, hospitalizations, and deaths. This information is important to inform mitigation and policy decisions, but there are limitations with these types of data, and some data are not tracked routinely. In a Perspective, Pagel and Yates discuss what the data do and do not tell us about epidemics and highlight the data that are missing, including counting Long Covid and breakthrough infections in vaccinated individuals. —GKA

Science, abi6602, this issue p. 403

NANOPHOTONICS

Exploiting light with a twist

Light has several degrees of freedom that can be exploited to encode information. In addition to wavelength, amplitude, phase, polarization, and pulse length, light can also be structured into vortices. These vortices carry optical angular momentum, effectively opening up many other modes as communication channels. Ni *et al*. review advances made in the nanophotonic platform where optical vortices can be generated and detected on-chip. These developments are expected to find application in classical and quantum information processing. —ISO

Science, abj0039, this issue p. 418

CLIMATE CHANGE

The impacts of marine heatwaves

Extreme warming events in the world's oceans are becoming more widespread and frequent: Eight of the 10 most severe recorded events have taken place in the past decade. Smith *et al*. review how these marine heatwaves are severely altering ecosystem service provision, with widespread socioeconomic impacts. Heatwave effects, including range shifts and mass mortality of marine species and harmful algal blooms, have knock-on economic consequences that already run into billions of US dollars. In addition to reviewing the consequences of these events, the authors discuss the mitigation and adaptation measures that are needed to alleviate the risks and damaging impacts. —AMS

Science, abj3593, this issue p. 419

REVIEW SUMMARY

NANOPHOTONICS

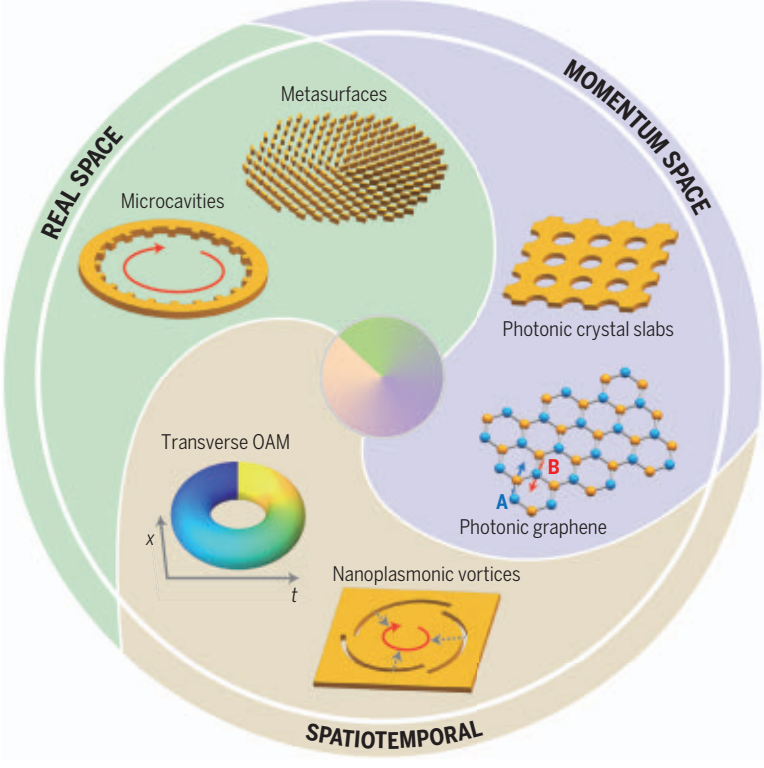
Multidimensional phase singularities in nanophotonics

Jincheng Ni†, Can Huang†, Lei-Ming Zhou, Min Gu, Qinghai Song*, Yuri Kivshar*, Cheng-Wei Qiu*

BACKGROUND: Optical vortices provide one of the promising solutions for enhancing the capacity of data multiplexing to meet an unprecedented growth in big data and internet traffic information. The optical vortex possesses a phase singularity—an undefined phase in the beam center—and has attracted a special attention for carrying the orbital angular momentum (OAM) of light. The photonic OAM states supply new degrees of freedom for light, which can be used in optical manipulation, enantiomer-selective sensing, microscopic imaging, and optical communications. In particular, the unlimited OAM states can potentially provide more channels for data transmission in optical communications and quantum information processing. Although the benchtop optical systems composed of bulky optical components can stably realize the generation and detection of vortex beams, many practical applications of photonic OAM require shrinking optical systems to on-chip devices for matching the scale of optical fibers. Toward this aim, the phase singularity of light has raised considerable attention in multiple dimensions, with the vortex generators and detectors being desired at the micro- and nanoscale for integration with compact nanophotonic devices.

ADVANCES: Here, we discuss optical vortices in subwavelength photonic structures, generated in multiple dimensions such as real space, momentum space, and the spatiotemporal domain. In real space, stable OAM beams can be generated by microscopic devices such as microscopic spiral phase plates, metasurfaces, and planar spiral nanostructures with a helical phase delay. The vortex generators in the form of whispering-gallery–mode microcavities can also achieve nondegenerate OAM modes from incident light with spin-momentum locking. In momentum space, phase singularities also appear in photonic crystal slabs near the condition of bound states in the continuum being induced by the Pancharatnam-Berry phase. Such noncenter periodic nanostructures can flexibly generate vortex beams without nanoscale aligning problems. Vortex beams generated by nanophotonic structures can still exhibit features identical to those of conventional bulk devices, including phase modulation and an efficient optical switch. On the receiving end, nanophotonic structures can well distinguish the OAM eigenstates in light beams through interference patterns or the photocurrent effect. Particularly, these microscopic vortex generators and detectors allow convenient integration with optical fibers or on-chip photonic devices for optical communications and quantum information processing. Furthermore, the multidimensional vortex generators with nonlinear, ultrafast, time-varying, and transverse OAM make optical communications possible with higher capacity and improved security.

OUTLOOK: Optical vortices being fundamentally analogous with superfluid vortices in nature, the mindset around their physics has undergone a revolution in the past three decades. The study of vortex beams generated by micro- and nanoscale structures has been established as a new research direction in compact and integrated nanophotonics. Meanwhile, the research on optical vortices is extended from the conventional real-space dimension to the momentum space and time- and frequency-varying synthetic dimensions. The extension of dimensionality brings entirely new physics and design philosophies to nanophotonic vortex devices. In addition, on-chip vortex generation, transmission, and detection demonstrate unprecedented functionalities in broader potential applications, especially in optical communications and quantum information processing. From the maturity of semiconductor technology for integrating optical and electrical components on a single chip, we expect to see very soon many practical applications and commercial products based on these research results. ■



Multidimensional phase singularities with on-chip nanophotonics. Schematic illustration of phase singularities from real space, momentum space, and the spatiotemporal domain emerged in integrated micro- and nanoscale photonics, with applications in optical communications and quantum information processing.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: qinghai.song@hit.edu.cn (Q.S.); ysk@internode.on.net (Y.K.); chengwei.qiu@nus.edu.sg (C.Q.)
†These authors contributed equally to this work.
Cite this article as: J. Ni *et al.*, *Science* **374**, eabj0039 (2021). DOI: 10.1126/science.abj0039

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abj0039

REVIEW

NANOPHOTONICS

Multidimensional phase singularities in nanophotonics

Jincheng Ni[1]†, Can Huang[2]†, Lei-Ming Zhou[1], Min Gu[3,4], Qinghai Song[2,5]*, Yuri Kivshar[6]*, Cheng-Wei Qiu[1]*

Rapid progress in miniaturizing vortex devices is driven by their integration with optical sensing, micromanipulation, and optical communications in both classical and quantum realms. Many such efforts are usually associated with on-chip micro- or nanoscale structures in real space and possess a static orbital angular momentum. Recently, a new branch of singular optics has emerged that seeks phase singularities in multiple dimensions, realizing vortex beams with compact nanodevices. Here, we review the topological phase singularities in real space, momentum space, and the spatiotemporal domain for generating vortex beams; discuss recent developments in theoretical and experimental research for generation, detection, and transmission of vortex beams; and provide an outlook for future opportunities in this area, ranging from fundamental research to practical applications.

Optical devices have been widely developed to manipulate light's different degrees of freedom (such as amplitude, wavelength, phase, polarization, and angular momentum) for modern optics and photonic techniques. In particular, the photonic angular momentum is demonstrated as a new degree of freedom in light with chiral states, which lies at the heart of both fundamental theoretical research and advanced applications in modern physics (*1*, *2*). More than 100 years ago, Poynting (*3*) reasoned that the spin angular momentum (SAM) of light, limited by two polarization states ($\mathbf{S} = \pm\hbar$, where $\hbar$ is Planck's constant h divided by 2π), is related to the left- or right-handed circular polarization with helical electric vectors. By contrast, the photonic orbital angular momentum (OAM) is intrinsically carried by vortex beams with phase singularities, which can be evaluated by $\mathbf{L} = \ell\hbar$, where the integer ℓ stands for the topological charge, and its sign indicates the handedness of the helical wavefront (*4*). Analogous to ubiquitous vortices in nature such as hurricanes and whirlpools, the vortex beams are screwed along the propagating direction, resulting into a dark core at the beam center (*5*). Recently, the photonic OAM has been widely applied in the fields of optical manipulation (*6*, *7*), imaging and microscopy (*8*, *9*), enantiomer-selective sensing (*10*, *11*), and optical communications (*12*–*14*). Although the wavelength and polarization of light have been widely used to increase the transmission capacity through data multiplexing, the fast development of information processing has seen the capacity limit in conventional technology. By virtue of the unlimited OAM states, in principle, the OAM of light can dramatically boost the channels for data transmission and storage in optical communications and quantum information processing (*15*), which are considered to be the next generation of high-capacity photonics (*12*, *16*–*18*). The comprehensive investigations on photonic OAM states even created a new chapter of modern optics—singular optics—which is a great leap forward in the development of classical and quantum optics (*19*, *20*).

Up to now, the extensive applications of photonic OAM, however, remain challenging with bulky devices such as spiral phase plates (*21*) and spatial light modulators (*22*). Recent decades have witnessed a growing interest in the study of micro- and nanoscale vortex generators and detectors on compact devices. The compact OAM devices endow vortex beams with the ability of compatible, high-efficiency manipulation and transmission at the micro- and nanometer scale. Generally, in the original sources of vortex beams, phase singularities are considered to extract from real space. To improve the singularity properties for practical applications, other dimensions of phase singularities also need to be explored such as in momentum space and spatiotemporal domain. In addition, the continuous and nonadjustable OAM modes would hinder the process of optical information communications. To this end, the generated vortex beams should demonstrate dynamically tunable properties with nonlinear, time-varying, and even transverse OAM states.

In this Review, we first summarize the fundamental concepts of phase singularities in real space, momentum space, and the spatiotemporal domain, providing an introduction of generating phase singularities in multiple dimensions. We then mainly review recent theoretical and experimental efforts in generation, detection, and transmission of vortex beams by use of micro- and nanoscale optical devices. The interested reader may find more detailed applications of optical vortices in classical and quantum optics in recent review papers (*23*–*25*). Last, we provide an outlook for the future of the phase singularity realm and applications.

Observation of phase singularities in nanostructures

The predominant properties of vortex beams are reflected on the helical wavefronts and phase singularities. Optical vortices possess an azimuthal gradient phase distribution, which demonstrates an undefined phase at the center of light beams (Fig. 1A). Therefore, optical vortices have a donut-shape intensity profile, yielding a dark core on the transverse plane of light beam because of the certain wave function. The topological charge of vortex beam can be defined as

$$\ell = \frac{1}{2\pi}\oint_C d\boldsymbol{r}\cdot\nabla_{\boldsymbol{r}}\phi(\boldsymbol{r}) \qquad (1)$$

where C is a closed loop around the phase singularity, and ϕ is the phase distribution. For a light beam with phase factor of $e^{i\ell\varphi}$, ℓ is the topological charge, and φ is the azimuthal angle.

With conventional bulky optical components, the polarization of a generated vortex beam (the oscillation direction of its electric field) is normally spatially homogeneous. The scalar-approach generations of vortex beams are empowered by a dynamic phase, in which the wavefront is tailored by accumulated phases along the propagating direction. However, using artificially subwavelength structures, it was also proposed that vector vortex beams can be produced by a geometric phase or Pancharatnam-Berry phase (*26*). The phase singularity, coexistent with polarization singularity, emerges from the spatially distributed polarization states of light, represented as different paths on the Poincaré sphere. The vector vortex beams have demonstrated many advantages in compact on-chip devices such as metasurfaces (Fig. 1B) (*27*–*30*). In particular, the on-chip devices with subwavelength

[1]Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117583, Singapore. [2]State Key Laboratory on Tunable Laser Technology, Ministry of Industry and Information Technology Key Laboratory of Micro-Nano Optoelectronic Information System, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518055, China. [3]Institute of Photonic Chips, University of Shanghai for Science and Technology, Shanghai, China. [4]Centre for Artificial-Intelligence Nanophotonics, School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai, China. [5]Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan, 030006 Shanxi, China. [6]Nonlinear Physics Centre, Research School of Physics, Australian National University, Canberra ACT 2601, Australia.
***Corresponding author. Email: qinghai.song@hit.edu.cn (Q.S.); ysk@internode.on.net (Y.K.); chengwei.qiu@nus.edu.sg (C.Q.)**
†These authors contributed equally to this work.

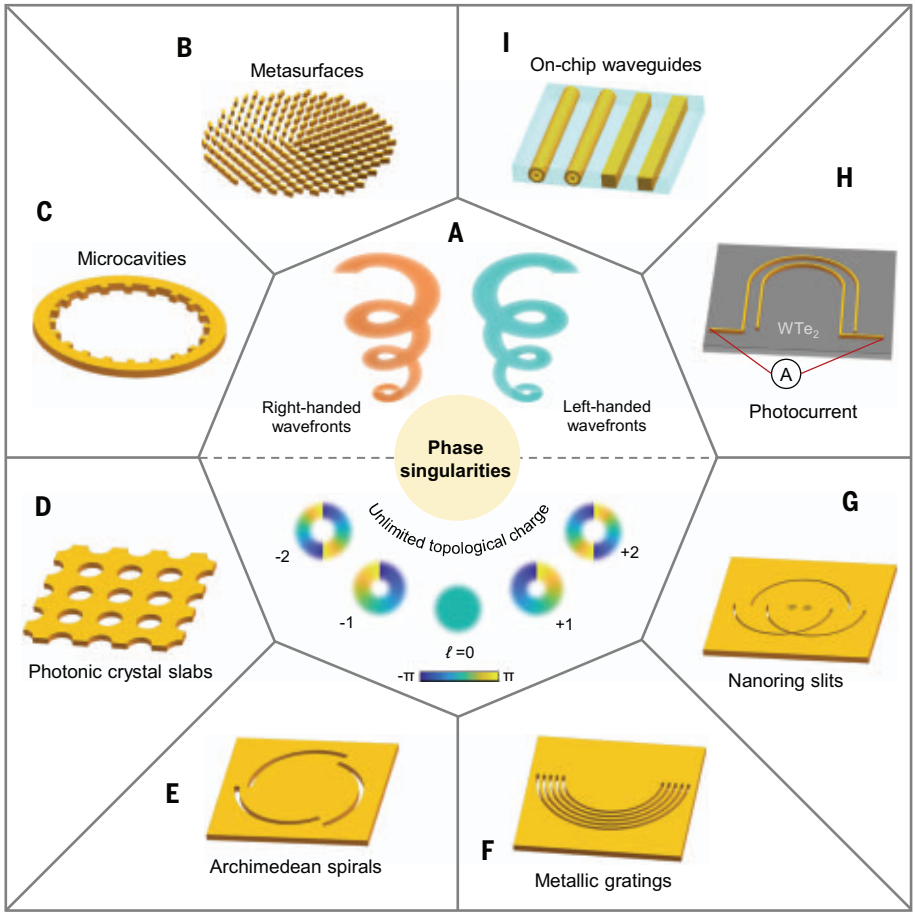


Fig. 1. Vortex beams carrying phase singularities in nanophotonics. (**A**) Schematic illustration of vortex beams with helical wavefronts and unlimited topological charge ℓ. (**B**) Metasurfaces with geometric phase provide phase singularities for generating vortex beams. (**C**) Vortex beams extracted in unidirectional WGM microcavities. (**D**) Phase singularities in momentum space generated by photonic crystal slabs. (**E**) Planar Archimedean nanostructures excited phase singularities in spatiotemporal domain. (**F**) On-chip discrimination of OAM beams by metallic gratings. (**G**) Nanoring slits for both SAM and OAM multiplexing of broadband light. (**H**) Photocurrent measurements for discriminating OAM of vortex beams. (**I**) Transmission of OAM beams in on-chip waveguides.

structures can efficiently generate vector vortex beams with high-purity OAM modes (*31*–*33*).

To illustrate the phase singularities of light beams, we consider the generation of vortex beams in various micro- and nanoscale photonic devices. In real space, it is widely demonstrated that vortex beams can be achieved through chiral light-matter interaction. For example, the helical wavefront of a vortex beam can be extracted from a whispering-gallery–mode (WGM) microcavity (Fig. 1C). The unidirectional OAM mode can be controlled by the direction of incident light. Additionally, the momentum-space phase singularity is contained in photonic crystal slabs (Fig. 1D) (*34*). In the spatiotemporal domain, vortex beams can be provided by planar-chiral nanostructures such as Archimedean spirals (Fig. 1E), yielding phase delays along the nanostructures (*35*, *36*). On the receiving end, miniaturized vortex detectors are required for discriminating OAM modes with metallic gratings, nanoring slits, and even photocurrent measurements (Figs. 1, F to H). To avoid the atmospheric turbulence in free space, the stable transmission of vortex beams is realized in on-chip integrated waveguides (Fig. 1I) (*37*, *38*). These on-chip vortex devices could find applications in the next generation of integrated optoelectronic devices for optical communications in both classical and quantum realms.

OAM generators in real space

In real space, although the conventional generation of vortex beams usually relies on bulky devices, miniaturized, robust, and high-efficiency OAM generators are proposed for facilitating on-chip functionalities of optical communications. The micro- and nanoscale vortex beams have been widely realized through micro- and nanostructures such as WGM microcavities and metasurfaces (*39*). A microcavity is likely the most simple and efficient structure with which to generate vortex beams. To extract the confined OAM modes in a WGM microcavity, angular grating structures are designed into the microring resonator (*18*). However, the resonant microcavity usually coexists with clockwise and counterclockwise WGM, circulating with opposite topological charges. Some efforts have been used to avoid the mode degeneracy. For example, by use of directional input couplers (*40*, *41*), only clockwise or counterclockwise OAM modes are extracted into the free space to generate cylindrical vector vortex beams (Fig. 2A). In the antenna-decorated microcavity, the generation of vortex beams can be realized with pure polarization states (*42*). Moreover, by exploiting the emerging non-Hermitian photonics design at an exceptional point, a microring resonator was reported to produce a single-mode OAM beam with specific topological charge (Fig. 2B) (*43*). The exception-point operation breaks the mirror-symmetry in the microcavity and facilitates the unidirectional power oscillation. In addition, single OAM modes can be excited by circularly polarized light with optical breaking of time-reversal symmetry (*44*), implying an optical-controlled chirality of the helical wavefront (Fig. 2C). Using the photonic quantum Hall effect, the single OAM mode can also be obtained at the topological interface under a static magnetic field (*45*).

Given the wavelength-dependent WGM microcavity, the magnitude of the OAM value is typically not tunable for the integrated device at a single wavelength. Recently, it was reported that an OAM-tunable vortex microlaser was provided at a single telecommunication wavelength by using a radial polarizer, resulting in the OAM-SAM-wavelength division multiplexing for high-capacity optical communications (Fig. 2D) (*46*). The dynamic switching between different topological charges in the time dimension can extremely increase the capacity and security of a communication system. The WGM microcavity is also an efficient platform for achieving single photons carrying OAM, which may enable high-dimensional quantum information processing. It was reported that a bright solid-state source of single photons with an OAM superposition state was realized by use of embedded quantum dots in a microcavity (Fig. 2E) (*47*). Obvious azimuthal patterns with multiple antinodes can be found in the far field because of the coexistence of opposite OAM modes in the WGM microcavity.

By virtue of a direct-laser-writing technique, microscopic spiral phase plates can be fabricated for realizing the miniaturization of vortex generators (*48*, *49*). Using the microscopic spiral phase plates, the phase singularity is still generated by a dynamic phase delay, in which

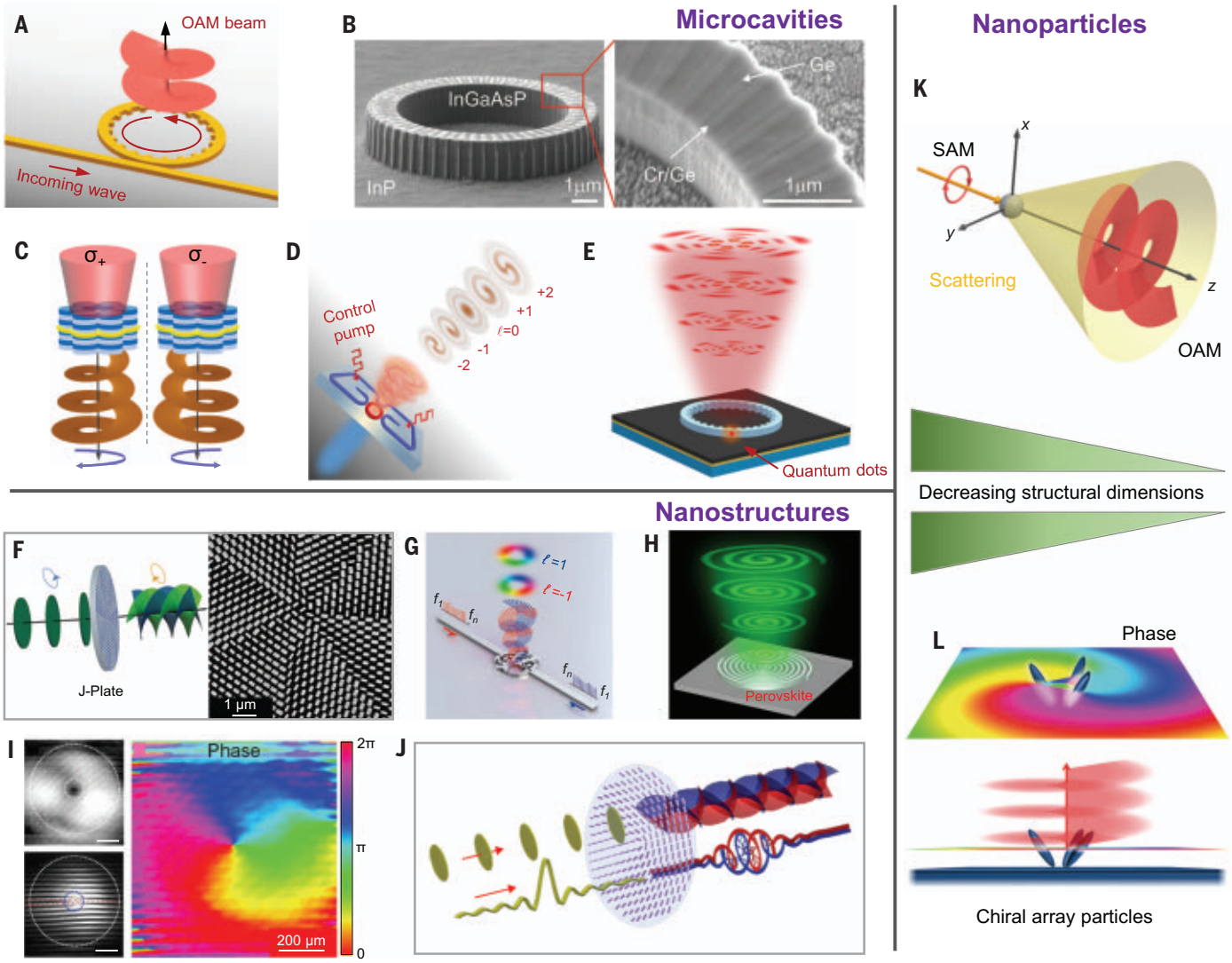


Fig. 2. Generation of vortex beams in real space. (**A**) Illustration of a vortex generator with angular gratings distributed along the inner wall of a microring resonator. (**B**) Scanning electron microscope (SEM) images of WGM microlaser fabricated on the InGaAsP/InP platform. [Adapted with permission from (*43*)] (**C**) Schematic representation of optically controllable OAM microlaser in a benzene-shape microcavity (*44*). (**D**) The non–Hermitian-controlled vortex generator with OAM states switched by the externally applied control pump. [Adapted with permission from (*46*)] (**E**) Microcavity embedded with quantum dots for the generation of single photons carrying quantum OAM superposition states (*47*). (**F**) Concept of arbitrary SAM-to-OAM transformation and SEM image of the fabricated metasurface. [Adapted with permission from (*52*)] (**G**) Schematic illustration of nanostructure design of the broadband multiplexed OAM emitter. [Adapted with permission from (*53*)] (**H**) Planar-Archimedean-spiral gratings for vortex microlasers (*54*). (**I**) Vortex beam generated by a Huygens dielectric metasurface. (Left) Optical intensity profile of (top) generated OAM beam and (bottom) interference pattern. (Right) Reconstructed phase of the beam. Scale bars, 200 μm. [Adapted with permission from (*57*)] (**J**) Schematic of entanglement between SAM and OAM on a single photon by means of metasurface. [Adapted with permission from (*61*)] (**K**) Scattering by a single dipole nanoparticle to generate a helical phase vortex. (**L**) Chiral array of three particles supports vortex emission. [Adapted with permission from (*67*)]

the polarization of the vortex beam is homogeneous. By contrast, associated with polarization states, optical components can transform the incident circularly polarized light into vortex beams by means of a *q*-plate or metasurfaces (*30*, *50*). The SAM-to-OAM converters based on the Pancharatnam-Berry phase permit the conversion of left- and right-handed circular polarizations into opposite OAM states (*27*, *51*). As shown in Fig. 2F, it was reported that arbitrary SAM states can be converted into a superposition of total angular momentum states of light with metasurfaces (*52*). Such metasurfaces can convert left- and right-circular polarizations into states with independent OAM values, using titanium dioxide (TiO_2) as a material platform with high fidelity and high efficiency at visible wavelengths (*33*).

Integrated with a waveguide, an on-chip OAM emitter with nanoscale features can generate a broadband multiplexed OAM beam (Fig. 2G). Coaxial OAM beams are emitted by using a joint path-resonance phase control concept, realizing the operating wavelength across the entire telecommunication band from 1450 to 1650 nm (*53*). Imposing a phase shift along a circle, the Archimedean spiral grating is an effective nanodevice for vortex generation (Fig. 2H). The spiral perovskite nanostructures can excite a microlaser at the wavelength of around 540 nm (*54*). In high-efficiency polarization-sensitive dielectric metasurfaces, it was reported that OAM beams can

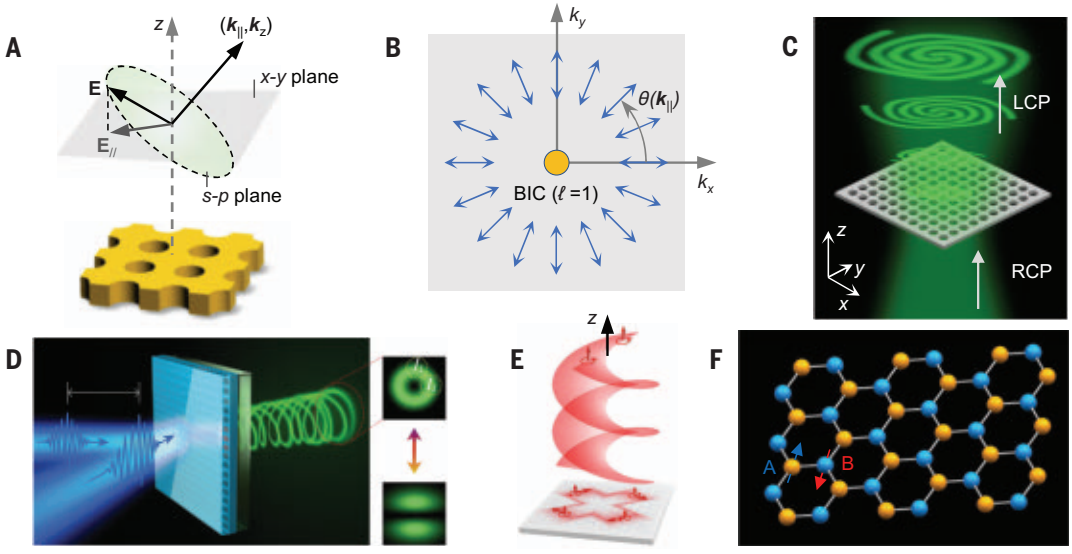


Fig. 3. Optical vortices generated in momentum space. (**A**) Schematic of photonic crystal slab at bound states in the continuum for vortex centers of polarization vectors (*34*). (**B**) Polarization states distribution near a BIC with a topological charge of 1 in momentum space. Double-headed arrows indicate the states of polarizations, projecting onto the momentum-space plane. (**C**) Passive vortex generator on a photonic crystal slab excited by circularly polarized light (*71*). (**D**) Schematic of two-beam pumping experiment for switching vortex beams on photonic crystal slabs. Two beams are spatially detuned with a distance, shifted temporally with a delay time. (Insets) The far-field emission patterns from the perovskite metasurface under both symmetric and asymmetric excitations. [Adapted with permission from (*72*)] (**E**) Schematic of the topological vortex laser with spin-momentum–locked edge mode for generating circularly polarized vortex beams. [Adapted with permission from (*73*)] (**F**) Schematic of pseudospin-mediated vortex generator in photonic graphene (*74*). The arrows indicate the pseudospin representation of honeycomb sublattice (A or B) degree of freedom.

be obtained from incident light with specific polarization states for phase control (*55*, *56*). By varying the lattice periodicity in a nanodisk array (*57*), polarization-independent vortex beams are generated by a silicon Huygens' metasurface (Fig. 2I). The specific phase change can also be achieved by tailoring the diameter of structures in an ultracompact array of nanowaveguides (*58*). Distinguishing from conventional spiral phase plates by phase accumulation in a medium, compact metasurfaces with nanoscale thickness are convenient to integrate with nanophotonic devices.

More recently, OAM states have been suggested as a new platform for quantum optics, in which the entanglement of photon pairs in OAM states can be generated through spontaneous parametric down-conversion in nonlinear crystals (*15*, *59*, *60*). From the perspective of photonic integrated devices, it is of great importance to further miniaturize the quantum system. The dielectric metasurfaces can efficiently generate entanglement between the SAM and OAM of photons (Fig. 2J). The generation of four Bell states on a single photon are demonstrated by using the geometric phase that arises from the photonic spin-orbit interaction, showing nonlocal correlations between two photons that interacted with the metasurface (*61*). Besides, high-index dielectric metasurfaces have been proposed to efficiently facilitate the generation of vortex beams with high transmission and low optical losses (*62*–*64*). These results illustrate a general material-mediated connection between SAM and OAM of light and may find applications in producing complex structured light and optical communications. Further decreasing the size of OAM generators, the SAM-to-OAM conversion can be realized with a single nanoparticle (Fig. 2K) (*20*, *65*, *66*). Recently, it was also reported that vortex beams can be generated from a chiral array structure with three identical nanoscale emitters (Fig. 2L) (*67*). These nanoscale OAM generators are convenient for the integration and miniaturization of vortex-related nanodevices.

Phase singularities in momentum space

In contrast to the real-space phase singularities, vortex beams can be obtained in momentum space excited by photonic nanostructures. Recently, it has been clarified that photonic crystal slabs can possess nontrivial vectorial topological properties in momentum space (*68*). The dielectric slabs have a periodic modulation of refractive index at the wavelength, giving rise to photonic bands. Optical bound states in the continuum (BICs), being the most famous example, can be realized in photonic crystal slabs, in which the out-of-plane radiation perfectly turns into guided modes with infinite lifetimes (*69*). A fourfold rotational-symmetric photonic crystal slab has a vortex center in the far-field radiation, which inherently exists near BICs (Fig. 3A) (*34*). The symmetry-protected BIC with conserved and quantized topological charges demonstrates robust properties, which can only be generated or annihilated by large variations in the structural parameters. In momentum space, the polarization vector forms a vortex around each BIC with a corresponding topological charge (*69*). Distinguishing from the real-space condition, observing topological polarization singularity in momentum space generally is in the Fourier plane (*70*). Perturbations can turn a BIC into a leaky resonance, resulting in a far-field radiation. Therefore, using the Fourier plane image technology, the vector vortex beam of degenerate state in momentum space can be observed intuitively. The topological charge in momentum space can be defined as

$$\ell = \frac{1}{2\pi}\oint_C d\boldsymbol{k}_{\parallel}\cdot\nabla_{\boldsymbol{k}_{\parallel}}\theta(\boldsymbol{k}_{\parallel}) \qquad (2)$$

where C is a closed loop around the singularity, and $\boldsymbol{k}_{\parallel}$ is the in-plane wave vector. The polarization vectors, which are almost linear, create a circulation of 2π in a closed path around the BIC in momentum space (Fig. 3B). Because the polarization vector must come back to itself after the closed loop, the overall angle change has to be an integer multiple of 2π, and the topological charge ℓ must be an integer.

By introducing the Pancharatnam-Berry phase in momentum space, vortex beams can be generated on the photonic crystal slabs (Fig. 3C). Owing to the period of photonic crystal slabs, the vortex generators have no center of structure to align at the incident beam center (*71*). Although the BIC or quasi-BIC of photonic crystal slab is robust to the global change, it is extremely sensitive to symmetric breaking for BIC at the Γ point. Taking into account the symmetry requirement, it was demonstrated that a perovskite-based vortex microlaser can realize ultrafast all-optical switching at room temperature (Fig. 3D) (*72*). The far-field pattern of the emission light from BIC mode can switch from the donut-shaped

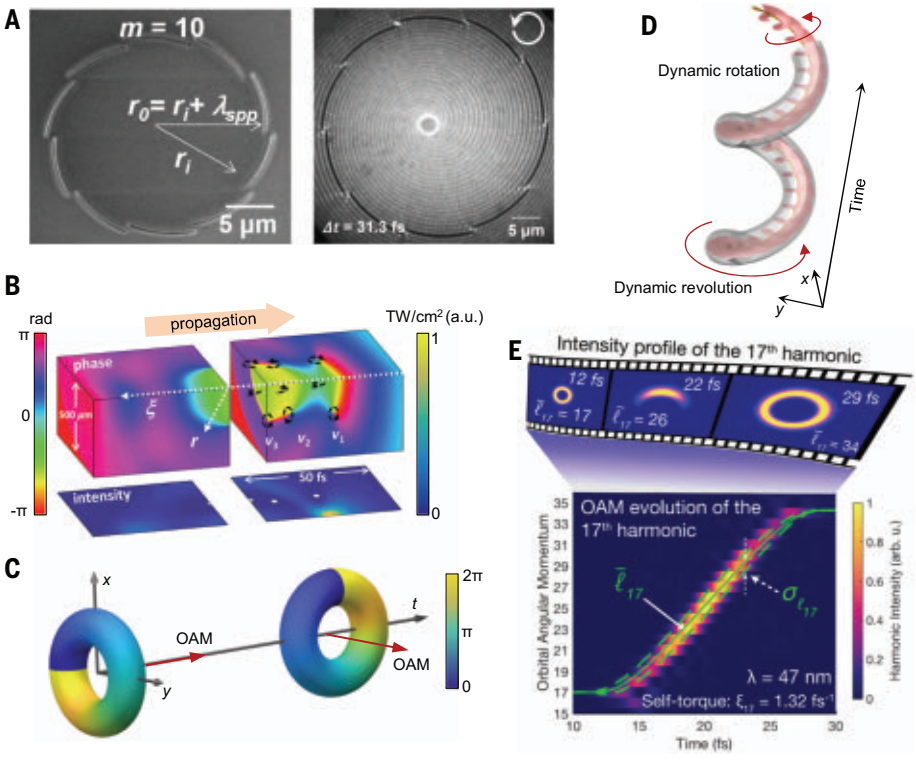


Fig. 4. Spatiotemporal vortex generation. (**A**) Subfemtosecond dynamics of nanoplasmonic vortices. (Left) SEM image of a plasmonic vortex generator with 10 spirals milled in a gold layer. (Right) Experimental snapshot from the plasmonic-vortex generator by means of time-resolved two-photon photoemission electron microscopy. The white arrow indicates the incident right-handed circularly polarized light. [Adapted with permission from (*36*)] (**B**) Phase and intensity projections of simulated spatiotemporal pulse with transverse OAM. Black dashed arrows indicate phase singularities. [Adapted with permission from (*84*)] (**C**) 3D profile of a spatiotemporal vortex pulse with the longitudinal or transverse OAM (*85*). The color bar describes the azimuthal phase distribution. (**D**) Illustration of a spatiotemporal vortex beam exhibiting both dynamic rotation and revolution (*89*). (**E**) Generation of extreme-ultraviolet vortex with self-torque. (Top) Predicted evolution of the intensity distribution of the 17th harmonic at three instants in time during the emission process. (Bottom) Temporal evolution of the OAM of the 17th harmonic. [Adapted with permission from (*21*)]

profiles to two side lobes after breaking the spatial distribution symmetry. The microlaser can switch from a vortex beam to a regular linearly polarized beam in ~1.5 ps. By changing the symmetry and scale of unit cells, vortex beams with high-order topological charges can also be realized at visible and near-infrared operating wavelengths (*71*).

Considering the in-plane phase shift in a topological interface between two topologically dissimilar photonic structures (*73*), a spin-momentum–locked topological edge mode can yield vortex patterns in momentum space (Fig. 3E). A high-performance vortex laser with a single mode operation, narrow emission linewidth, and vertical emission directionality was observed with the out-of-plane lasing radiation. In momentum space, the high-quality vortex beam was observed to be robust against the shape of microcavity. Analogous to the pseudospin in graphene, it was reported that vortex beams can be observed by selectively exciting each of two sublattices of photonic graphene, where the degeneracy in honeycomb lattice is broken by the two inequivalent atomic sites (Fig. 3F) (*74*). The sign of topological charge was flipped as the probe beam moved from one sublattice to another, resulting in a momentum–to–real-space mapping of topological phase singularity (*75*).

Spatiotemporal phase singularities of vortex beams

Optical vortices have been heavily investigated by means of monochromatic waves that have phase singularities purely in the spatial domain. Recently, vortex beams with specific characteristics of ultrafast, transverse, and time-varying OAM were proposed in the spatiotemporal domain, in which the ultrafast light pulses were shaped simultaneously in space and time. For example, it was reported that the detailed spatiotemporal evolution of nanoscale vortices was realized by using time-resolved two-photon photoemission electron microscopy (*36*). The ultrafast plasmonic vortex is generated on a metallic plate with Archimedean-spiral–shaped slits illuminated by a circularly polarized ultrashort pulse (Fig. 4A). The nanoscale vortices can be confined to deep subwavelength scales of about $\lambda_{SPP}/4$ and subfemtosecond time-step resolution. The OAM values of subfemtosecond vortices can be directly accessed by measuring the angular velocity of rotating vortex lobes. Furthermore, the SAM components of light interacting with the Archimedean coupling structures form a plasmonic vortex wave, creating a meron-like quasiparticle of topological plasmonic spin texture (*76*). By using an optically excited femtosecond chiral plasmonic in the near field, it was also reported that an electron vortex beam can be generated and dynamically controlled (*77*).

Generally, the angular momentum of light is longitudinal with respect to the propagation direction. However, recent advances in photonic SAM have led to the notion of electromagnetic fields carrying angular momenta transverse to the direction of motion—transverse SAM (*78*–*80*)—which enables numerous applications in nano-optics, near-field microscopy, and chiral nanophotonic interfaces (*20*, *81*). Analogous to the transverse SAM, spatially transverse OAM can be obtained with tightly focused vortex segmented beams (*82*). Theoretically, a transverse OAM was also expected in the spatiotemporal domain (*83*). In the nonlinear interaction between extremely high-power pulse and air, a spatiotemporal vortex with transverse OAM was experimentally found in the meridional plane with phase and energy circulation (Fig. 4B) (*84*). However, the produced transverse OAM accounts for a very small proportion of the propagating pulse. Until recently, the spatiotemporal optical vortex with a controllable purely transverse OAM is realized in a three-dimensional (3D) wave packet (Fig. 4C) (*85*, *86*). The spatiotemporal vortex is generated by applying a helical phase in the space-time domain. In contrast to the transverse SAM (*87*), the magnitude of the transverse OAM carried by the spatiotemporal vortex is scalable to a larger value through simple adjustments of the topological charge of input vortex beam. Although the spatial size of generated spatiotemporal vortex with transverse OAM cannot be directly shrunk to the micro- or nanoscale by use of a high-numerical-aperture objective lens owing to the lack of temporal focusing, subwavelength transverse OAM can be obtained with a preconditioned spatiotemporal vortex beam (*88*). Recently, spatiotemporal vortex beams that combine dynamic rotation and revolution of OAM by coherently adding multiple frequency comb lines demonstrate high mode purity (Fig. 4D) (*89*). The controllable generation of spatiotemporal optical vortices may promote applications in free-space optical communications and nonlinear wave propagation (*85*).

The conventional continuous-wave vortex beam typically has a static topological

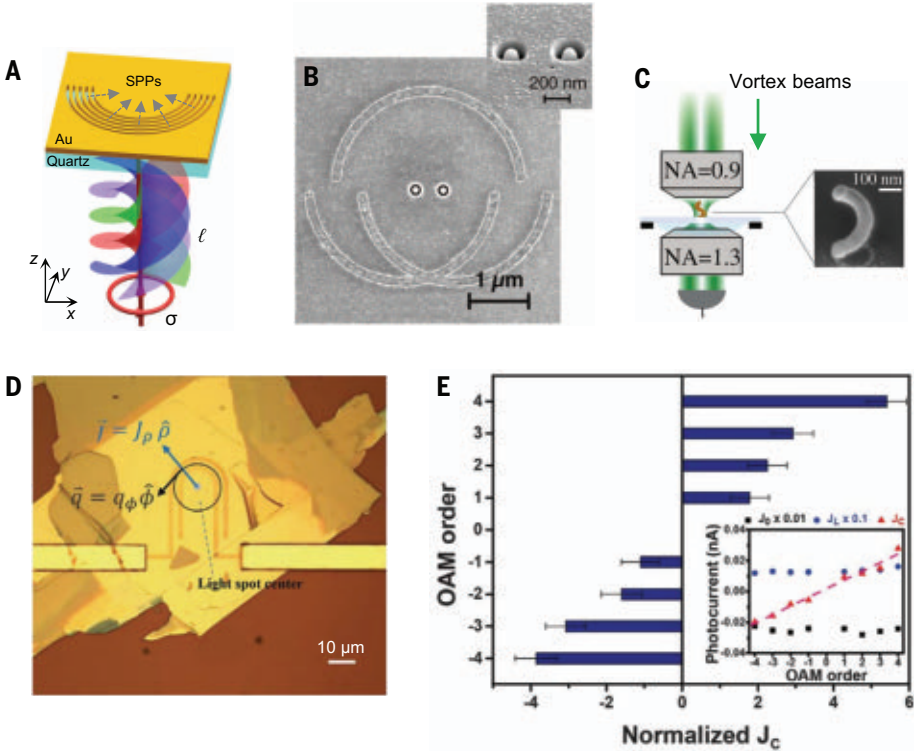


Fig. 5. OAM detectors in nanophotonics. (**A**) Schematic of sorting multiplexed vortex beams by use of nanoslits (*92*). (**B**) SEM images of the double nanoring slits for discriminating photonic SAM and OAM simultaneously. [Adapted with permission from (*94*)]. (**C**) 3D chiral nanostructure for distinguishing the handedness of incident vortex beams. [Adapted with permission from (*98*)] (**D** and **E**) Discrimination of OAM by means of photocurrent measurements. (D) Optical image of a photo-detector device with U-shaped electrodes on WTe_2. (E) Normalized photocurrent that switches with circular polarization for vortex beams with OAM order ranging from −4 to 4. [(D) and (E) adapted with permission from (*104*)]

charge ℓ. However, spatiotemporal OAM beams with some disruptive characteristics, such as time-varying OAM values, have also been realized by means of ultrafast laser. For example, it was reported that the self-torque of light manifests as a temporal OAM variation along a pulse (Fig. 4E) (*21*). The extreme-ultraviolet self-torqued beams arise in a nonlinear process of high-harmonic generation driven by two ultrafast time-delayed laser pulses with different OAM. The high-harmonic generation of vortex beams contains a time-varying OAM along the extreme-ultraviolet pulses. The inherent properties of vortex beams open additional routes for light-matter interaction at femtosecond or attosecond time and nanometer spatial scales. Furthermore, combined with nanophotonics, spatiotemporal vortex beams are expected to bring abundant research opportunities in areas such as quantum entanglement, photonic spin-orbit coupling effect, and new optoelectronic devices (*85*).

OAM detectors in compact devices

Optical vortex with discrete unlimited states has emerged as a physically orthogonal multiplexing method in high-capacity optical information technologies. However, typical bulky elements used for information retrieval from the overall diffracted field, based on the interference method, impose a fundamental limit toward realizing on-chip communication systems (*90*, *91*). By using the interaction between light and plasmon in metal surfaces, it was reported that a semi-ring plasmonic nanoslit can directly and spatially sort various OAM modes (Fig. 5A) (*92*). The spiral phase–modulated surface plasmon polaritons (SPPs) excited by a vortex beam can be directly and spatially classified by the mode spacing of ~120 nm. The disadvantage of this method lies in the lack of external coupling mechanism of SPPs' transfer to the far-field detector, and the observation of linear OAM dispersive plasma needs to be characterized with a near-field scanning optical microscope (*92*). To directly measure the angular momentum–related spatial displacement in the far-field region, holographic plasmonic interfaces provide a powerful strategy to detect the OAM eigenstate of the incident beam, which is coupled to a particular SPP mode (*93*). Moreover, with mode-sorting nanoring slits (*94*, *95*), SAM and OAM of broadband light exhibit efficient parallel multiplexing with noninterference over a bandwidth of 150 nm in the visible wavelength range (Fig. 5B). The reported OAM multiplexing allows lensless reconstruction of a range of distinctive OAM-dependent holographic images. By resolving the splitting angle, the plasma grating structures effectively distinguish SAM and OAM mode with two separated propagating SPPs (*96*, *97*).

Through chiral light-matter interactions with a chiral dipolar nanostructure (*98*, *99*), the handedness of incident vortex beams can be distinguished with chiroptical responses (Fig. 5C). For tightly focused vortex beams, the vectorial electric field distributions can be probed and reconstructed with a single spherical gold nanoparticle (*50*, *100*). By contrast, vortex beams can probe the phase properties of metasurfaces, introduced by small and spatially inhomogeneous nanostructures (*101*). Furthermore, the OAM information can be embedded in a holographic metasurface, which needs corresponding OAM beams to extract the original information, demonstrating an optical encryption (*102*). Intriguingly, considering the principle of reciprocity, the OAM generators may also be used to receive the OAM signals through spin-orbit coupling effects (*40*, *103*).

Recently, it was reported that different OAM modes can also be directly detected with photocurrent to extract the phase information of electromagnetic fields (Fig. 5D) (*104*). Such a direct electrical readout of OAM is reflected by the direction and amplitude of photocurrent on 2D materials (Fig. 5E). The photocurrent generation from vector OAM beams has been demonstrated in a high-order Poincaré sphere, implying the sensitivity of an orbital photogalvanic effect on both SAM and OAM states of optical field. The on-chip vortex detectors for an OAM beam offer the possibility of ultrahigh-capacity and miniaturized nanophotonic devices harnessing angular momentum division. For polarization-coupled spatial modes, it was reported that the OAM states of light can be directly discriminated by human eyes with the interaction polarization topologies and the radially symmetric dichroic elements in the macula (*105*). These studies unlock a new dimension for studying the interaction between the OAM of light and matter, and the advanced OAM-selection rule is expected to provide new physical insights and detection paradigms.

Transmission of vortex beams

The stable transmission of vortex beams has substantial impacts in practical applications of optical micromanipulation, microscopic imaging, and optical communications. As a subset of spatial-division multiplexing in optical communications, mode-division multiplexing with different orthogonal spatial modes has attracted much attention for increasing the transmission capacity (*14*, *17*). Among the mode-division

Table 1. Comparison of vortex nanodevices. *d*, diameter; *w*, width; *h*, height; *p*, period.

Dimensional type	Devices	Material	Size (μm): rings: *d*/*w*/*h*, photonic crystal: *d*/*p*/*h*, other: *h*	Operating wavelength (nm)	OAM values	Reference
Real space	Microring	InGaAsP	9/1.1/1.5	1474	−1	(*43*)
Real space	Microring	Silicon	15/0.5/0.22	1470 to 1580	0, ±1, ±2, ±3, ±4	(*18*)
Real space	Microring	InGaAsP	7/0.65/0.2	1492.6	0, ±1, ±2	(*46*)
Real space	Microring	GsAs	10/0.3/0.16	890 to 920	4, 5, 6, 7, 8*	(*47*)
Real space	Metasurface	TiO_2	0.6	532	10, 100	(*33*)
Real space	Metasurface	TiO_2	0.6	532	3, 4	(*52*)
Real space	Nanoparticle	Silicon	0.176	640, 710	±1	(*65*)
Momentum space	Photonic crystal	Perovskite	0.21/0.28/0.22	552	1	(*72*)
Momentum space	Photonic crystal	Si_3N_4	0.28/0.38/0.12	532	±2	(*71*)
Momentum space	Photonic crystal	Si_3N_4	0.53/0.85/0.1	923	±4	(*71*)
Spatiotemporal	Spiral slits	Au	0.2	800†	10	(*36*)
Spatiotemporal	Spiral slits	Ag	0.1	550†	3/2	(*76*)
Real space	Ring slits	Au	0.2	633	1, 2, 3, 4, 5	(*92*)
Real space	Ring slits	Au	0.2	580, 640, 700	±2, ±4	(*94*)
Real space	Photocurrent	WTe_2	~0.05 to 0.2	1000	±1, ±2, ±3, ±4	(*104*)
Real space	Metasurface	Ag	0.06	633	±1, ±2, ±3	(*97*)
Real space	Nanohelix	Au	0.23	1350 to 1550	±1	(*98*)
Real space	Nanosieves	Au	0.12	445, 532, 633	+13, −39, +52, −91	(*149*)

*Superposition states of OAMs with different topological charges. †The central wavelength of ultrafast laser pulse.

multiplexing, the unlimited OAM modes are considered as a promising basis for a future communication system. In particular, the OAM multiplexing and demultiplexing are compatible with current techniques such as wavelength-division multiplexing and polarization-division multiplexing to further multiply capacity of optical communications (*17*). Moreover, OAM modes are conveniently distinguished by only recovering a full central circle, getting rid of the beam's full radial extent (*106*). To realize OAM-related optical communications, the transmission of vortex beams in free space has made great progress (*12*, *16*, *107*, *108*). However, the vortex beams propagating in free space are affected by atmospheric turbulence, resulting in phase distortion and cross-talk between different OAM modes (*107*–*113*). The propagation properties of vortex beams are characterized in nanocolloidal media to improve the structural stability (*111*–*113*). By virtue of the integration with nanodevices, transmission of vortex beams has also been developed in optical fibers and on-chip waveguides (*17*).

Because the conventional fibers are considered unsuitable for stably supporting the transmission of OAM beams, fibers with ring-shaped index profiles have been demonstrated to have vortex stability for enhancing data capacity through OAM mode-division multiplexing (*17*, *114*). Recently, it was also reported that helically twisted photonic crystal fibers can cause the propagation constants of left- and right-handed optical vortices to be non-degenerate for the same magnitude of OAM modes (*115*). To improve the transmission capacity and miniaturize the fiber-optic communication system, it will be important to directly realize fiber vector eigenmode multiplexing and generation of fiber-eigenmode–like polarization vortices (radially or azimuthally polarized beams) by using photonic integrated devices. For example, by seeding an integrated vector vortex transmitter to an optical fiber, low-level mode cross-talk has been demonstrated for a long transmission length of 2 km in a large-core fiber (*116*).

Inspired by the research on the propagation of a vortex beam in optical fibers, the optical-vortex transmission in an on-chip waveguide can make the system more compact for quantum-information processing. For example, in a rectangular waveguide, a linearly polarized incident beam can produce a vector beam in which the transverse component is circularly polarized and the longitudinal component is phase vortices (*117*). In addition, OAM beams were reported to generate and stably propagate in hybrid plasmonic waveguides (*118*) and trench silicon waveguides (*38*). These produced OAM modes in rectangular waveguides generally have homogeneous polarization states. Some efforts have also been made to prepare the waveguide with cylindrical symmetric cross-section for the propagation of the vector vortex beam. Recently, such a waveguide by means of laser direct writing technology was demonstrated to generate a vector vortex beam inside a photonic chip (*37*). The vortex beams of different OAM modes can be coupled into the ring waveguide embedded in the photonic chip to achieve well-preserved output states. The on-chip waveguide has been proved to be effective even at the single-photon level, which can promote applications in the fields of classical optical communications and integrated quantum information technologies.

Further perspectives and outlook

The peculiar features of phase singularities, associated with orthogonal OAM channels and donut-shaped intensity profiles, enable exciting opportunities for a wide range of applications in optical communications and quantum information processing. To satisfy the specific demand of vortex sources, the phase singularities have been considered in real space, momentum space, and the spatiotemporal domain associated with nanophotonics (Table 1). In practice, the misalignment between OAM devices and light beams results in decreased purity of OAM modes and increased signal cross-talk for generation and detection, respectively. Generated by phase singularities in momentum space, the vortex beams can get rid of the optical misaligning problem for nanoscale devices, which could lead to robust engineering applications in both on-chip generation and detection. Moreover, dynamic switching between different OAM modes in time not only enables high-density data transmission but also further increases the security of wired and wireless communication networks (*119*–*121*). Therefore, it is of great value to analyze how the on-chip micro- and nanostructures can be used for generating and detecting optical phase singularities in multiple dimensions such as ultrafast, time-varying, and

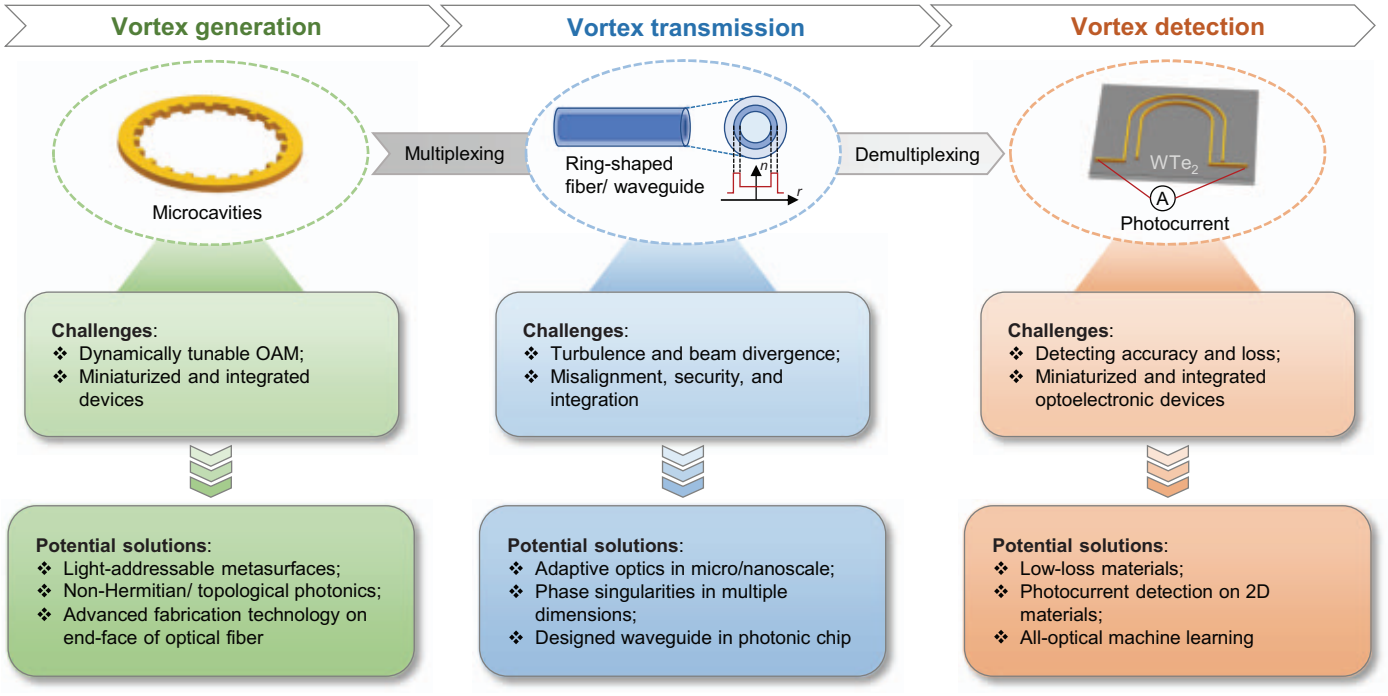


Fig. 6. Challenges and solutions in miniaturized vortex devices for optical communications and quantum information processing.

nonlinear OAM. These new dimensions of phase singularities can be exploited to improve the quality and controllability of vortex beams, which is crucial in the practical applications of high-capacity optical communications and high-dimensional quantum information processing.

The rapid extension of vortex beams puts forward considerable challenges for the design and fabrication of vortex devices, whereas it can also provide more opportunities for applications in chiroptical spectroscopy (*10*, *65*, *98*, *122*), acoustic vortices (*123*–*126*), micro and nanofabrication (*127*–*129*), nonlinear optics (*130*–*135*), and information security both in classical and quantum regimes (*136*, *137*). Multiple challenges need to be overcome to miniaturize the vortex devices for integrating on-chip optical communications and quantum information processing (Fig. 6). To achieve this goal, we believe the following aspects should continue to be promoted. For vortex generation, compact, flexible, and functionalized micro and nanodevices, such as light-addressable metasurfaces (*138*) and non-Hermitian topological photonics (*46*, *139*), are required to generate dynamically tunable OAM modes without bulky components and much energy dissipation. The development of micro- and nanofabrication technology may bring more possibilities for the integration of nanoscale devices such as 3D micro- and nanofabrication technology on the end-face of optical fibers (*140*). To decrease the crosstalk between OAM modes from atmospheric turbulence, misalignment, and beam divergence, vortex beams with high-order topological charges can propagate in optical fibers or on-chip waveguides, compensated by adaptive optics. Moreover, the phase singularities proposed in more dimensions can provide sufficient variances of OAM modes such as vectorial, ultraviolet, nonlinear, and time-varying vortex beams to enhance communication capacity and security (*141*). On the receiving end, the development of micro- and nanoscale optoelectronic devices, distinguishing OAM modes by SPPs on low-loss materials (*142*) or photocurrent detection on varying 2D materials (*104*), may have important practical relevance for the next-generation on-chip photonic circuits and optical communications.

Last, it is worth mentioning the influence of computational tools and artificial intelligence on the research of optical vortices. The optimization algorithm with machine learning offers an efficient mean to design nanodevices (*143*). For example, using a global optimization algorithm, an on-chip OAM emitter with designed nanoscale features can be applied in optical communications (*53*). By using convolutional neural networks and principal components analysis, vector vortex beams can also be identified and classified specifically (*144*, *145*). This concept can provide a deeper understanding of the underlying geometrical properties of the experimentally generated states, without requiring prior knowledge about the physics of the generation apparatus. Also, identification of the high-dimensional mode indices in OAM holography for free-space optical communication has been demonstrated with convolutional neural networks based on computers and proposed with optical diffractive neural networks (*146*). In fact, optical computation process can be directly implemented with optical vortices. The OAM states can potentially be used as a new dimension for enhancing the density of connections of artificial neuronal networks—the computational power as well as the optical inference power (*147*). Moreover, the realization of multiplexing holographic images with photonic OAM provides the new horizon for holography-based artificial neural networks (*148*). We envision that these technologies are a powerful tool, from automatic design to experimental characterization, and can potentially promote the application of vortex beams in classical and quantum information technologies.

REFERENCES AND NOTES

1. D. L. Andrews, M. Babiker, *The Angular Momentum of Light* (Cambridge Univ. Press, 2012).
2. A. M. Yao, M. J. Padgett, Orbital angular momentum: Origins, behavior and applications. *Adv. Opt. Photonics* **3**, 161–204 (2011). doi: 10.1364/AOP.3.000161
3. J. H. Poynting, The wave motion of a revolving shaft, and a suggestion as to the angular momentum in a beam of circularly polarised light. *Proc. R. Soc. Lond., A Contain. Pap. Math. Phys. Character* **82**, 560–567 (1909). doi: 10.1098/rspa.1909.0060
4. L. Allen, M. W. Beijersbergen, R. J. C. Spreeuw, J. P. Woerdman, Orbital angular momentum of light and the transformation of Laguerre-Gaussian laser modes. *Phys. Rev. A* **45**, 8185–8189 (1992). doi: 10.1103/PhysRevA.45.8185; pmid: 9906912
5. L. Ge, Twisted light on a chip. *Science* **368**, 707–708 (2020). doi: 10.1126/science.abb8091
6. D. G. Grier, A revolution in optical manipulation. *Nature* **424**, 810–816 (2003). doi: 10.1038/nature01935; pmid: 12917694
7. T. Arikawa *et al.*, Transfer of orbital angular momentum of light to plasmonic excitations in metamaterials. *Sci. Adv.* **6**, eaay1977 (2020). doi: 10.1126/sciadv.aay1977; pmid: 32582843

8. K. I. Willig, S. O. Rizzoli, V. Westphal, R. Jahn, S. W. Hell, STED microscopy reveals that synaptotagmin remains clustered after synaptic vesicle exocytosis. *Nature* **440**, 935–939 (2006). doi: 10.1038/nature04592; pmid: 16612384
9. M. P. J. Lavery, F. C. Speirits, S. M. Barnett, M. J. Padgett, Detection of a spinning object using light's orbital angular momentum. *Science* **341**, 537–540 (2013). doi: 10.1126/science.1239936; pmid: 23908234
10. J. Mun *et al*., Electromagnetic chirality: From fundamentals to nontraditional chiroptical phenomena. *Light Sci. Appl.* **9**, 139 (2020). doi: 10.1038/s41377-020-00367-8
11. W. Brullot, M. K. Vanbel, T. Swusten, T. Verbiest, Resolving enantiomers using the optical angular momentum of twisted light. *Sci. Adv.* **2**, e1501349 (2016). doi: 10.1126/sciadv.1501349; pmid: 26998517
12. J. Wang *et al*., Terabit free-space data transmission employing orbital angular momentum multiplexing. *Nat. Photonics* **6**, 488–496 (2012). doi: 10.1038/nphoton.2012.138
13. T. Lei *et al*., Massive individual orbital angular momentum channels for multiplexing enabled by Dammann gratings. *Light Sci. Appl.* **4**, e257–e257 (2015). doi: 10.1038/lsa.2015.30
14. J. Wang, Advances in communications using optical vortices. *Photon. Res.* **4**, B14–B28 (2016). doi: 10.1364/PRJ.4.000B14
15. A. Mair, A. Vaziri, G. Weihs, A. Zeilinger, Entanglement of the orbital angular momentum states of photons. *Nature* **412**, 313–316 (2001). doi: 10.1038/35085529; pmid: 11460157
16. G. Gibson *et al*., Free-space information transfer using light beams carrying orbital angular momentum. *Opt. Express* **12**, 5448–5456 (2004). doi: 10.1364/OPEX.12.005448; pmid: 19484105
17. N. Bozinovic *et al*., Terabit-scale orbital angular momentum mode division multiplexing in fibers. *Science* **340**, 1545–1548 (2013). doi: 10.1126/science.1237861; pmid: 23812709
18. X. Cai *et al*., Integrated compact optical vortex beam emitters. *Science* **338**, 363–366 (2012). doi: 10.1126/science.1226528; pmid: 23087243
19. H. Rubinsztein-Dunlop *et al*., Roadmap on structured light. *J. Opt.* **19**, 013001 (2016). doi: 10.1088/2040-8978/19/1/013001
20. K. Y. Bliokh, F. J. Rodriguez-Fortuno, F. Nori, A. V. Zayats, Spin-orbit interactions of light. *Nat. Photonics* **9**, 796–808 (2015). doi: 10.1038/nphoton.2015.201
21. L. Rego *et al*., Generation of extreme-ultraviolet beams with time-varying orbital angular momentum. *Science* **364**, eaaw9486 (2019). doi: 10.1126/science.aaw9486; pmid: 31249031
22. J. Ni *et al*., Three-dimensional chiral microstructures fabricated by structured optical vortices in isotropic material. *Light Sci. Appl.* **6**, e17011 (2017). doi: 10.1038/lsa.2017.11; pmid: 30167269
23. Y. Shen *et al*., Optical vortices 30 years on: OAM manipulation from topological charge to multiple singularities. *Light Sci. Appl.* **8**, 90 (2019). doi: 10.1038/s41377-019-0194-2; pmid: 31645934
24. X. Wang *et al*., Recent advances on optical vortex generation. *Nanophotonics* **7**, 1533–1556 (2018). doi: 10.1515/nanoph-2018-0072
25. M. Erhard, R. Fickler, M. Krenn, A. Zeilinger, Twisted photons: New quantum perspectives in high dimensions. *Light Sci. Appl.* **7**, 17146–17146 (2018). doi: 10.1038/lsa.2017.146; pmid: 30839541
26. G. Biener, A. Niv, V. Kleiner, E. Hasman, Formation of helical beams by use of Pancharatnam-Berry phase optical elements. *Opt. Lett.* **27**, 1875–1877 (2002). doi: 10.1364/OL.27.001875; pmid: 18033387
27. L. Marrucci, C. Manzo, D. Paparo, Optical spin-to-orbital angular momentum conversion in inhomogeneous anisotropic media. *Phys. Rev. Lett.* **96**, 163905 (2006). doi: 10.1103/PhysRevLett.96.163905; pmid: 16712234
28. N. Yu *et al*., Light propagation with phase discontinuities: Generalized laws of reflection and refraction. *Science* **334**, 333–337 (2011). doi: 10.1126/science.1210713; pmid: 21885733
29. M. Q. Mehmood *et al*., Visible-frequency metasurface for structuring and spatially multiplexing optical vortices. *Adv. Mater.* **28**, 2533–2539 (2016). doi: 10.1002/adma.201504532; pmid: 26833667
30. E. Maguid *et al*., Photonic spin-controlled multifunctional shared-aperture antenna array. *Science* **352**, 1202–1206 (2016). doi: 10.1126/science.aaf3417; pmid: 27103668
31. D. Naidoo *et al*., Controlled generation of higher-order Poincaré sphere beams from a laser. *Nat. Photonics* **10**, 327–332 (2016). doi: 10.1038/nphoton.2016.37
32. Q. Zhan, Cylindrical vector beams: From mathematical concepts to applications. *Adv. Opt. Photonics* **1**, 1–57 (2009). doi: 10.1364/AOP.1.000001
33. H. Sroor *et al*., High-purity orbital angular momentum states from a visible metasurface laser. *Nat. Photonics* **14**, 498–503 (2020). doi: 10.1038/s41566-020-0623-z
34. B. Zhen, C. W. Hsu, L. Lu, A. D. Stone, M. Soljačić, Topological nature of optical bound states in the continuum. *Phys. Rev. Lett.* **113**, 257401 (2014). doi: 10.1103/PhysRevLett.113.257401; pmid: 25554906
35. Y. Gorodetski, A. Niv, V. Kleiner, E. Hasman, Observation of the spin-based plasmonic effect in nanoscale structures. *Phys. Rev. Lett.* **101**, 043903 (2008). doi: 10.1103/PhysRevLett.101.043903; pmid: 18764330
36. G. Spektor *et al*., Revealing the subfemtosecond dynamics of orbital angular momentum in nanoplasmonic vortices. *Science* **355**, 1187–1191 (2017). doi: 10.1126/science.aaj1699; pmid: 28302854
37. Y. Chen *et al*., Vector vortex beam emitter embedded in a photonic chip. *Phys. Rev. Lett.* **124**, 153601 (2020). doi: 10.1103/PhysRevLett.124.153601; pmid: 32357035
38. S. Zheng, J. Wang, On-chip orbital angular momentum modes generator and (de)multiplexer based on trench silicon waveguides. *Opt. Express* **25**, 18492–18501 (2017). doi: 10.1364/OE.25.018492; pmid: 28789334
39. C.-W. Qiu, Y. Yang, Vortex generation reaches a new plateau. *Science* **357**, 645–645 (2017). doi: 10.1126/science.aan6359; pmid: 28818929
40. Z. Shao, J. Zhu, Y. Chen, Y. Zhang, S. Yu, Spin-orbit interaction of light induced by transverse spin angular momentum engineering. *Nat. Commun.* **9**, 926 (2018). doi: 10.1038/s41467-018-03237-5; pmid: 29500340
41. J. Zhang *et al*., An InP-based vortex beam emitter with monolithically integrated laser. *Nat. Commun.* **9**, 2652 (2018). doi: 10.1038/s41467-018-05170-z; pmid: 29985405
42. K. G. Cognée, H. M. Doeleman, P. Lalanne, A. F. Koenderink, Generation of pure OAM beams with a single state of polarization by antenna-decorated microdisk resonators. *ACS Photonics* **7**, 3049–3060 (2020). doi: 10.1021/acsphotonics.0c01081
43. P. Miao *et al*., Orbital angular momentum microlaser. *Science* **353**, 464–467 (2016). doi: 10.1126/science.aaf8533; pmid: 27471299
44. N. Carlon Zambon *et al*., Optically controlling the emission chirality of microlasers. *Nat. Photonics* **13**, 283–288 (2019). doi: 10.1038/s41566-019-0380-z
45. B. Bahari *et al*., Photonic quantum Hall effect and multiplexed light sources of large orbital angular momenta. *Nat. Phys.* **17**, 700–703 (2021). doi: 10.1038/s41567-021-01165-8
46. Z. Zhang *et al*., Tunable topological charge vortex microlaser. *Science* **368**, 760–763 (2020). doi: 10.1126/science.aba8996; pmid: 32409473
47. B. Chen *et al*., Bright solid-state sources for single photons with orbital angular momentum. *Nat. Nanotechnol.* **16**, 302–307 (2021). doi: 10.1038/s41565-020-00827-7; pmid: 33432207
48. G. Knöner *et al*., Integrated optomechanical microelements. *Opt. Express* **15**, 5521–5530 (2007). doi: 10.1364/OE.15.005521; pmid: 19532808
49. E. Brasselet, M. Malinauskas, A. Žukauskas, S. Juodkazis, Photopolymerized microscopic vortex beam generators: Precise delivery of optical orbital angular momentum. *Appl. Phys. Lett.* **97**, 211108 (2010). doi: 10.1063/1.3517519
50. T. Bauer *et al*., Optics. Observation of optical polarization Möbius strips. *Science* **347**, 964–966 (2015). doi: 10.1126/science.1260635; pmid: 25636796
51. E. Karimi *et al*., Generating optical orbital angular momentum at visible wavelengths using a plasmonic metasurface. *Light Sci. Appl.* **3**, e167–e167 (2014). doi: 10.1038/lsa.2014.48
52. R. C. Devlin, A. Ambrosio, N. A. Rubin, J. P. B. Mueller, F. Capasso, Arbitrary spin-to-orbital angular momentum conversion of light. *Science* **358**, 896–901 (2017). doi: 10.1126/science.aao5392; pmid: 29097490
53. Z. Xie *et al*., Ultra-broadband on-chip twisted light emitter for optical communications. *Light Sci. Appl.* **7**, 18001 (2018). doi: 10.1038/lsa.2018.1; pmid: 30839547
54. W. Sun *et al*., Lead halide perovskite vortex microlasers. *Nat. Commun.* **11**, 4862 (2020). pmid: 32978397
55. M. I. Shalaev *et al*., High-efficiency all-dielectric metasurfaces for ultracompact beam manipulation in transmission mode. *Nano Lett.* **15**, 6261–6266 (2015). doi: 10.1021/acs.nanolett.5b02926; pmid: 26280735
56. J. Sun, J. Zeng, N. M. Litchinitser, Twisting light with hyperbolic metamaterials. *Opt. Express* **21**, 14975–14981 (2013). doi: 10.1364/OE.21.014975; pmid: 23787685
57. K. E. Chong *et al*., Polarization-independent silicon metadevices for efficient optical wavefront control. *Nano Lett.* **15**, 5369–5374 (2015). doi: 10.1021/acs.nanolett.5b01752; pmid: 26192100
58. J. Sun *et al*., Spinning light on the nanoscale. *Nano Lett.* **14**, 2726–2729 (2014). doi: 10.1021/nl500658n; pmid: 24697576
59. S. Franke-Arnold, S. M. Barnett, M. J. Padgett, L. Allen, Two-photon entanglement of orbital angular momentum states. *Phys. Rev. A* **65**, 033823 (2002). doi: 10.1103/PhysRevA.65.033823
60. R. F. Offer, D. Stulga, E. Riis, S. Franke-Arnold, A. S. Arnold, Spiral bandwidth of four-wave mixing in Rb vapour. *Commun. Phys.* **1**, 84 (2018). doi: 10.1038/s42005-018-0077-5
61. T. Stav *et al*., Quantum entanglement of the spin and orbital angular momentum of photons using metamaterials. *Science* **361**, 1101–1104 (2018). doi: 10.1126/science.aat9042; pmid: 30213909
62. A. I. Kuznetsov, A. E. Miroshnichenko, M. L. Brongersma, Y. S. Kivshar, B. Luk'yanchuk, Optically resonant dielectric nanostructures. *Science* **354**, aag2472 (2016). doi: 10.1126/science.aag2472; pmid: 27856851
63. L. Jin *et al*., Dielectric multi-momentum meta-transformer in the visible. *Nat. Commun.* **10**, 4789 (2019). doi: 10.1038/s41467-019-12637-0; pmid: 31636259
64. H. Ren *et al*., Metasurface orbital angular momentum holography. *Nat. Commun.* **10**, 2986 (2019). doi: 10.1038/s41467-019-11030-1; pmid: 31324755
65. J. S. Eismann, M. Neugebauer, P. Banzer, Exciting a chiral dipole moment in an achiral nanostructure. *Optica* **5**, 954–959 (2018). doi: 10.1364/OPTICA.5.000954
66. S. Nechayev, J. S. Eismann, G. Leuchs, P. Banzer, Orbital-to-spin angular momentum conversion employing local helicity. *Phys. Rev. B* **99**, 075155 (2019). doi: 10.1103/PhysRevB.99.075155
67. M. M. Coles, M. D. Williams, K. Saadi, D. S. Bradshaw, D. L. Andrews, Chiral nanoemitter array: A launchpad for optical vortices. *Laser Photonics Rev.* **7**, 1088–1092 (2013). doi: 10.1002/lpor.201300117
68. M. Notomi, Topology in momentum space becomes real. *Nat. Photonics* **14**, 595–596 (2020). doi: 10.1038/s41566-020-0693-y
69. C. W. Hsu, B. Zhen, A. D. Stone, J. D. Joannopoulos, M. Soljačić, Bound states in the continuum. *Nat. Rev. Mater.* **1**, 16048 (2016). doi: 10.1038/natrevmats.2016.48
70. H. M. Doeleman, F. Monticone, W. den Hollander, A. Alù, A. F. Koenderink, Experimental observation of a polarization vortex at an optical bound state in the continuum. *Nat. Photonics* **12**, 397–401 (2018). doi: 10.1038/s41566-018-0177-5
71. B. Wang *et al*., Generating optical vortex beams by momentum-space polarization vortices centred at bound states in the continuum. *Nat. Photonics* **14**, 623–628 (2020). doi: 10.1038/s41566-020-0658-1
72. C. Huang *et al*., Ultrafast control of vortex microlasers. *Science* **367**, 1018–1021 (2020). doi: 10.1126/science.aba4597; pmid: 32108108
73. Z.-Q. Yang, Z.-K. Shao, H.-Z. Chen, X.-R. Mao, R.-M. Ma, Spin-momentum-locked edge mode for topological vortex lasing. *Phys. Rev. Lett.* **125**, 013903 (2020). doi: 10.1103/PhysRevLett.125.013903; pmid: 32678624
74. D. Song *et al*., Unveiling pseudospin and angular momentum in photonic graphene. *Nat. Commun.* **6**, 6272 (2015). doi: 10.1038/ncomms7272; pmid: 25687645
75. X. Liu *et al*., Universal momentum-to-real-space mapping of topological singularities. *Nat. Commun.* **11**, 1586 (2020). doi: 10.1038/s41467-020-15374-x; pmid: 32221307
76. Y. Dai *et al*., Plasmonic topological quasiparticle on the nanometre and femtosecond scales. *Nature* **588**, 616–619 (2020). doi: 10.1038/s41586-020-3030-1; pmid: 33361792
77. G. M. Vanacore *et al*., Ultrafast generation and control of an electron vortex beam via chiral plasmonic near fields. *Nat. Mater.* **18**, 573–579 (2019). doi: 10.1038/s41563-019-0336-1; pmid: 31061485
78. K. Y. Bliokh, F. Nori, Transverse and longitudinal angular momenta of light. *Phys. Rep.* **592**, 1–38 (2015). doi: 10.1016/j.physrep.2015.06.003
79. A. Aiello, N. Lindlein, C. Marquardt, G. Leuchs, Transverse angular momentum and geometric spin Hall effect of light. *Phys. Rev. Lett.* **103**, 100401 (2009). doi: 10.1103/PhysRevLett.103.100401; pmid: 19792286
80. P. Banzer *et al*., The photonic wheel-demonstration of a state of light with purely transverse angular momentum. *J. Eur. Opt. Soc. Rapid Publ.* **8**, 13032 (2013). doi: 10.2971/jeos.2013.13032
81. K. Y. Bliokh, F. Nori, Transverse spin of a surface polariton. *Phys. Rev. A* **85**, 061801 (2012). doi: 10.1103/PhysRevA.85.061801
82. M. N. Neugebauer, A. Aiello, P. Banzer, Linear and angular momenta in tightly focused vortex segmented beams of light. *Chin. Opt. Lett.* **15**, 030003–030007 (2017). doi: 10.3788/COL201715.030003

83. K. Y. Bliokh, F. Nori, Spatiotemporal vortex beams and angular momentum. *Phys. Rev. A* **86**, 033824 (2012). doi: 10.1103/PhysRevA.86.033824
84. N. Jhajj *et al*., Spatiotemporal optical vortices. *Phys. Rev. X* **6**, 031037 (2016). doi: 10.1103/PhysRevX.6.031037
85. A. Chong, C. Wan, J. Chen, Q. Zhan, Generation of spatiotemporal optical vortices with controllable transverse orbital angular momentum. *Nat. Photonics* **14**, 350–354 (2020). doi: 10.1038/s41566-020-0587-z
86. S. Hancock, S. Zahedpour, A. Goffin, H. Milchberg, Free-space propagation of spatiotemporal optical vortices. *Optica* **6**, 1547–1553 (2019). doi: 10.1364/OPTICA.6.001547
87. K. Y. Bliokh, D. Smirnova, F. Nori, OPTICS. Quantum spin Hall effect of light. *Science* **348**, 1448–1451 (2015). doi: 10.1126/science.aaa9519; pmid: 26113717
88. J. Chen, C. Wan, A. Chong, Q. Zhan, Subwavelength focusing of a spatio-temporal wave packet with transverse orbital angular momentum. *Opt. Express* **28**, 18472–18478 (2020). doi: 10.1364/OE.394428; pmid: 32680045
89. Z. Zhao *et al*., Dynamic spatiotemporal beams that combine two independent and controllable orbital-angular-momenta using multiple optical-frequency-comb lines. *Nat. Commun.* **11**, 4099 (2020). doi: 10.1038/s41467-020-17805-1; pmid: 32796838
90. G. C. Berkhout, M. P. Lavery, J. Courtial, M. W. Beijersbergen, M. J. Padgett, Efficient sorting of orbital angular momentum states of light. *Phys. Rev. Lett.* **105**, 153601 (2010). doi: 10.1103/PhysRevLett.105.153601; pmid: 21230900
91. M. Mirhosseini, M. Malik, Z. Shi, R. W. Boyd, Efficient separation of the orbital angular momentum eigenstates of light. *Nat. Commun.* **4**, 2781 (2013). doi: 10.1038/ncomms3781; pmid: 24216691
92. S. Mei *et al*., On-chip discrimination of orbital angular momentum of light with plasmonic nanoslits. *Nanoscale* **8**, 2227–2233 (2016). doi: 10.1039/C5NR07374J; pmid: 26742094
93. P. Genevet, J. Lin, M. A. Kats, F. Capasso, Holographic detection of the orbital angular momentum of light with plasmonic photodiodes. *Nat. Commun.* **3**, 1278 (2012). doi: 10.1038/ncomms2293; pmid: 23232408
94. H. Ren, X. Li, Q. Zhang, M. Gu, On-chip noninterference angular momentum multiplexing of broadband light. *Science* **352**, 805–809 (2016). doi: 10.1126/science.aaf1112; pmid: 27056843
95. Z. Yue, H. Ren, S. Wei, J. Lin, M. Gu, Angular-momentum nanometrology in an ultrathin plasmonic topological insulator film. *Nat. Commun.* **9**, 4413 (2018). doi: 10.1038/s41467-018-06952-1; pmid: 30356063
96. J. Chen, X. Chen, T. Li, S. Zhu, On-chip detection of orbital angular momentum beam by plasmonic nanogratings. *Laser Photonics Rev.* **12**, 1700331 (2018). doi: 10.1002/lpor.201700331
97. F. Feng, G. Si, C. Min, X. Yuan, M. Somekh, On-chip plasmonic spin-Hall nanograting for simultaneously detecting phase and polarization singularities. *Light Sci. Appl.* **9**, 95 (2020). doi: 10.1038/s41377-020-0330-z; pmid: 32528669
98. P. Woźniak, I. De Leon, K. Höflich, G. Leuchs, P. Banzer, Interaction of light carrying orbital angular momentum with a chiral dipolar scatterer. *Optica* **6**, 961–965 (2019). doi: 10.1364/OPTICA.6.000961
99. J. Ni *et al*., Giant helical dichroism of single chiral nanostructures with photonic orbital angular momentum. *ACS Nano* **15**, 2893–2900 (2021). doi: 10.1021/acsnano.0c08941; pmid: 33497201
100. T. Bauer, S. N. Khonina, I. Golub, G. Leuchs, P. Banzer, Ultrafast spinning twisted ribbons of confined electric fields. *Optica* **7**, 1228–1231 (2020). doi: 10.1364/OPTICA.392772
101. Y. Xu, J. Sun, W. Walasik, N. M. Litchinitser, Probing metamaterials with structured light. *Opt. Express* **24**, 26249–26254 (2016). doi: 10.1364/OE.24.026249; pmid: 27857361
102. H. Ren *et al*., Complex-amplitude metasurface-based orbital angular momentum holography in momentum space. *Nat. Nanotechnol.* **15**, 948–955 (2020). doi: 10.1038/s41565-020-0768-4; pmid: 32958936
103. Y. Guo *et al*., Spin-decoupled metasurface for simultaneous detection of spin and orbital angular momenta via momentum transformation. *Light Sci. Appl.* **10**, 63 (2021). doi: 10.1038/s41377-021-00497-7; pmid: 33767137
104. Z. Ji *et al*., Photocurrent detection of the orbital angular momentum of light. *Science* **368**, 763–767 (2020). doi: 10.1126/science.aba9192; pmid: 32409474
105. D. Sarenac *et al*., Direct discrimination of structured light by humans. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 14682–14687 (2020). doi: 10.1073/pnas.1920226117; pmid: 32546523
106. A. E. Willner, OAM light for communications. *Opt. Photonics News* **32**, 34–41 (2021). doi: 10.1364/OPN.32.6.000034
107. M. Krenn *et al*., Twisted light transmission over 143 km. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 13648–13653 (2016). doi: 10.1073/pnas.1612023113; pmid: 27856744
108. M. P. J. Lavery *et al*., Free-space propagation of high-dimensional structured optical fields in an urban environment. *Sci. Adv.* **3**, e1700552 (2017). doi: 10.1126/sciadv.1700552; pmid: 29075663
109. J. B. Götte *et al*., Light beams with fractional orbital angular momentum and their vortex structure. *Opt. Express* **16**, 993–1006 (2008). doi: 10.1364/OE.16.000993; pmid: 18542173
110. W. Li *et al*., Rapidly tunable orbital angular momentum (OAM) system for higher order Bessel beams integrated in time (HOBBIT). *Opt. Express* **27**, 3920–3934 (2019). doi: 10.1364/OE.27.003920; pmid: 30876016
111. S. Z. Silahli, W. Walasik, N. M. Litchinitser, Necklace beam generation in nonlinear colloidal engineered media. *Opt. Lett.* **40**, 5714–5717 (2015). doi: 10.1364/OL.40.005714; pmid: 26670494
112. R. J. Watkins *et al*., Experimental probing of turbulence using a continuous spectrum of asymmetric OAM beams. *Opt. Express* **28**, 924–935 (2020). doi: 10.1364/OE.380405; pmid: 32121812
113. J. Sun *et al*., Nanoscale orbital angular momentum beam instabilities in engineered nonlinear colloidal media. *Opt. Express* **26**, 5118–5125 (2018). doi: 10.1364/OE.26.005118; pmid: 29529719
114. S. Ramachandran, P. Kristensen, Optical vortices in fiber. *Nanophotonics* **2**, 455–474 (2013). doi: 10.1515/nanoph-2013-0047
115. P. S. J. Russell, R. Beravat, G. K. Wong, Helically twisted photonic crystal fibres. *Philos. Trans. A Math. Phys. Eng. Sci.* **375**, 20150440 (2017). pmid: 28069771
116. J. Liu *et al*., Direct fiber vector eigenmode multiplexing transmission seeded by integrated optical vortex emitters. *Light Sci. Appl.* **7**, 17148 (2018). doi: 10.1038/lsa.2017.148; pmid: 30839539
117. G. Maltese *et al*., Towards an integrated AlGaAs waveguide platform for phase and polarisation shaping. *J. Opt.* **20**, 05LT01 (2018). doi: 10.1088/2040-8986/aabc1d
118. Y. Liang, H. W. Wu, B. J. Huang, X. G. Huang, Light beams with selective angular momentum generated by hybrid plasmonic waveguides. *Nanoscale* **6**, 12360–12365 (2014). doi: 10.1039/C4NR03606A; pmid: 25192324
119. X. Fang, H. Ren, M. Gu, Orbital angular momentum holography for high-security encryption. *Nat. Photonics* **14**, 102–108 (2020). doi: 10.1038/s41566-019-0560-x
120. H. Zhou *et al*., Polarization-encrypted orbital angular momentum multiplexed metasurface holography. *ACS Nano* **14**, 5553–5559 (2020). doi: 10.1021/acsnano.9b09814; pmid: 32348122
121. K. Huang *et al*., Spiniform phase-encoded metagratings entangling arbitrary rational-order orbital angular momentum. *Light Sci. Appl.* **7**, 17156 (2018). doi: 10.1038/lsa.2017.156; pmid: 30839520
122. J. Ni *et al*., Gigantic vortical differential scattering as a monochromatic probe for multiscale chiral structures. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2020055118 (2021). doi: 10.1073/pnas.2020055118; pmid: 33372145
123. Y. Fu *et al*., Sound vortex diffraction via topological charge in phase gradient metagratings. *Sci. Adv.* **6**, eaba9876 (2020). doi: 10.1126/sciadv.aba9876; pmid: 33008907
124. Z. Hong, J. Zhang, B. W. Drinkwater, Observation of orbital angular momentum transfer from bessel-shaped acoustic vortices to diphasic liquid-microparticle mixtures. *Phys. Rev. Lett.* **114**, 214301 (2015). doi: 10.1103/PhysRevLett.114.214301; pmid: 26066437
125. C. Shi, M. Dubois, Y. Wang, X. Zhang, High-speed acoustic communication by multiplexing orbital angular momentum. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 7250–7253 (2017). doi: 10.1073/pnas.1704450114; pmid: 28652341
126. S. Wang, G. Ma, C. T. Chan, Topological transport of sound mediated by spin-redirection geometric phase. *Sci. Adv.* **4**, eaaq1475 (2018). doi: 10.1126/sciadv.aaq1475; pmid: 29464210
127. J. C. Ni *et al*., Multifurcate assembly of slanted micropillars fabricated by superposition of optical vortices and application in high-efficiency trapping microparticles. *Adv. Funct. Mater.* **27**, 1701939 (2017). doi: 10.1002/adfm.201701939
128. K. Toyoda *et al*., Transfer of light helicity to nanostructures. *Phys. Rev. Lett.* **110**, 143603 (2013). doi: 10.1103/PhysRevLett.110.143603; pmid: 25166987
129. N. M. Litchinitser, Structured light meets structured matter. *Science* **337**, 1054–1055 (2012). doi: 10.1126/science.1226204; pmid: 22936768
130. Y. Tang *et al*., Harmonic spin–orbit angular momentum cascade in nonlinear optical crystals. *Nat. Photonics* **14**, 658–662 (2020). doi: 10.1038/s41566-020-0691-0
131. D. Gauthier *et al*., Tunable orbital angular momentum in high-harmonic generation. *Nat. Commun.* **8**, 14971 (2017). doi: 10.1038/ncomms14971; pmid: 28378741
132. X. Zhang *et al*., Generation of intense high-order vortex harmonics. *Phys. Rev. Lett.* **114**, 173901 (2015). doi: 10.1103/PhysRevLett.114.173901; pmid: 25978234
133. R. Géneaux *et al*., Synthesis and characterization of attosecond light vortices in the extreme ultraviolet. *Nat. Commun.* **7**, 12583 (2016). doi: 10.1038/ncomms12583; pmid: 27573787
134. T. Roger, J. J. Heitz, E. M. Wright, D. Faccio, Non-collinear interaction of photons with orbital angular momentum. *Sci. Rep.* **3**, 3491 (2013). doi: 10.1038/srep03491; pmid: 24336441
135. M. I. Shalaev, Z. A. Kudyshev, N. M. Litchinitser, Twisted light in a nonlinear mirror. *Opt. Lett.* **38**, 4288–4291 (2013). doi: 10.1364/OL.38.004288; pmid: 24177075
136. N. Zhao, X. Li, G. Li, J. M. Kahn, Capacity limits of spatially multiplexed free-space communication. *Nat. Photonics* **9**, 822–826 (2015). doi: 10.1038/nphoton.2015.214
137. R. Fickler *et al*., Quantum entanglement of high angular momenta. *Science* **338**, 640–643 (2012). doi: 10.1126/science.1227193; pmid: 23118185
138. X. G. Zhang *et al*., An optically driven digital metasurface for programming electromagnetic functions. *Nat. Electron.* **3**, 165–171 (2020). doi: 10.1038/s41928-020-0380-5
139. Z. Chen, M. Segev, Highlighting photonics: Looking into the next decade. *eLight* **1**, 2 (2021). doi: 10.1186/s43593-021-00002-y
140. Y. Hong *et al*., Solvent-free nanofabrication based on ice-assisted electron-beam lithography. *Nano Lett.* **20**, 8841–8846 (2020). doi: 10.1021/acs.nanolett.0c03809; pmid: 33185450
141. F. Kong *et al*., Controlling the orbital angular momentum of high harmonic vortices. *Nat. Commun.* **8**, 14970 (2017). doi: 10.1038/ncomms14970; pmid: 28378823
142. Y. Wang *et al*., Stable, high-performance sodium-based plasmonic devices in the near infrared. *Nature* **581**, 401–405 (2020). doi: 10.1038/s41586-020-2306-9; pmid: 32461649
143. W. Ma *et al*., Deep learning for the design of photonic structures. *Nat. Photonics* **15**, 77–90 (2021). doi: 10.1038/s41566-020-0685-y
144. T. Giordani *et al*., Machine learning-based classification of vector vortex beams. *Phys. Rev. Lett.* **124**, 160401 (2020). doi: 10.1103/PhysRevLett.124.160401; pmid: 32383956
145. H. Ren, W. Shao, Y. Li, F. Salim, M. Gu, Three-dimensional vectorial holography based on machine learning inverse design. *Sci. Adv.* **6**, eaaz4261 (2020). doi: 10.1126/sciadv.aaz4261; pmid: 32494614
146. H. Luan *et al*., 768-ary Laguerre-Gaussian-mode shift keying free-space optical communication based on convolutional neural networks. *Opt. Express* **29**, 19807–19818 (2021). doi: 10.1364/OE.420176; pmid: 34266083
147. E. Goi *et al*., Nanoprinted high-neuron-density optical linear perceptrons performing near-infrared inference on a CMOS chip. *Light Sci. Appl.* **10**, 40 (2021). doi: 10.1038/s41377-021-00483-z; pmid: 33654061
148. X. Lin *et al*., All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018). doi: 10.1126/science.aat8084; pmid: 30049787
149. Z. Lin *et al*., Phyllotaxis-inspired nanosieves with multiplexed orbital angular momentum. *eLight* **1**, 5 (2021). doi: 10.1186/s43593-021-00005-9

ACKNOWLEDGMENTS

We thank the reviewers for their insightful comments that improved the paper. **Funding:** This project was supported by the National Research Foundation, Prime Minister's Office, Singapore under Competitive Research Program Award NRF-CRP22-2019-0006, and the Advanced Research and Technology Innovation Centre (ARTIC), the National University of Singapore (grant R-261–518-004-720). M.G. acknowledges the support from the Zhangjiang National Innovation Demonstration Zone (ZJ2019-ZD-005). Q.S. acknowledges the financial support from National Natural Science Foundation of China (grants 12025402, 11934012, and 61975041) and the Fundamental Research Funds for the Central Universities. Y.K. acknowledges the support by Australian Research Council (grant DP200101168) and Russian Science Foundation (grant 21-72-30018). **Author contributions:** J.N. and C.H. wrote the manuscript; L.-M.Z. and M.G. contributed to the discussion of the content; and Q.S., Y.K., and C.-W.Q. supervised the project. All authors edited the paper. **Competing interests:** The authors declare no competing interests.

10.1126/science.abj0039

REVIEW SUMMARY

CLIMATE CHANGE

Socioeconomic impacts of marine heatwaves: Global issues and opportunities

Kathryn E. Smith, Michael T. Burrows, Alistair J. Hobday, Alex Sen Gupta, Pippa J. Moore, Mads Thomsen, Thomas Wernberg, Dan A. Smale*

BACKGROUND: Anthropogenic climate change is altering the structure and functioning of ecosystems globally. The upper layers of the oceans have absorbed >90% of the excess heat arising from human activities and, as a result, surface waters have warmed at an average rate of 0.15°C per decade over the past 40 years, with global ocean heat content currently at a record high. Many marine species have shifted their distributions in response to this ubiquitous warming, causing community reconfigurations and changes to entire ecosystems. In addition to gradual warming, the frequency and severity of short-term oceanic warming events, marine heatwaves (MHWs), has increased substantially, largely as a consequence of rising anthropogenic carbon emissions and ocean warming. The global annual number of MHW days has risen by 54% over the past century, with eight of the 10 most extreme MHWs ever recorded occurring after 2010. MHWs are forceful agents of disturbance with wide-ranging impacts on marine ecosystems, including driving species range shifts and mass mortalities and altering foodwebs and species interactions. Regional case studies have shown that ecological responses to MHWs can have socioeconomic implications such as loss of fisheries income, erosion of essential ecosystem services, mass mortalities of iconic species, and conflict. Although our understanding of the physical drivers and biological impacts from MHWs are becoming increasingly robust and predictive, only a few studies have examined their consequences for human-ocean interactions, a crucial step in understanding the associated costs and risks and in developing potential adaptation and mitigation measures.

ADVANCES: Using a large number of case studies, we provide a global perspective on the impacts of MHWs on the provision of ecosystem services and examine the socioeconomic implications. We show that these discrete warming events are altering ecosystem functioning with present-day ramifications for human-ocean interactions. Ecological impacts commonly include harmful algal blooms, species range shifts, mass mortality events, and ecosystem reconfiguration lasting from weeks to years or longer. Such ecological impacts affect humans by changing provisioning, habitat, regulating, and cultural ecosystem services. Clear socioeconomic consequences have been linked to at least 34 MHWs occurring across a wide range of ecosystems in all major ocean basins globally. Impacts on provisioning services relating to fisheries and habitat services relating to loss of foundation species (e.g., corals, kelps, and seagrasses) have been particularly prevalent. Impacts on cultural services relating to mass mortality of charismatic (e.g., sea lions, sea birds, and corals) and spiritual (e.g., certain kelp) species, and on regulating services relating to carbon sequestration and altered nutrient cycling were also common. Biological responses to extreme warming events can, however, also increase human-ocean interactions and offer short-term socioeconomic opportunities through, for example, increased tourism or fisheries prospects. Economic losses of single MHW events to date exceed US$800 million in direct losses and in excess of US$3.1 billion per annum in indirect losses for multiple consecutive years. The true costs are, however, likely to be much greater because many socioeconomic effects likely remain unknown and underreported, particularly in lower-income countries.

OUTLOOK: The likelihood and intensity of MHWs have increased substantially in recent decades, largely due to broad-scale warming attributed to anthropogenic climate change, and are projected to intensify in coming decades. Concurrently, reported ecological impacts, changes in ecosystem services, and economic losses have also increased markedly. We discuss potential adaptation and mitigation measures to reduce risks and consequences for human-ocean interactions and examine knowledge gaps in our current understanding of MHW impacts, for example, in relation to the likely exacerbating effects of other anthropogenic stressors. Globally, marine ecosystems are threatened by a diverse range of anthropogenic stressors, and policy, management, and conservation measures must prioritize approaches to alleviating them to ensure continued provision of ecosystem services. We demonstrate the far-reaching impacts of MHWs on present-day human-ocean interactions and highlight the urgent need to develop a coherent understanding of the linked social-ecological impacts of MHWs and robust approaches to mitigation and adaptation. ■

PHOTO: TORY CHASE/ARC CENTRE OF EXCELLENCE FOR CORAL REEF STUDIES DISTRIBUTED/FLICKR UNDER CC-BY-ND

MHWs have caused widespread loss of habitat-forming species, such as reef-building corals, severely compromising ecosystem service provision.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: dansma@mba.ac.uk
Cite this article as K. E. Smith *et al.*, *Science* 374, eabj3593 (2021). DOI: 10.1126/science.abj3593

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abj3593

REVIEW

CLIMATE CHANGE

Socioeconomic impacts of marine heatwaves: Global issues and opportunities

Kathryn E. Smith[1], Michael T. Burrows[2], Alistair J. Hobday[3], Alex Sen Gupta[4], Pippa J. Moore[5], Mads Thomsen[6,7], Thomas Wernberg[8,9], Dan A. Smale[1]*

Extreme climatic events, including marine heatwaves (MHWs), are altering ecosystems globally, often with profound socioeconomic impacts. We examine how MHWs have affected the provision of ecosystem services and evaluate the socioeconomic consequences for human society. Ecological impacts range from harmful algal blooms and mass mortality events to reconfigurations of entire ecosystems, affecting provisioning, habitat, regulating, and cultural ecosystem services globally. Reported economic costs of individual MHW events exceed US$800 million in direct losses or >US$3.1 billion in indirect losses of ecosystem services for multiple years. However, biological responses to MHWs can also increase human-ocean interactions, providing opportunities for coastal societies. Our study provides a global perspective on the far-reaching impacts of MHWs on human societies and highlights the urgent need to develop robust approaches to mitigation and adaptation.

Anthropogenic climate change is altering the structure and functioning of ecosystems globally (*1*, *2*). The upper layers of the oceans have absorbed >90% of the excess heat arising from human activities and, as a result, surface waters have warmed at an average rate of 0.15°C per decade over the past 40 years (*3*), with rates two to three times higher in some regions (*4*). Marine species have shifted their distributions in response to gradual warming trends, causing community reconfiguration and changes to entire ecosystems (*5*, *6*). In addition to gradual long-term ocean warming, the frequency and severity of discrete warming events, marine heatwaves (MHWs), have increased substantially, with a rise in global annual MHW days of 54% over the past century (*7*). Of the 10 most severe MHWs ever recorded, eight occurred in the most recent decade (*8*). Although MHWs are driven by a range of complex oceanographic and atmospheric processes (*9*), their increasing severity over time can largely be explained by long-term increases in mean ocean temperature caused by anthropogenic greenhouse gas emissions, and they are therefore unequivocally projected to intensify in coming decades (*7*, *10*, *11*).

High temperatures associated with MHWs have been shown to induce physiological stress as thermal thresholds are increasingly exceeded, leading to reduced viability or increased mortality, changes in population structure and species' distributions, and altered ecosystem structure and functioning (*11*). Indeed, ecosystem-wide impacts of MHWs have been reported from a number of regions where foundation species that support a wide range of associated biodiversity have been affected, including loss of kelp forests in the northeast and southwest Pacific and southeast Indian Oceans (*12*–*15*), coral bleaching in the Indo-Pacific and southwest Indian Ocean (*16*–*18*), widespread mortality of habitat-forming invertebrates in the Mediterranean Sea (*19*), and depleted extent and resilience of seagrass meadows in the southeast Indian Ocean (*20*, *21*). MHWs have also been linked to mass dies-offs of invertebrates, fish, seabirds, and marine mammals (*22*–*26*); redistributions of commercially important fisheries species (*27*, *28*); and disrupted food webs and carbon and nitrogen cycles (*21*, *29*). An analysis of eight high-profile MHWs showed that all had negative impacts across a range of taxa and ecological functions (*30*). Responses to extreme climatic events are highly variable, however, with some species or populations showing little response and others even responding positively by extending their ranges (*28*, *31*) or proliferating as a result of increasingly favorable conditions or competitive release (*13*, *28*, *32*, *33*). Broadly, populations found toward their species' poleward range edge are more likely to respond positively to MHW events (*30*), and species exhibiting certain life history traits (e.g., high mobility, generalist feeding, or wide distribution) may proliferate more rapidly with warming (*34*). Crucially, regional species losses or gains may be either temporary or persistent, depending on a variety of factors such as the characteristics of the MHW, the wider environmental and climatological context, and ecological interactions (*11*). Although our understanding of the physical drivers and biological impacts of MHWs is becoming increasingly robust and predictive, only a few scattered regional studies have examined their consequences on linked human-ocean interactions [e.g., (*35*–*37*)], and those that have typically reported regional losses in excess of hundreds of millions USD from individual MHW events, often with acute societal responses (*28*, *38*, *39*). Conversely, MHWs have also resulted in economic gains, for example, through increased fisheries or tourism opportunities (*28*, *31*, *32*, *37*).

We draw on a large number of case studies to provide a global perspective on MHW-associated risks, costs, and opportunities, and use it to guide priorities and management decisions that counter or take advantage of more frequent and severe MHWs in the future. We review the linked impacts of MHWs globally to quantitatively examine associated changes in ecosystem service provision and evaluate socioeconomic consequences. Through analyses of global datasets, we identify a broad range of biological response variables related to ecosystem services and investigate variability across 10-year periods spanning prominent MHWs. We classify responses by associated ecosystem services using The Economics of Ecosystems and Biodiversity framework [TEEB; (*40*)] alongside a coherent framework to identify and describe MHWs (*41*, *42*), allowing for direct comparisons of events, regions, and type of impacts. Specifically, an event is considered a MHW if it lasts for 5 or more days, with temperatures warmer than the 90th percentile based on a 30-year historical climatology (*41*). We then examine the impacts of MHWs in the context of United Nations Sustainable Development Goals (SDGs), because these extreme events will affect the achievement of several of these goals that seek improved socioeconomic outcomes for the planet. Finally, we explore possible adaptation measures and highlight future directions for research.

Impacts of MHWs on ecosystem services

We found compelling evidence for MHW impacts globally, on all major types of ecosystem services, with extensive socioeconomic consequences (Tables 1 and 2 and Fig. 1). Over the past 25 years, at least 34 MHW events have affected provisioning, regulating, habitat, and cultural ecosystem services (Fig. 1). For example, MHWs spanning the Indo Pacific associated with the extreme 1997/1998 El Niño caused

[1]Marine Biological Association of the United Kingdom, The Laboratory, Citadel Hill, Plymouth, PL1 2PB, UK. [2]Scottish Association for Marine Science, Oban, Argyll PA37 1QA, UK. [3]CSIRO Oceans and Atmosphere, Hobart, Tasmania 7000, Australia. [4]Climate Change Research Centre, University of New South Wales, Sydney, New South Wales 2052, Australia. [5]School of Natural and Environmental Sciences, Newcastle University, Newcastle-Upon-Tyne NE1 7RU, UK. [6]The Marine Ecology Research Group, Centre of Integrative Ecology, School of Biological Sciences, University of Canterbury, 8041 Christchurch, New Zealand. [7]Department of Bioscience, Aarhus University, 4000 Roskilde, Denmark. [8]University of Western Australia, Oceans Institute and School of Biological Sciences, Crawley, Western Australia 6009, Australia. [9]Institute of Marine Research, Floedevigen, 4817 His, Norway.
*Corresponding author. Email: dansma@mba.ac.uk

Table 1. Examples of direct economic outcomes of MHWs.

Location	Year(s)	Socioeconomic impact	Location of impact	Value/loss/gain	Primary industry affected	Affected TEEB ecosystem service	Reference(s)
Gulf of Alaska	2014–2019	Population decline and low recruitment affects Gulf of Alaska commercial Pacific cod fishery catch limits	Alaska, USA	Fishery valued at US$103 million per annum; loss unknown; funding relief: In 2018, Congress appropriated US$24.4 million in federal disaster relief funding.*	Commercial fisheries	Provisioning and habitat services	(*29*)
	2020	Population decline and low recruitment closes Gulf of Alaska commercial Pacific cod fishery for the season	Alaska, USA	Fishery valued at US$103 million per annum; comparative loss likely; federal disaster relief funding requested (pending as of 5 May 2021)*	Commercial fisheries	Provisioning and habitat services	(*29*, *48*)
Northeast Pacific Ocean	2015	Harmful algal blooms cause closure of recreational razor clamming	Washington, USA	US$40 million loss in tourist spending	Tourism and recreational fisheries	Regulating and cultural	(*49*, *66*)
	2015–2016	Harmful algal blooms cause closure of commercial Dungeness crab fisheries	West Coast, USA	US$97.5 million loss; funding relief: Congress to appropriate US$27.3 million in federal disaster relief funding including US$1.5 million to the native American Quileute Tribe*	Commercial fisheries	Provisioning and regulating	(*67*, *103*)
	2015–present	Population decline of red sea urchin due to kelp loss causes closure of commercial fishery	California, USA	Fishery valued at US$3 million per annum; funding relief: Congress appropriated US$3.3 million in federal disaster relief funding for 2016 and 2017*	Commercial fisheries	Provisioning, regulating, habitat, and cultural	(*14*)
	2018–present	Population decline of abalone due to kelp loss causes closure of recreational fishery	California, USA	US$44 million loss per annum	Tourism and recreational fishing	Provisioning, regulating, habitat, and cultural	(*14*)
Southeast Pacific Ocean	2016–2017	Mass mortality of farmed salmon due to harmful algal blooms	Southern Chile	Export loss of US$800 million	Aquaculture	Provisioning and regulating	(*38*)
Coral Sea	2016,,–2018	Bleaching and mass mortalities of corals	Great Barrier Reef, Australia	The Great Barrier Reef is valued at US$4.2 billion annually, with a total value estimated at US$41 billion; loss related to bleaching unknown; gains related to "last chance tourism" also unknown	Tourism	Habitat and cultural	(*69*)

continued on next page

Location	Year(s)	Socioeconomic impact	Location of impact	Value/loss/gain	Primary industry affected	Affected TEEB ecosystem service	Reference(s)
Tasman Sea	2016	Pacific oyster mortality syndrome closed hatcheries and decimated juvenile stocks	East coast of Tasmania, Australia	Aquaculture valued at US$19 million; loss unknown	Aquaculture	Provisioning and habitat	(*104*)
	2016	Poor performance of salmon limited production	East coast of Tasmania, Australia	Aquaculture valued at US$545 million; loss unknown	Aquaculture	Provisioning and habitat	(*104*)
	2016	Mortality of wild caught abalone led to smaller catch and reduced quotas	South/southeast coast of Tasmania, Australia	Wild-caught fishery valued at US$62 million; loss unknown	Fisheries	Provisioning and habitat	(*104*)
Southeast Asia seas	2010	Bleaching and mass mortalities of corals	Indonesia, Malaysia, Thailand	US$49 to 74 million loss	Tourism	Habitat and cultural	(*73*)
Southeast Indian Ocean	2011–present	Closure of abalone fishery due to mass mortality event	West Coast, Australia	Estimated loss per annum ~US$0.16 million	Commercial and recreational fisheries	Provisioning, habitat, and cultural	(*51*)
	2011–2013	Closure of commercial swimmer crab fishery for 18 months to protect breeding stock after low catch rates	Shark Bay, Australia	US$3.1 million loss	Commercial fisheries	Provisioning and habitat	(*105*)
	2011–2016	Closure of commercial scallop fishery for 3 years (Shark Bay) or 5 years (Abrolhos Islands and Midwest) due to low recruitment	West Coast, Australia	Estimated losses per annum: Shark Bay ~US$8.2 million†; Abrolhos Islands and Midwest ~US$3 million†	Commercial fisheries	Provisioning and habitat	(*51*)
	2011	Loss of carbon storage and other ecosystem services provided by seagrass	Shark Bay, Western Australia, Australia	US$3.1 billion loss per annum for multiple consecutive years	Multiple ecosystem services	Provisioning, regulating, habitat, and cultural	(*21*)
Gulf of Maine	2012	Early lobster migration led to record landings and drop in lobster value	Gulf of Maine, USA	US$38 million loss	Commercial fisheries	Provisioning and habitat	(*77*)
	2016	Proactive management of lobster fishery after 2012 MHW led to economic gains	Gulf of Maine, USA	US$108 million gain	Commercial fisheries	Provisioning and habitat	(*77*)

*Funding relief values in the United States are from the National Oceanic and Atmospheric Administration (NOAA; https://www.fisheries.noaa.gov/national/funding-and-financial-services/fishery-disaster-determinations). †Values estimated from the Government of Western Australia State of the fisheries annual reports (https://www.fish.wa.gov.au/About-Us/Publications/Pages/State-of-the-Fisheries-report.aspx).

extensive coral bleaching and depleted zooplankton abundance, leading to mass mortalities of a range of marine mammals and seabirds. Ultimately, the disruption of ecosystem functioning had major ramifications for tourism and fishing across the Pacific (*43*). Similarly, the Mediterranean Sea has been repeatedly affected by strong MHWs, causing mass mortality of seagrass, corals, invertebrates, and fish and population explosions of invasive species and pathogens (*19*, *24*, *44*, *45*). These responses have markedly reduced provisioning (e.g., fisheries), cultural (e.g., mass mortalities of iconic species), regulating (e.g., carbon capture, water quality), and habitat (or supporting, e.g., restructuring, disease) ecosystem services (*19*, *24*, *43*–*45*) (Fig. 1), which, collectively, are of global socioeconomic importance.

The most pervasive socioeconomic impacts of MHWs reported to date relate to provisioning services, particularly fisheries, which directly influence markets, employment, food availability, and wider ecosystem services (Figs. 1 and 2). Impacts on provisioning services often closely align with cultural services by affecting tourism and recreational fisheries. In the northeast Pacific, "the blob" was associated with multiple regional MHWs between 2014 and 2016 that affected ecosystem services (*46*). In the Gulf of Alaska, the zooplankton community shifted

Table 2. Reported impacts of MHW events. Responses in **bold** indicate positive socioeconomic consequences, those in *italics* are negative consequences, and those in roman are both negative and positive.

Location	Year(s)	Recorded response variable	Reference(s)
Southeast Atlantic (Benguela Niño)	1995	*Mass mortalities of fish*, southward displacement of fish species	(*22*)
South Indian Ocean	1997	*Mortality and reproductive failure in king penguins, range shift in prey species*	(*106*, *107*)
East Pacific El Niño	1997–1998	Shift in fisheries species, *mass mortalities and reproductive failure of Galapagos sea lions, decline in kelp forests, low zooplankton abundance, coral bleaching, ecosystem disruption*	(*26*, *27*, *43*, *108*, *109*)
West Indian Ocean	1998	*Coral bleaching, restructuring of ecosystems*	(*61*, *110*)
Red Sea	1998, 2007, 2010, 2012, 2015,	*Coral bleaching*	(*111*)
	2017	*Mass mortalities of coral reef fish, fish disease*	(*112*)
Mediterranean Sea	1999	*Mass mortalities of benthic organisms, habitat loss, reduced growth and reproduction, coral bleaching*	(*19*, *24*, *113*, *114*)
Mediterranean Sea	2003	*Mortalities in mollusk fisheries (multiple life stages), mass mortalities of benthic organisms, coral bleaching, seagrass flowering and shoot mortality, reduced growth and reproduction, disease*	(*24*, *44*, *45*, *114–116*)
Caribbean Sea	2005	*Coral bleaching, loss of fisheries*	(*16*, *117*, *118*)
Mediterranean Sea	2006	*Mortalities in mollusk fisheries (multiple life stages), mass mortalities of benthic organisms, shoot mortality in seagrass*	(*24*, *44*, *45*)
Canadian Arctic	2009	Earlier sea ice breakup, decreased sea ice, **increased productivity**, **increased cod reproduction**	(*32*, *33*)
Bay of Bengal	2010	*Coral bleaching*	(*16*, *17*)
Southeast Indian Ocean (Ningaloo Niño)	2011	*Loss of seagrass and kelp, coral bleaching, fisheries closures, reduced fisheries quotas*, shifts in recruitment, **range expansion of tropical fish**	(*15*, *31*, *36*, *50*, *51*, *119–121*)
Northwest Atlantic	2012	Spring boom in lobster landings, range shift of a range of commercially valuable fisheries species (e.g., squid, cod, flounder, hake)	(*35*, *39*, *77*, *84*)
Great Australian Bight	2013	*Mass mortalities of fish and abalone, harmful algal blooms, disease*	(*122*)
Northeast Pacific (The Blob)	2014–2016	*Mass mortalities of California sea lions, seals, seabirds, and marine invertebrates; sea star wasting disease*; range shifts in a variety of species; *harmful algal blooms*; *fisheries disruptions and closures*; *shifts in kelp forest ecosystems*; **increase in orca births**; *coral bleaching*; **increased abundance of tuna**; **unusual sightings of warm water species**; increases and decreases in fisheries recruitment; **increased observations of whales**; *increased whale entanglement*	(*25*, *28*, *38*, *123–125*)
Northwest Atlantic	2015–2016	*Loss of seagrass*, **early migration of lobsters**	(*58*, *77*)
Southwest Atlantic	2015	*Harmful algal blooms, mass fish mortalities, closure of fisheries, flags flown on beaches to indicate potential health risks*	(*126*)
Central Pacific	2015–2016	*Mass coral bleaching and mortality, shift from coral to encrusting macroalgae and crustose coralline algae, decreased fish biomass, reduction in seabirds and seabird breeding*	(*61*, *127*, *128*)

continued on next page

Location	Year(s)	Recorded response variable	Reference(s)
Southeast Indian Ocean	2015–2016	*Coral bleaching*	(*121*, *129*)
Tasman Sea	2015–2016	*Loss of fisheries*, *loss of kelp*, *POMS*, *mortality of abalone*, fish species noted outside of their range	(*74*, *130*)
Bering Sea	2016	*Range expansion of bowhead whales*, *mass mortality of puffins*, *low sea ice*, *reduced ice-based fishing*, *low snow and unsafe river ice making winter transportation difficult*, *topicalization of copepods*	(*131*)
Southern Ocean	2016	*Decreased sea ice*	(*132*)
Southeast Pacific	2016–2017	*Mass mortalities of farmed salmon*, *harmful algal blooms*	(*38*, *65*)
Southeast Asia seas	2016–2017	*Mass coral bleaching*, *reefs closed to tourism*	(*132*)
Coral Sea	2016–2018	*Mass coral bleaching*, *mass coral mortalities*, *failed coral recruitment*, *species range shifts*, *loss of seagrass*, *poor reproductive effort in seagrass*, increased abundance and biomass of coral trout	(*37*, *53*, *121*, *133–138*)
Tasman Sea	2017	*Loss of kelp*, range shift of fish and jellyfish, early spawning in snapper, *salmon mortalities in fish farms*	(*13*, *139*)
East China Sea	2017	*Harmful algal blooms*, *coral bleaching*, *mass mortalities of farmed fish*	(*140–142*)
Southwest Atlantic	2017	*Harmful algal blooms*, *mass mortalities of fish*, *closure of recreational beaches*	(*143*, *144*)
West Pacific	2017	*Mass mortality of coral after successive bleaching events in 2013*, *2014*, *2016*, *and 2017*	(*61*, *145*)
Persian Gulf	2017	*Mass coral bleaching*, *coral mortality*, *increase in turf algae*	(*146*)
Northeast Pacific	2018	*Loss of kelp*	(*12*)
Northwest Atlantic	2018	*Loss of kelp forest*	(*55*)
North and Baltic Seas	2018	*Loss of kelp forest*	(*55*)
Northeast Pacific	2019–2021	*Harmful algal blooms*, *Dungeness crab fisheries closures*, *recreational razor clam fisheries closures*, *fish mortalities at fish farms*, olive Ridley sea turtles observed off British Colombia, *coral bleaching*	(*38*, *147–150*)

from cold-water, lipid-rich copepods to less nutritious warm-water species, reducing food availability and thus the abundances of groundfish, including Pacific cod (*Gadus macrocephalus*) (*29*) and Alaskan pollock (*Gadus chalcogrammus*) (*47*), and resulting in a massive die-off of fish-eating seabirds, including common murres (*Uria aalge*) (*25*) (Fig. 2A). The effect of the MHW on the life cycle of cod persisted for >5 years, as recruitment rates and spawning biomass remained well below pre-MHW levels (*29*, *48*). This had major implications for the regional fishery and ecosystem services worth US$103 million per year, with reduced quotas for several years and closure of the federal fishery in 2020 (*49*). Impacts of the blob were widespread: Throughout the West Coast, mass die-offs of finfish, shellfish, seabirds, sea lions, and seals were reported (*25*, *28*); harmful algal blooms were extensive; and commercial and recreational fisheries were closed (*14*, *49*). Moreover, off the coast of California, a poleward range shift of market squid [*Doryteuthis (Loligo) opalescens*] resulted in higher and lower landings in northern and southern Californian counties, respectively, thereby altering supply chains and market values.

Other recent MHWs that have (positively or negatively) affected provisioning and cultural services include a MHW in the Canadian Arctic in 2009, which led to early sea ice breakup, resulting in increased primary productivity and enhanced polar cod (*Boreogadus saida*) recruitment (*32*). The Ningaloo Niño MHW in the southeast Indian Ocean during 2010–2011 (*15*, *50*) caused poleward range extensions of tropical fish species, leading to a short-term boom in landings of valuable aquarium species (West Australian butterflyfish; *Chaetodon assarius*) off the coast of Perth, Western Australia (Fig. 2B), and the opening of a new commercial rabbitfish fishery (*Siganus* spp.) in Cockburn Sound, Western Australia (*31*). At the same time, reduced growth rates of abalone (*Haliotis* spp.) led to a 50% drop in recreational landings near Perth (Fig. 2B), whereas farther north, mortalities and reduced recruitment of Roe's abalone (*Haliotis roei*), scallops (*Amusium balloti*), and blue swimmer crabs (*Portunus armatus*) necessitated closures affecting these fisheries for several years (*36*, *51*). In the Gulf of Maine, a MHW in 2012 resulted in an early lobster (*Homarus americanus)* migration from deep to shallow water and an early molt to legal size (*35*). This led to high spring landings that were beyond processing capacity, causing a drop in lobster value (*35*, *39*). Civil unrest followed as Canadian protests and blockades attempted to prevent imports of lobster from the United States into Canada for processing, and values

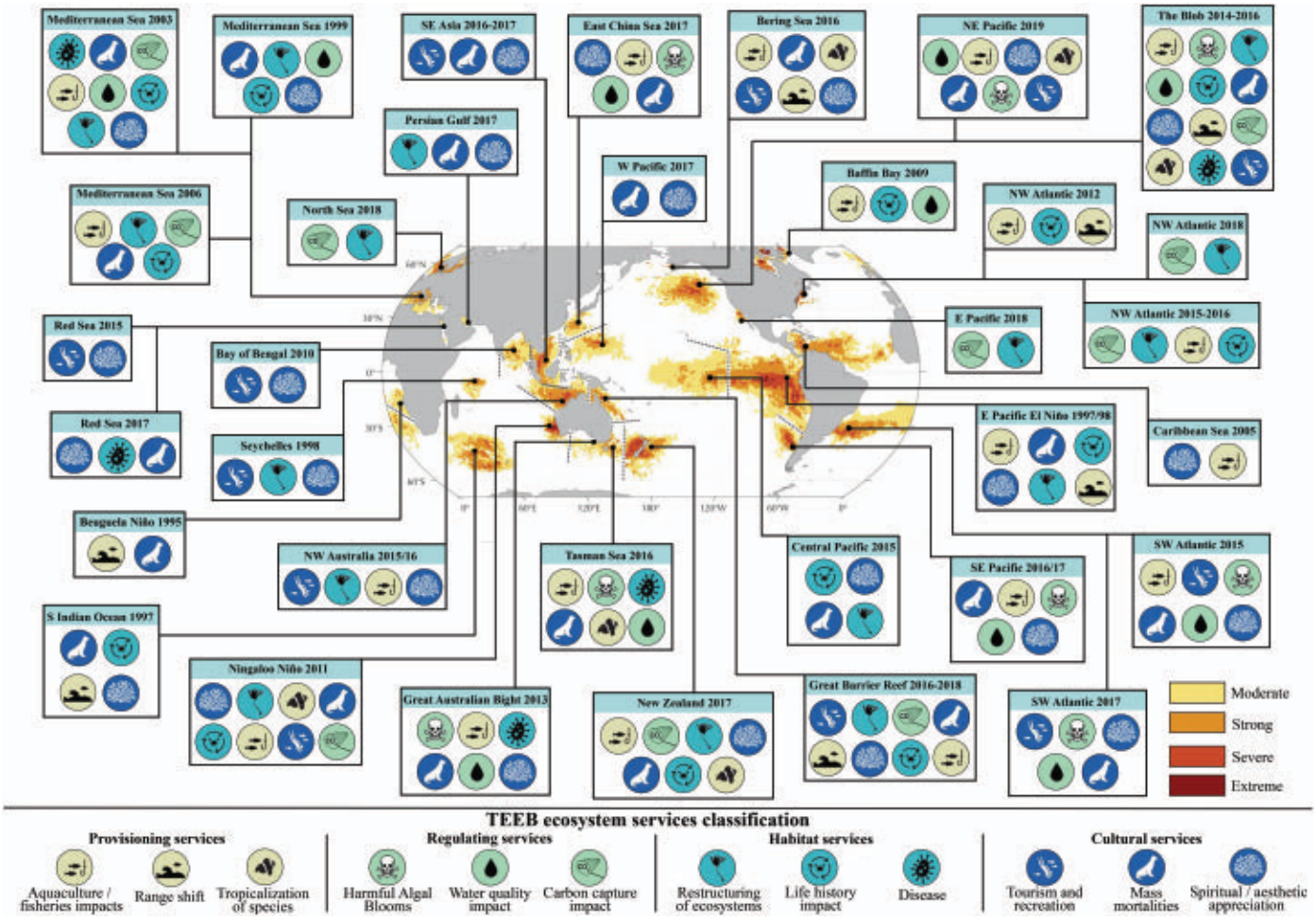


Fig. 1. Socioeconomic impacts of major MHW events since 1995. An event was considered a MHW if sea surface temperatures (SSTs) exceeded the 90th percentile for ≥5 days. The intensity scale of each event (moderate to extreme) represents conditions corresponding to the peak date of the event as defined by (*41*). Intensity categories are identified as sequential multiples of the difference between the 90th percentile and the mean climatology as defined by (*42*). For each event, biological responses have been linked to ecosystem services based on TEEB ecosystem services classifications [provisioning services, cultural services, regulating services, and habitat services; (*40*)], which recognize the economic value of nature. For further details on positive and negative biological responses and a full reference list, see Table 2.

fell on either side of the border (*52*). By contrast, the same MHW shifted longfin squid [*Doryteuthis (Amerigo) pealelli*] distribution poleward, enabling the Maine fishing fleet to develop a profitable new market within the season (*35*). Recent MHWs during 2016–2017 along the Great Barrier Reef have resulted in bleaching, and ultimately mortality, of coral (*53*), with the unexpected side effect of an increase in the presence of coral trout (*Plectropomus* and *Variola* spp.) near reefs, resulting in higher catch-per-unit effort by the commercial fishery (*37*).

Kelp forests, seagrass meadows, and coral reefs are globally distributed coastal marine habitats that deliver regulating services via nutrient cycling and carbon storage (blue carbon) and habitat services by offering biogenic structure for socioeconomically important species and by maintaining high levels of biodiversity (*54*). However, species that underpin these habitats are susceptible to MHWs (*30*), particularly where populations are situated toward upper thermal limits (i.e., equatorward range edges). Off the coast of Baja California, Mexico, giant kelp (*Macrocystis pyrifera*) populations were negatively affected by multiple MHWs between 2014 and 2018, although the magnitude of responses varied between locations within the species range (*12*) (Fig. 2C). The loss of giant kelp populations at the equatorward range edges after the MHWs affected several regulating (i.e., carbon capture and transfer, nutrient cycling, and storm protection) and habitat (i.e., ecosystem restructuring and habitat for associated species) services. Similarly, kelp populations on either side of the North Atlantic (the eastern United States and southern Norway) suffered mortalities during MHWs in 2018 (*55*), with similar implications for ecosystem service provision. Loss of southern bull kelp in New Zealand after the 2017–2018 Tasman Sea MHW (*13*) have, in addition to the loss of the aforementioned kelp-forest services, also affected traditional knowledge and spirituality because indigenous tribes on the South Island rely on abundant bull kelp to make "poha" bags for food storage. Indeed, ancient stories repeated across generations detail "sacred poha kelp bags" that symbolize the regions wealth, "mana," and authority over food supplies. MHW-associated loss of bull kelp can thereby reduce tribal spiritual power and their long-term connection with the land and sea (*56*). Widespread loss of seagrass meadows as a result of MHWs have been reported off eastern and western Australia (*21*, *57*), off the east coast of the United States (*58*), and in the Mediterranean Sea (*44*), reducing habitat for economically and culturally important associated species and compromising long-term carbon stocks (*59*). Reef-building corals are particularly sensitive to extreme heat (*60*), and between

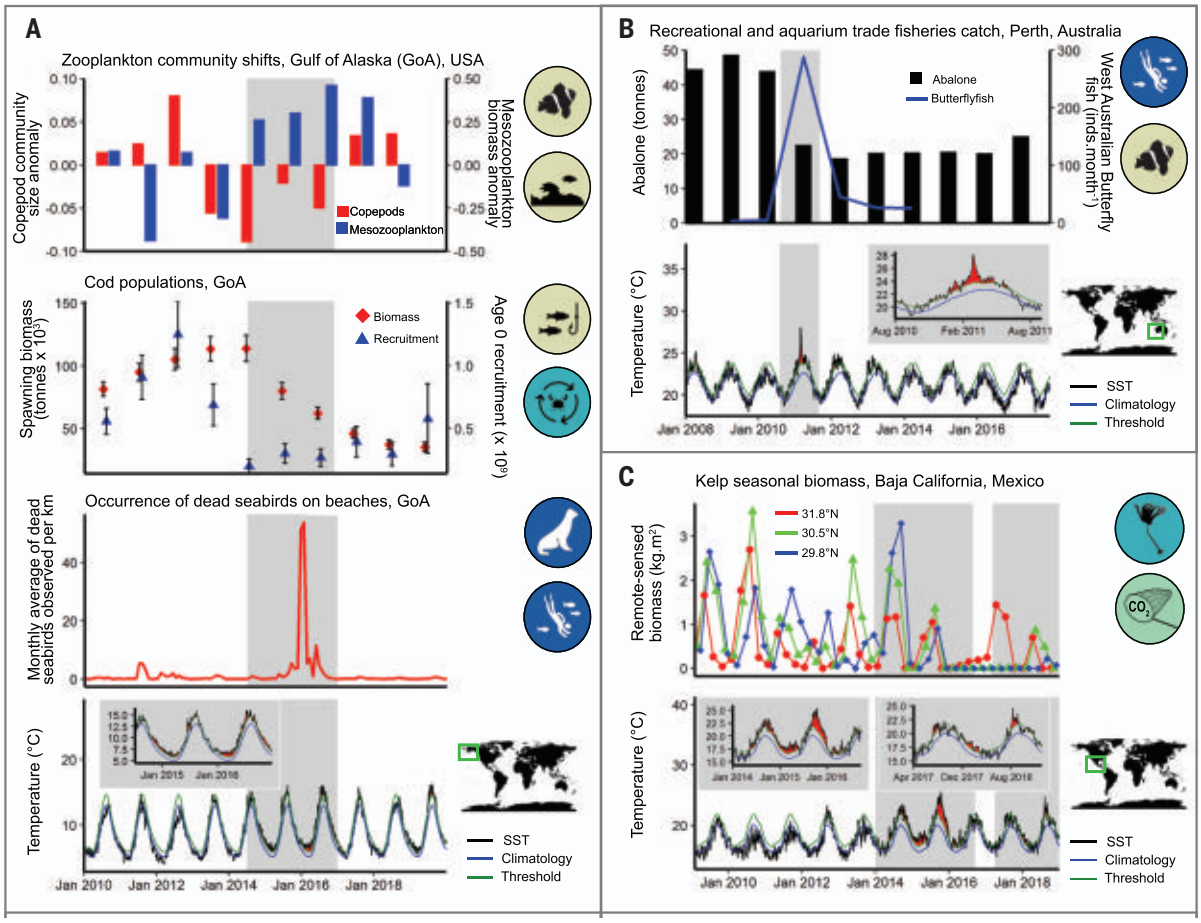


Fig. 2. Impacts of MHWs on ecosystem services. (**A**) Between 2014 and 2016, MHWs in the northeast Pacific led to a shift in zooplankton community in the Gulf of Alaska from lipid-rich, cold water species to smaller warm-water species [top plot; data extracted from (*100*)], reducing food supplies for groundfish. Concurrently, the spawning biomass and recruitment in cod (*Gadus macrocephalus*) decreased [second plot; data from (*48*)], and a mass die-off of fish-eating seabirds occurred due to starvation (third plot; data from https://COASST.org). Error bars represent 1 SD. (**B**) The 2011 MHW off Western Australia led to a short-term increase in butterflyfish [*Chaetodon assarius*, valuable to the aquarium trade; data extracted from (*31*)] and a long-term decrease in abalone growth, resulting in an ~50% reduction in recreational landings for at least 7 years (data from the Government of Western Australia Department of Fisheries Annual reports; https://www.fish.wa.gov.au/About-Us/Publications/Pages/State-of-the-Fisheries-report.aspx). (**C**) Giant kelp (*Macrocystis pyrifera*) off the coast of Baja California has been affected by multiple MHWs since 2014, with equatorward populations being most severely affected [data extracted from (*12*)]. The SST climatology (blue), 90th percentile (green), SST time series (black), and periods of time with identified MHWs (red fill) for each location are shown in the bottom plots. Gray shading and inset climatologies indicate the timing of respective MHWs. An event was considered a MHW if SSTs exceeded the 90th percentile for periods of ≥5 days (*41*). All circled icons are as identified in Fig. 1. Climatology timelines were generated in RStudio (*101*) using the HeatwaveR package [R Core Team, https://www.R-project.org, R package version 0.4.4, https://CRAN.R-project.org/package=heatwaveR) (*102*)].

2014 and 2017, MHWs were responsible for a global-scale coral-bleaching event (*61*) that caused habitat loss and ecosystem reconfiguration in many regions (*53*). As MHWs increase in frequency (*7*), the opportunity for recovery between successive events is further reduced, exacerbating ecosystem degradation and affecting regulating, provisioning, habitat, and cultural services, with wide-ranging socio-economic consequences.

Economic consequences of MHWs

Conservative estimates value global ocean assets at ~US$24 trillion (*3*). For example, coral reefs are worth ~US$35.8 billion per year globally in tourism alone (*62*), Australian kelp forests ≥US$7.8 billion per year in fishing and tourism (*63*), and U.S. saltwater fishing industries generate >US$210 billion in sales annually and support 1.7 million jobs (*64*). Where economic estimates of the impacts of MHWs have been made, reported gains are typically rare and relatively small, whereas losses are common and often extensive (Table 1). In Chile, MHWs associated with the El Niño in 2016 induced harmful algal blooms, which in turn caused the largest fish farm mortality ever recorded globally, resulting in an export loss of US $800 million (*38*, *65*). Similarly, off the West Coast of the United States, a combination of harmful algal blooms and loss of kelp forest habitat (both caused by MHWs) led to closures of shellfisheries in 2015, costing the economy in excess of US$185 million in that year alone (*14*, *66*, *67*). The commercial tristate Dungeness crab (*Metacarcinus magister*) fishery recorded a loss of ~US$ 97.5 million, affecting both tribal and nontribal fisheries (*67*), and Washington and Californian coastal communities lost a combined ~ US$84 million in tourist spending due to the closure of recreational razor clam (*Siliqua patula*) and abalone (*Haliotis* spp.) fisheries (*14*, *66*). By contrast, the higher temperatures were associated with an increased abundance of warmer water species such as the Pacific bluefin tuna (*Thunnus orientalis*), resulting in an increase in sports fishing charters, and the whale-watching industry saw a surge

in wildlife sightings with associated economic gains from increased demand (*28*, *68*). A surprising outcome of the 2016–2017 coral reef bleaching on the Great Barrier Reef was that, despite expectations, tourism in the area was not affected. Instead, "last chance tourism" motivated visitors, saw an increase in protective attitudes toward the reef, and illustrated the potential for the loss of iconic ecosystems to mobilize the public toward tackling climate change (*69*).

Loss of biogenic coastal marine habitats are particularly costly because of their role in blue carbon sequestration and as nursery habitat for economically and culturally important species, thereby underpinning provisioning and habitat services. Estimates place the annual per-hectare ecosystem service value of seagrass and coral reefs at ~US$28,916, and US$352,249, respectively (*70*). Scaling this up, the 106,900 hectares of seagrass completely lost in Shark Bay, Western Australia, after the Ningaloo Niño (*57*) is equivalent to ~US$3.1 billion in lost ecosystem services per year, with little seagrass recovery observed after 5 years (*57*). In terms of climate regulation alone, the loss of seagrass habitat in Shark Bay, which was the largest seagrass mortality event in the world recorded to date, resulted in the potential release of between 2 and 9 Tg CO_2 from sedimentary carbon stocks (*21*).

Interpreting the impacts of MHWs in the context of the SDGs

Contextualizing biological responses within the framework provided by the SDGs further highlights the socioeconomic impacts of MHWs. Clearly, MHWs affect our ability to protect life below water (SDG 14), but there are multiple other SDGs that are interlinked with the impacts of these extreme warming events. Most responses relating to fisheries or aquaculture hinder our ability to achieve zero hunger (SDG 2) or decent work and economic growth (SDG 8), with local effects sometimes having global impacts. For example, the Peruvian anchovy (*Engraulis ringens*) is by far the most caught fish species globally (*71*). Anchovies are almost exclusively processed into fishmeal and fish oil, which are used extensively in aquaculture and livestock feeds, with Peru providing as much as 50 and 33% of total global fishmeal and fish oil production, respectively (*72*). An intense MHW associated with the 1997–1998 El Niño and lasting ~8 months caused a temporary range shift in anchovies, ultimately leading to a decrease in landings from 7.5 million tons in 1996 to 1.2 million tons in 1998 (data from FishStatJ, available at www.fao.org/fishery/statistics/software/fishstatj/en), affecting availability of fishmeal and fish oil worldwide and resulting in a reduction in global fishmeal production alone by ~25% (data from seafish, available at www.seafish.org/document/?id=1b08b6d5-75d9-4179-9094-840195ceee4b).

Coastal and island communities globally rely heavily on human-ocean interactions for income, further highlighting the impacts of MHWs on SDGs 2 and 8. For example, in Southeast Asia, a MHW in 2010 caused a coral-bleaching event costing >US$49 to 74 million (based on 2010 USD values) in SCUBA-related tourism, (*73*), and likely impacts on artisanal fisheries. Further, >10% of the Tasmanian Pacific oyster (*Magallana gigas*) industry workforce lost their jobs within months of Tasmania's first outbreak of Pacific oyster mortality syndrome (POMS), which was linked to the Tasman Sea 2015–2016 MHW (*74*). The Tasmanian POMS outbreak also affected SDG 12 (manage responsible consumption and production) by preventing the movement of spat (juvenile oysters) from Tasmania to seed the Australian oyster industry, affecting production in ~300 farms (*75*).

Mortality events associated with MHWs in the Mediterranean Sea have been widely recorded, particularly for habitat-forming filter- or suspension-feeding invertebrates such as cnidarians, sponges, and bivalves, thereby reducing water quality and local biodiversity (*24*). Widespread losses of vast seagrass habitats have been linked with MHWs in Western Australia and the Mediterranean (*44*, *57*), reducing natural carbon sequestration capacity and, thus, the likelihood of meeting SDG13 (climate action). Similarly, kelp forests generally support a significant standing stock of carbon (*76*) and their loss exemplifies impacts of MHWs on SDG13.

Adaptation, mitigation, and resilience

Human societies are beginning to adapt to the impacts of MHWs to protect ecosystem services and limit socioeconomic consequences. Local industries in several areas have responded to limit economic losses from MHWs in the short-term. For example, during MHWs in North and South America, fisheries efforts were refocused to target transiently abundant species (*27*, *35*, *77*). Longer term, fishermen are developing aquaculture programs as a secondary source of income (*78*) and governments are encouraging diversification to focus on fisheries species expected to increase in abundance with climate warming (*79*). Approaches also include increasing demand for previously rare or undesirable species through marketing efforts such as renaming and rebranding (*80*). Although shifting to newly arrived species might ease short-term fishery disruption, it may also slow the population growth of these shifting species such that larger catches are delayed for many years. Different exploitation pathways should be explored with ecosystem models to aid fishery decision makers. Furthermore, aquaculture industries are responding globally to MHWs by considering climatic extremes when selecting new aquaculture sites, for example, by using remote sensing data (*81*, *82*), by adjusting stock practices to reduce the likelihood of exposure to extreme climatic events (*83*), by regulating geographic production based on temperature-related lice and disease outbreaks (*82*), and by selective breeding for disease-resistant aquaculture strains (*75*).

Early management intervention and shifts in management approaches can also help to limit the impacts of MHWs. For example, after a US$38 million loss from the lobster price crash caused by the 2012 MHW in the northwest Atlantic, processing capacity was expanded and a proactive management system was applied, which included quotas determined by climatological forecasts (*35*, *84*). Such efforts to increase the resilience of the fishery proved effective during a subsequent MHW in 2016, when the shift in management approach led to a gain of US$108 million (*39*, *77*). Similarly, in many regions, the combination of stock assessments and flexible harvesting strategies allow management to intervene early to reduce MHW impacts, and a variety of approaches, including quotas, fisheries closures, and restocking, have been implemented to facilitate species recovery and stabilize stocks after MHWs (*39*, *51*). For example, low female spawning biomass identified during stock assessments in 2019 led to the closure of the Gulf of Alaska Pacific cod federal fishery in 2020. Removing fishing pressure for a season enabled the cod population to begin recovering and helped to ensure that adequate prey were available for the endangered western stock of the Steller sea lion (48). However, consistent with increasing MHWs, the United States has seen a significant rise in federal fishery disasters related to climate warming over the past 30 years (*85*), suggesting that whereas proactive management can prove beneficial, it may be more effectively used in combination with other mitigation strategies.

Natural populations may also be showing resilience to MHWs, with emerging evidence suggesting that moderate climatic events affect population genetics (*86*). For example, the 2011 Ningaloo Niño caused substantial changes to the genetic structure of kelp forests in the southeast Indian Ocean, with a loss of genetic diversity (*87*) and an increase in warm-affiliated genotypes (*88*, *89*), consistent with genomic evidence for adaptive changes to increased temperature tolerance (*90*). Therefore, moderate MHWs are likely to accelerate selection for heat tolerance. Similarly, coral populations exposed to previous MHWs and extreme temperature variation show increased resilience to warming events (*91*).

It remains uncertain whether marine protected areas (MPAs) can buffer the impacts of

extreme climatic events in a warming ocean (*92*). Evidence certainly indicates that warming events occurring inside MPAs have comparable, or sometimes even greater, impacts as events in nonprotected areas [(*92*) and references therein]. This is partly because, historically, climate change and extreme climatic events have not been considered during the design and implementation phases of MPAs (*92*). MPAs are more likely to buffer the impacts of MHWs if climate change responses are considered in their design. Planning larger networks of no-take MPAs may provide some protection for species tracking their thermal niches compared with smaller, isolated MPAs. Likewise, targeting locations with steep depth (and therefore often temperature) gradients or strong upwelling may offer climatic refugia. Furthermore, examining relevant scales of spatial variability in sea temperature may identify climatic refugia for the foundation species that underpin biodiversity (*92*, *93*). Certainly, resilience to and recovery from MHW events vary spatially, and what drives this is not yet well understood but may be related to local environmental conditions (*58*, *94*). For example, giant kelp populations off Bahia Tortugas, Mexico, close to the equatorward range edge of the species, have shown greater resilience to and more rapid recovery from multiple MHWs compared with populations both north and south of this location (*94*, *95*), suggesting that this area could serve as an important refuge for giant kelp and its associated communities (and therefore should be prioritized in designs of future MPAs). Regardless, MPAs are likely to be more effective at mitigating the impacts of MHWs if combined with other management tools such as restoration or stock enhancement (*96*).

Future directions

Efforts to improve mechanistic understanding of the physical drivers of MHWs will undoubtedly lead to better climatological forecasting and predictions (*39*), which in turn will provide opportunities to implement mitigation measures. It is clear that MHWs induce both direct (e.g., coral reef bleaching reduces tourism) and indirect (e.g., loss of kelp forests results in lack of nursery grounds for fisheries species, which leads to reduced landings) impacts on ecosystem services, which can last from weeks to decades or more. The realization that heritable genotypic variation likely underpins responses to MHWs opens the possibility to harness "future-proof" populations through selective breeding and assisted evolution of resilience (*97*, *98*). However, current knowledge of (direct or indirect) biological impacts of MHWs is lacking for many species and populations, which limits our capacity to safeguard marine resources against future extreme climatic events (*99*). Moreover, understanding of how MHWs interact with other stressors, such as oxygen, nutrient, and light availability, to affect marine organisms, populations, and ecosystems remains limited. MHWs have intensified over the past century largely due to anthropogenic climate change and will continue to do so as warming intensifies further, with wide-ranging consequences for marine ecosystems globally (*11*). There is a pressing need to develop a toolbox of adaptation and mitigation measures, including improved climatological forecasting, proactive resource management, and enhanced resilience, in response to increasing MHW impacts in the coming decades.

REFERENCES AND NOTES

1. G. T. Pecl *et al*., Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science* **355**, eaai9214 (2017). doi: 10.1126/science.aai9214; pmid: 28360268
2. C. Nolan *et al*., Past and future global transformation of terrestrial ecosystems under climate change. *Science* **361**, 920–923 (2018). doi: 10.1126/science.aan5360; pmid: 30166491
3. Intergovernmental Panel on Climate Change, "Climate change 2021: The physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change," V. Masson-Delmotte *et al*., Eds. (Cambridge Univ. Press, 2021); https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Full_Report_smaller.pdf.
4. L. Wu *et al*., Enhanced warming over the global subtropical western boundary currents. *Nat. Clim. Chang.* **2**, 161 (2012). doi: 10.1038/nclimate1353
5. E. S. Poloczanska *et al*., Global imprint of climate change on marine life. *Nat. Clim. Chang.* **3**, 919–925 (2013). doi: 10.1038/nclimate1958
6. A. Vergés *et al*., The tropicalization of temperate marine ecosystems: Climate-mediated changes in herbivory and community phase shifts. *Proc. Biol. Sci.* **281**, 20140846 (2014). doi: 10.1098/rspb.2014.0846; pmid: 25009065
7. E. C. J. Oliver *et al*., Longer and more frequent marine heatwaves over the past century. *Nat. Commun.* **9**, 1324 (2018). doi: 10.1038/s41467-018-03732-9; pmid: 29636482
8. A. Sen Gupta *et al*., Drivers and impacts of the most extreme marine heatwaves events. *Sci. Rep.* **10**, 19359 (2020). doi: 10.1038/s41598-020-75445-3; pmid: 33168858
9. N. J. Holbrook *et al*., A global assessment of marine heatwaves and their drivers. *Nat. Commun.* **10**, 2624 (2019). doi: 10.1038/s41467-019-10206-z; pmid: 31201309
10. E. C. J. Oliver *et al*., Projected marine heatwaves in the 21st century and the potential for ecological impact. *Front. Mar. Sci.* **6**, 734 (2019). doi: 10.3389/fmars.2019.00734
11. M. Collins, M. Sutherland, L. Bouwer, "Extremes, abrupt changes and managing risk," in: *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*, H. O. Portner. D. C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, N.M. Weyer, Eds. (IPCC, 2019), pp. 589–656.; https://www.ipcc.ch/site/assets/uploads/sites/3/2019/11/10_SROCC_Ch06_FINAL.pdf.
12. N. Arafeh-Dalmau *et al*., Extreme marine heatwaves alter kelp forest community near its equatorward distribution limit. *Front. Mar. Sci.* **6**, 499 (2019). doi: 10.3389/fmars.2019.00499
13. M. S. Thomsen *et al*., Local extinction of bull kelp (*Durvillaea* spp.) due to a marine heatwave. *Front. Mar. Sci.* **6**, 84 (2019). doi: 10.3389/fmars.2019.00084
14. L. Rogers-Bennett, C. A. Catton, Marine heat wave and multiple stressors tip bull kelp forest to sea urchin barrens. *Sci. Rep.* **9**, 15050 (2019). doi: 10.1038/s41598-019-51114-y; pmid: 31636286
15. T. Wernberg *et al*., Climate-driven regime shift of a temperate marine ecosystem. *Science* **353**, 169–172 (2016). doi: 10.1126/science.aad8745; pmid: 27387951
16. S. D. Donner, G. J. M. Rickbeil, S. F. Heron, A new, high-resolution global mass coral bleaching database. *PLOS ONE* **12**, e0175490 (2017). doi: 10.1371/journal.pone.0175490; pmid: 28445534
17. A. P. Krishnan *et al*., Elevated sea surface temperature during May 2010 induces mass bleaching of corals in the Andaman. *Curr. Sci.* **100**, 111–117 (2011); www.currentscience.ac.in/Volumes/100/01/0111.pdf.
18. J. A. Y. Moore *et al*., Unprecedented mass bleaching and loss of coral across 12° of latitude in Western Australia in 2010-11. *PLOS ONE* **7**, e51807 (2012). doi: 10.1371/journal.pone.0051807; pmid: 23284773
19. C. Cerrano *et al*., A catastrophic mass-mortality episode of gorgonians and other organisms in the Ligurian Sea (North-western Mediterranean), summer 1999. *Ecol. Lett.* **3**, 284–293 (2000). doi: 10.1046/j.1461-0248.2000.00152.x
20. M. W. Fraser *et al*., Extreme climate events lower resilience of foundation seagrass at edge of biogeographical range. *J. Ecol.* **102**, 1528–1536 (2014). doi: 10.1111/1365-2745.12300
21. A. Arias-Ortiz *et al*., A marine heatwave drives massive losses from the world's largest seagrass carbon stocks. *Nat. Clim. Chang.* **8**, 338–344 (2018). doi: 10.1038/s41558-018-0096-y
22. T. Gammelsrød, C. H. Bartholomae, D. C. Boyer, V. L. L. Filipe, M. J. O'Toole, Intrusion of warm surface water along the Angolan-Namibian coast in February-March 1995: The 1995 Benguela Nino. *S. Afr. J. Mar. Sci.* **19**, 41–56 (1998). doi: 10.2989/025776198784126719
23. T. Jones *et al*., Massive mortality of a planktivorous seabird in response to a marine heatwave. *Geophys. Res. Lett.* **45**, 3193–3202 (2018). doi: 10.1002/2017GL076164
24. J. Garrabou *et al*., Collaborative database to track mass mortality events in the Mediterranean Sea. *Front. Mar. Sci.* **6**, 707 (2019). doi: 10.3389/fmars.2019.00707
25. J. F. Piatt *et al*., Extreme mortality and reproductive failure of common murres resulting from the northeast Pacific marine heatwave of 2014-2016. *PLOS ONE* **15**, e0226087 (2020). doi: 10.1371/journal.pone.0226087; pmid: 31940310
26. S. P. Salazar, R. H. Bustamante, Effects of the 1997-1998 El Niño on population size and diet of the Galápagos sea lion (*Zalophus wollebaeki*). *Not. Galapagos* **62**, 40–45 (2003).
27. M. Ñiquen, M. Bouchon, Impact of El Niño events on pelagic fisheries in Peruvian waters. *Deep. Res. Part II Top. Stud. Oceanogr.* **51**, 563–574 (2004). doi: 10.1016/j.dsr2.2004.03.001
28. L. M. Cavole *et al*., Biological impacts of the 2013–2015 warm-water anomaly in the northeast Pacific: Winners, losers, and the future. *Oceanography (Wash. D.C.)* **29**, 273–285 (2016). doi: 10.5670/oceanog.2016.32
29. S. J. Barbeaux, K. Holsman, S. Zador, Marine heatwave stress test of ecosystem-based fisheries management in the Gulf of Alaska Pacific cod fishery. *Front. Mar. Sci.* **7**, 703 (2020). doi: 10.3389/fmars.2020.00703
30. D. A. Smale *et al*., Marine heatwaves threaten global biodiversity and the provision of ecosystem services. *Nat. Clim. Chang.* **9**, 306–312 (2019). doi: 10.1038/s41558-019-0412-1
31. R. C. J. Lenanton, C. E. Dowling, K. A. Smith, D. V. Fairclough, G. Jackson, Potential influence of a marine heatwave on range extensions of tropical fishes in the eastern Indian Ocean—Invaluable contributions from amateur observers. *Reg. Stud. Mar. Sci.* **13**, 19–31 (2017). doi: 10.1016/j.rsma.2017.03.005
32. C. Bouchard *et al*., Climate warming enhances polar cod recruitment, at least transiently. *Prog. Oceanogr.* **156**, 121–129 (2017). doi: 10.1016/j.pocean.2017.06.008
33. M. LeBlanc *et al*., Pelagic production and the recruitment of juvenile polar cod *Boreogadus saida* in Canadian Arctic seas. *Polar Biol.* **43**, 1043–1054 (2020). doi: 10.1007/s00300-019-02565-6
34. J. M. Sunday *et al*., Species traits and climate velocity explain geographic range shifts in an ocean-warming hotspot. *Ecol. Lett.* **18**, 944–953 (2015). doi: 10.1111/ele.12474; pmid: 26189556
35. K. E. Mills *et al*., Fisheries management in a changing climate. *Oceanography (Wash. D.C.)* **26**, 191–195 (2013). doi: 10.5670/oceanog.2013.27
36. N. Caputi *et al*., Management adaptation of invertebrate fisheries to an extreme marine heat wave event at a global warming hot spot. *Ecol. Evol.* **6**, 3583–3593 (2016). doi: 10.1002/ece3.2137; pmid: 28725352
37. C. J. Brown, C. Mellin, G. J. Edgar, M. D. Campbell, R. D. Stuart-Smith, Direct and indirect effects of heatwaves on a coral reef fishery. *Glob. Chang. Biol.* **27**, 1214–1225 (2021). doi: 10.1111/gcb.15472; pmid: 33340216
38. V. L. Trainer *et al*., Pelagic harmful algal blooms and climate change: Lessons from nature's experiments with extremes.

Harmful Algae **91**, 101591 (2020). doi: 10.1016/j.hal.2019.03.009; pmid: 32057339
39. N. J. Holbrook *et al*., Keeping pace with marine heatwaves. *Nat. Rev. Earth Environ.* **1**, 482–493 (2020). doi: 10.1038/s43017-020-0068-4
40. R. De Groot, B. Fisher, M. Christie, "Integrating the ecological and economic dimensions in biodiversity and ecosystem service valuation," in *The Economics of Ecosystems and Biodiversity: The Ecological and Economic Foundations*, P. Kumar, Ed. (Cambridge Univ. Press, 2010); pp. 9–40.
41. A. J. Hobday *et al*., A hierarchical approach to defining marine heatwaves. *Prog. Oceanogr.* **141**, 227–238 (2016). doi: 10.1016/j.pocean.2015.12.014
42. A. J. Hobday *et al*., Categorizing and naming marine heatwaves. *Oceanography (Wash. D.C.)* **31**, 162–173 (2018). doi: 10.5670/oceanog.2018.205
43. D. B. Enfield, Evolution and historical perspective of the 1997-1998 El Niño-Southern Oscillation event. *Bull. Mar. Sci.* **69**, 7–25 (2001).
44. N. Marbà, C. M. Duarte, Mediterranean warming triggers seagrass (*Posidonia oceanica*) shoot mortality. *Glob. Change Biol.* **16**, 2366–2375 (2010). doi: 10.1111/j.1365-2486.2009.02130.x
45. L. C. Rodrigues *et al*., Sensitivity of Mediterranean Bivalve Mollusc Aquaculture to Climate Change, Ocean Acidification, and Other Environmental Pressures: Findings from a Producer Survey. *J. Shellfish Res.* **34**, 1161–1176 (2015). doi: 10.2983/035.034.0341
46. E. Di Lorenzo, N. Mantua, Multi-year persistence of the 2014/15 North Pacific marine heatwave. *Nat. Clim. Chang.* **6**, 1042–1047 (2016). doi: 10.1038/nclimate3082
47. L. A. Rogers, M. T. Wilson, J. T. Duffy-Anderson, D. G. Kimmel, J. F. Lamb, Pollock and "the Blob": Impacts of a marine heatwave on walleye pollock early life stages. *Fish. Oceanogr.* **30**, 142–158 (2020). doi: 10.1111/fog.12508
48. S. J. Barbeaux T. A'mar, W. A. Palsson, "Assessment of the Pacific cod stock in the Gulf of Alaska," in: *North Pacific Fishery Management Council Gulf of Alaska Stock Assessment and Fishery Evaluation Report* (NPFMC, 2020); https://www.researchgate.net/publication/311733967_Assessment_of_the_Pacific_cod_stock_in_the_Gulf_of_Alaska.
49. R. M. McCabe *et al*., An unprecedented coastwide toxic algal bloom linked to anomalous ocean conditions. *Geophys. Res. Lett.* **43**, 10366–10376 (2016). doi: 10.1002/2016GL070023; pmid: 27917011
50. T. Wernberg *et al*., An extreme climatic event alters marine ecosystem structure in a global biodiversity hotspot. *Nat. Clim. Chang.* **3**, 78–82 (2013). doi: 10.1038/nclimate1627
51. N. Caputi *et al*., Factors affecting the recovery of invertebrate stocks from the 2011 Western Australian extreme marine heatwave. *Front. Mar. Sci.* **6**, 484 (2019). doi: 10.3389/fmars.2019.00484
52. K. Sgueglia, "Too much Maine lobster prompts protests and court order in Canada," *CNN*, 10 August 2012; https://edition.cnn.com/2012/08/09/us/canada-lobster-injuction/index.html).
53. T. P. Hughes *et al*., Global warming transforms coral reef assemblages. *Nature* **556**, 492–496 (2018). doi: 10.1038/s41586-018-0041-2; pmid: 29670282
54. A. Himes-Cornell, L. Pendleton, P. Atiyah, Valuing ecosystem services from blue forests: A systematic review of the valuation of salt marshes, sea grass beds and mangrove forests. *Ecosyst. Serv.* **30**, 36–48 (2018). doi: 10.1016/j.ecoser.2018.01.006
55. K. Filbee-Dexter *et al*., Marine heatwaves and the collapse of marginal North Atlantic kelp forests. *Sci. Rep.* **10**, 13388 (2020). doi: 10.1038/s41598-020-70273-x; pmid: 32770015
56. Te Rūnanga o Kaikōura, "Te Poha o Tohu Raumati (Environmental Management Plan)" (Te Rūnanga o Kaikōura, 2007); https://ngaitahu.iwi.nz/wp-content/uploads/2013/08/Te-Runanga-o-Kaikoura-Environmental-Management-Plan.pdf.
57. S. Strydom *et al*., Too hot to handle: Unprecedented seagrass death driven by marine heatwave in a World Heritage Area. *Glob. Chang. Biol.* **26**, 3525–3538 (2020). doi: 10.1111/gcb.15065; pmid: 32129909
58. L. R. Aoki *et al*., Seagrass recovery following marine heat wave influences sediment carbon stocks. *Front. Mar. Sci.* **7**, 576784 (2021). doi: 10.3389/fmars.2020.576784
59. 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
60. T. P. Hughes *et al*., Coral reefs in the Anthropocene. *Nature* **546**, 82–90 (2017). doi: 10.1038/nature22901; pmid: 28569801
61. C. M. Eakin, H. P. A. Sweatman, R. E. Brainard, The 2014–2017 global-scale coral bleaching event: Insights and impacts. *Coral Reefs* **38**, 539–545 (2019). doi: 10.1007/s00338-019-01844-2
62. M. Spalding *et al*., Mapping the global value and distribution of coral reef tourism. *Mar. Policy* **82**, 104–113 (2017). doi: 10.1016/j.marpol.2017.05.014
63. S. Bennett *et al*., The "Great Southern Reef": Social, ecological and economic value of Australia's neglected kelp forests. *Mar. Freshw. Res.* **67**, 47–56 (2016). doi: 10.1071/MF15232
64. National Marine Fisheries Service, Fisheries Economics of the United States, 2016 (NOAA, 2018); https://www.fisheries.noaa.gov/resource/document/fisheries-economics-united-states-report-2016.
65. A. Clement *et al*., Exceptional summer conditions and HABs of Pseudochattonella in Southern Chile create record impacts on salmon farms. *Harmful Algae News* **53**, 1–3 (2016); www.e-pages.dk/ku/1217/.
66. B. E. Ferriss, D. J. Marcinek, D. Ayres, J. Borchert, K. A. Lefebvre, Acute and chronic dietary exposure to domoic acid in recreational harvesters: A survey of shellfish consumption behavior. *Environ. Int.* **101**, 70–79 (2017). doi: 10.1016/j.envint.2017.01.006; pmid: 28109640
67. S. K. Moore *et al*., An index of fisheries closures due to harmful algal blooms and a framework for identifying vulnerable fishing communities on the U.S. West Coast. *Mar. Policy* **110**, 103543 (2019). doi: 10.1016/j.marpol.2019.103543
68. J. A. Santora *et al*., Habitat compression and ecosystem shifts as potential links between marine heatwave and record whale entanglements. *Nat. Commun.* **11**, 536 (2020). doi: 10.1038/s41467-019-14215-w; pmid: 31988285
69. M. I. Curnock *et al*., Shifts in tourists' sentiments and climate risk perceptions following mass coral bleaching of the Great Barrier Reef. *Nat. Clim. Chang.* **9**, 535–541 (2019). doi: 10.1038/s41558-019-0504-y
70. R. Costanza *et al*., Changes in the global value of ecosystem services. *Glob. Environ. Change.* **26**, 152–158 (2014). doi: 10.1016/j.gloenvcha.2014.04.002
71. Food and Agriculture Organization of the United Nations, "FACT SHEET: The international fish trade and world fisheries" (FAO, 2010); http://www.fao.org/fileadmin/user_upload/newsroom/docs/fact_sheet_fish_trade_en.pdf.
72. P. Fréon *et al*., Harvesting for food versus feed: A review of Peruvian fisheries in a global context. *Rev. Fish Biol. Fish.* **24**, 381–398 (2014). doi: 10.1007/s11160-013-9336-4
73. A. Doshi *et al*., "Loss of economic value from coral bleaching in SE Asia," in 12th International Coral Reef Symposium, Cairns, Australia, 9–13 July 2012; https://www.academia.edu/12494037/Loss_of_economic_value_from_coral_bleaching_in_SE_Asia.
74. E. C. J. Oliver *et al*., The unprecedented 2015/16 Tasman Sea marine heatwave. *Nat. Commun.* **8**, 16101 (2017). doi: 10.1038/ncomms16101; pmid: 28706247
75. I. Catizone, "National impact from Tasmanian POMS outbreak" (Fisheries Research and Development Corporation, 2016); https://www.frdc.com.au/fish-vol-24-2/national-impact-tasmanian-poms-outbreak.
76. K. Filbee-Dexter, T. Wernberg, Substantial blue carbon in overlooked Australian kelp forests. *Sci. Rep.* **10**, 12341 (2020). doi: 10.1038/s41598-020-69258-7
77. A. Pershing, K. Mills, A. Dayton, B. Franklin, B. Kennedy, Evidence for adaptation from the 2016 marine heatwave in the Northwest Atlantic Ocean. *Oceanography (Wash. D.C.)* **31**, 152–161 (2018). doi: 10.5670/oceanog.2018.213
78. F. Bever, "Maine fishermen prepare for losses and gains in a climate changed ocean," *WBUR*, 23 September 2019; https://www.wbur.org/news/2019/09/23/maine-fishermen-prepare-for-losses-and-gains-in-a-climate-changed-ocean.
79. Oxford Business Group, "Peru Agriculture: Diversification efforts are up as instability over anchovy supply continues" (Oxford Business Group, 2016); https://oxfordbusinessgroup.com/analysis/choppy-waters-diversification-efforts-are-instability-over-anchovy-supply-continues.
80. T. Yusuke, N. Masato, "Marine heatwaves driving change in Hokkaido's fishing industry," *NHK World-Japan*, 10 December 2020; https://www3.nhk.or.jp/nhkworld/en/news/backstories/1404/.
81. Y. Liu, S. I. Saitoh, H. Igarashi, T. Hirawake, The regional impacts of climate change on coastal environments and the aquaculture of Japanese scallops in northeast Asia: Case studies from Dalian, China, and Funka Bay, Japan. *Int. J. Remote Sens.* **35**, 4422–4440 (2014). doi: 10.1080/01431161.2014.916435
82. H. T. Sandersen, J. Olsen, G. K. Hovelsrud, A. Gjertsen, "Climate change and Norwegian arctic aquaculture: Perception, relevance, and adaptation," in *Arctic Yearbook 2020*, L. Heininen, H. Exner-Pirot, J. Barnes, Eds. (2020); pp. 316–327.
83. L. Lebel *et al*., Innovation, practice, and adaptation to climate in the aquaculture sector. *Rev. Fish. Sci. Aquacult.* (2020); doi: 10.1080/23308249.2020.1869695
84. A. Le Bris *et al*., Climate vulnerability and resilience in the most valuable North American fishery. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 1831–1836 (2018). doi: 10.1073/pnas.1711122115; pmid: 29358389
85. L. Bellquist, V. Saccomanno, B. X. Semmens, M. Gleason, J. Wilson, The rise in climate change-induced federal fishery disasters in the United States. *PeerJ* **9**, e11186 (2021). doi: 10.7717/peerj.11186; pmid: 33981495
86. M. A. Coleman, T. Wernberg, The silver lining of extreme events. *Trends Ecol. Evol.* **35**, 1065–1067 (2020). doi: 10.1016/j.tree.2020.08.013; pmid: 32958366
87. T. Wernberg *et al*., Genetic diversity and kelp forest vulnerability to climatic stress. *Sci. Rep.* **8**, 1851 (2018). doi: 10.1038/s41598-018-20009-9; pmid: 29382916
88. M. A. Coleman, A. J. P. Minne, S. Vranken, T. Wernberg, Genetic tropicalisation following a marine heatwave. *Sci. Rep.* **10**, 12726 (2020). doi: 10.1038/s41598-020-69665-w; pmid: 32728196
89. C. F. D. Gurgel, O. Camacho, A. J. P. Minne, T. Wernberg, M. A. Coleman, Marine heatwave drives cryptic loss of genetic diversity in underwater forests. *Curr. Biol.* **30**, 1199–1206.e2 (2020). doi: 10.1016/j.cub.2020.01.051; pmid: 32109397
90. S. Vranken *et al*., Genotype-Environment mismatch of kelp forests under climate change. *Mol. Ecol.* **30**, 3730–3746 (2021). doi: 10.1111/mec.15993; pmid: 34018645
91. J. R. Guest *et al*., Contrasting patterns of coral bleaching susceptibility in 2010 suggest an adaptive response to thermal stress. *PLOS ONE* **7**, e33353 (2012). doi: 10.1371/journal.pone.0033353; pmid: 22428027
92. A. E. Bates *et al*., Climate resilience in marine protected areas and the 'Protection Paradox'. *Biol. Conserv.* **236**, 305–314 (2019). doi: 10.1016/j.biocon.2019.05.005
93. J. Verdura *et al*., Local-scale climatic refugia offer sanctuary for a habitat-forming species during a marine heatwave. *J. Ecol.* **109**, 1758–1773 (2021). doi: 10.1111/1365-2745.13599
94. K. C. Cavanaugh, D. C. Reed, T. W. Bell, M. C. N. Castorani, R. Beas-Luna, Spatial variability in the resistance and resilience of giant kelp in southern and Baja California to a multiyear heatwave. *Front. Mar. Sci.* **6**, 413 (2019). doi: 10.3389/fmars.2019.00413
95. M. S. Edwards, J. A. Estes, Catastrophe, recovery and range limitation in NE Pacific kelp forests: A large-scale perspective. *Mar. Ecol. Prog. Ser.* **320**, 79–87 (2006). doi: 10.3354/meps320079
96. R. M. Freedman, J. A. Brown, C. Caldow, J. E. Caselle, Marine protected areas do not prevent marine heatwave-induced fish community structure changes in a temperate transition zone. *Sci. Rep.* **10**, 21081 (2020). doi: 10.1038/s41598-020-77885-3; pmid: 33273514
97. M. J. H. van Oppen, J. K. Oliver, H. M. Putnam, R. D. Gates, Building coral reef resilience through assisted evolution. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 2307–2313 (2015). doi: 10.1073/pnas.1422301112; pmid: 25646461
98. M. A. Coleman *et al*., Restore or redefine: Future trajectories for restoration. *Front. Mar. Sci.* **7**, 237 (2020). doi: 10.3389/fmars.2020.00237
99. J. A. Benthuysen, E. C. J. Oliver, K. Chen, T. Wernberg, Editorial: Advances in understanding marine heatwaves and their impacts. *Front. Mar. Sci.* **7**, 147 (2020). doi: 10.3389/fmars.2020.00147
100. S. Zador, E. Yasumiishi, G. A. Whitehouse, eds., "Ecosystem Status Report 2019 Gulf of Alaska" (National Oceanic and Atmospheric Administration, 2019); https://www.fisheries.noaa.gov/resource/data/ecosystem-status-report-2019-gulf-alaska.
101. RStudio Team, "RStudio: Integrated development for R" (2020); https://www.rstudio.com/.
102. R. W. Schlegel, A. J. Smit, heatwaveR: A central algorithm for the detection of heatwaves and cold-spells. *J. Open Source Softw.* **3**, 821 (2018). doi: 10.21105/joss.00821
103. M. C. Fisher, S. K. Moore, S. L. Jardine, J. R. Watson, J. F. Samhouri, Climate shock effects and mediation in fisheries. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2014379117 (2021). doi: 10.1073/pnas.2014379117; pmid: 33397723

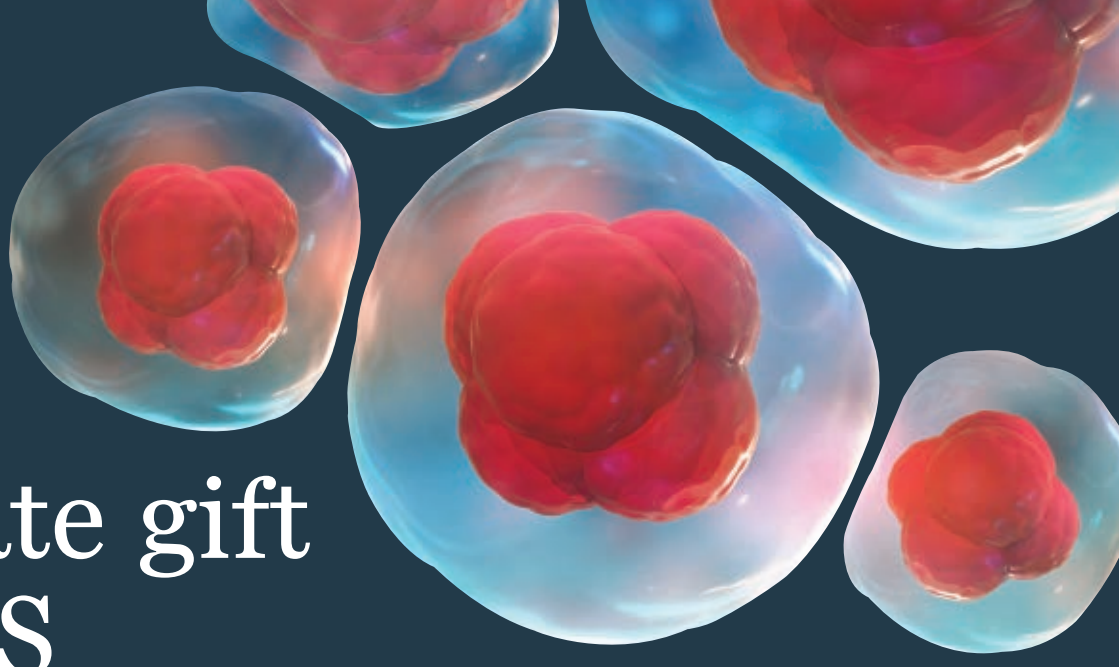
104. Australian Government Department of Aquaculture, Water and the Environment, "Australian fisheries and aquaculture statistics 2018" (ABARES, 2020); https://daff.ent.sirsidynix.net.au/client/en_AU/search/asset/1030241/0.
105. A. Chandrapavan, N. Caputi, M. I. Kangas, The decline and recovery of a crab population from an extreme marine heatwave and a changing climate. *Front. Mar. Sci.* **6**, 510 (2019). doi: 10.3389/fmars.2019.00510
106. C. A. Bost *et al*., Large-scale climatic anomalies affect marine predator foraging behaviour and demography. *Nat. Commun.* **6**, 8220 (2015). doi: 10.1038/ncomms9220; pmid: 26506134
107. R. Cristofari *et al*., Climate-driven range shifts of the king penguin in a fragmented ecosystem. *Nat. Clim. Chang.* **8**, 245–251 (2018). doi: 10.1038/s41558-018-0084-2
108. J. M. Alonso Vega, J. A. Vásquez, A. H. Buschmann, Population biology of the subtidal kelps *Macrocystis integrifolia* and *Lessonia trabeculata* (Laminariales, Phaeophyceae) in an upwelling ecosystem of northern Chile: Interannual variability and El Niño 1997-1998. *Rev. Chil. Hist. Nat.* **78**, 33–50 (2005). doi: 10.4067/S0716-078X2005000100004
109. F. Trillmich *et al*., On the challenge of interpreting census data: Insights from a study of an endangered pinniped. *PLOS ONE* **11**, e0154588 (2016). doi: 10.1371/journal.pone.0154588; pmid: 27148735
110. T. Spencer, K. A. Teleki, C. Bradshaw, M. D. Spalding, Coral bleaching in the Southern Seychelles during the 1997-1998 Indian Ocean warm event. *Mar. Pollut. Bull.* **40**, 569–586 (2000). doi: 10.1016/S0025-326X(00)00026-6
111. L. G. C. Genevier, T. Jamil, D. E. Raitsos, G. Krokos, I. Hoteit, Marine heatwaves reveal coral reef zones susceptible to bleaching in the Red Sea. *Glob. Chang. Biol.* **25**, 2338–2351 (2019). doi: 10.1111/gcb.14652; pmid: 30974020
112. A. Genin, L. Levy, G. Sharon, D. E. Raitsos, A. Diamant, Rapid onsets of warming events trigger mass mortality of coral reef fish. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 25378–25385 (2020). doi: 10.1073/pnas.2009748117; pmid: 32958634
113. C. N. Bianchi *et al*., Consequences of the marine climate and ecosystem shift of the 1980-90s on the Ligurian Sea biodiversity (NW Mediterranean). *Eur. Zool. J.* **86**, 458–487 (2019). doi: 10.1080/24750263.2019.1687765
114. R. Cupido *et al*., Sexual structure of a highly reproductive, recovering gorgonian population: Quantifying reproductive output. *Mar. Ecol. Prog. Ser.* **469**, 25–36 (2012). doi: 10.3354/meps09976
115. J. Garrabou *et al*., Mass mortality in Northwestern Mediterranean rocky benthic communities: Effects of the 2003 heat wave. *Glob. Change Biol.* **15**, 1090–1103 (2009). doi: 10.1111/j.1365-2486.2008.01823.x
116. E. Diaz-Almela, N. Marbà, C. M. Duarte, Consequences of Mediterranean warming events in seagrass (*Posidonia oceanica*) flowering records. *Glob. Change Biol.* **13**, 224–235 (2007). doi: 10.1111/j.1365-2486.2006.01260.x
117. C. M. Eakin *et al*., Caribbean corals in crisis: Record thermal stress, bleaching, and mortality in 2005. *PLOS ONE* **5**, e13969 (2010). doi: 10.1371/journal.pone.0013969; pmid: 21125021
118. Food and Agriculture Organization of the United Nations, "Climate change adaptation in the eastern Caribbean fisheries sector (CC4FISH) launching workshop, Barbados, 7–9 February 2017" (FAO Fisheries and Aquaculture Report No. 1189, 2017; http://www.fao.org/3/i8281en/I8281EN.pdf.
119. G. A. Kendrick *et al*., A systematic review of how multiple stressors from an extreme event drove ecosystem-wide loss of resilience in an iconic seagrass community. *Front. Mar. Sci.* **6**, 455 (2019). doi: 10.3389/fmars.2019.00455
120. D. A. Smale, T. Wernberg, Extreme climatic event drives range contraction of a habitat-forming species. *Proc. Biol. Sci.* **280**, 20122829 (2013). doi: 10.1098/rspb.2012.2829; pmid: 23325774
121. R. C. Babcock *et al*., Severe continental-scale impacts of climate change are happening now: Extreme climate events impact marine habitat forming communities along 45% of Australia's coast. *Front. Mar. Sci.* **6**, 411 (2019). doi: 10.3389/fmars.2019.00411
122. S. D. Roberts, P. D. Van Ruth, C. Wilkinson, S. S. Bastianello, M. S. Bansemer, Marine heatwave, harmful algae blooms and an extensive fish kill event during 2013 in South Australia. *Front. Mar. Sci.* **6**, 610 (2019). doi: 10.3389/fmars.2019.00610
123. C. D. Harvell *et al*., Disease epidemic and a marine heat wave are associated with the continental-scale collapse of a pivotal predator (*Pycnopodia helianthoides*). *Sci. Adv.* **5**, eaau7042 (2019). doi: 10.1126/sciadv.aau7042; pmid: 30729157
124. E. Sanford, J. L. Sones, M. García-Reyes, J. H. R. Goddard, J. L. Largier, Widespread shifts in the coastal biota of northern California during the 2014-2016 marine heatwaves. *Sci. Rep.* **9**, 4216 (2019). doi: 10.1038/s41598-019-40784-3; pmid: 30862867
125. R. M. Suryan *et al*., Ecosystem response persists after a prolonged marine heatwave. *Sci. Rep.* **11**, 6235 (2021). doi: 10.1038/s41598-021-83818-5; pmid: 33737519
126. S. M. Mendez, A. Martinez, L. Ortego, A. Fabre, High sea surface temperature, the potential trigger of mass mortality of fish, exceptional toxin producing HABs, and other socio-economic impacts in Uruguay. *Harmful Algae News* **51**, 1–2 (2015); www.e-pages.dk/ku/1141/.
127. B. Vargas-Ángel *et al*., El Niño-associated catastrophic coral mortality at Jarvis Island, central Equatorial Pacific. *Coral Reefs* **38**, 731–741 (2019). doi: 10.1007/s00338-019-01838-0
128. R. E. Brainard *et al*., Perspective, Special Supplement to the Bulletin of the American Meteorological Society, S. C. Herring, N. Christidis, A. Hoell, J. P. Kossin, C. J. Schreck III, P. A. Stott, Eds. (American Meteorological Society, 2018), vol. 99, pp. S21–S26; https://doi.org/10.1175/BAMS-ExplainingExtremeEvents2016.1.
129. M. Le Nohaïc *et al*., Marine heatwave causes unprecedented regional mass bleaching of thermally resistant corals in northwestern Australia. *Sci. Rep.* **7**, 14999 (2017). doi: 10.1038/s41598-017-14794-y; pmid: 29101362
130. K. Mathiesen, "Ocean heatwave destroys Tasmania's unique underwater jungle," *Climate Home News*, 14 October 2016; https://www.climatechangenews.com/2016/10/14/ocean-heatwave-destroys-tasmanias-unique-underwater-jungle/.
131. J. E. Walsh *et al*., 8. The high latitude marine heat wave of 2016 and its impacts on Alaska. *Bull. Am. Meteorol. Soc.* **99**, S39–S43 (2018). doi: 10.1175/BAMS-D-17-0105.1
132. G. A. Meehl *et al*., Sustained ocean changes contributed to sudden Antarctic sea ice retreat in late 2016. *Nat. Commun.* **10**, 14 (2019). doi: 10.1038/s41467-018-07865-9; pmid: 30600315
133. C. M. Eakin *et al*., Ding, dong, the witch is dead (?)— Three years of global coral bleaching 2014-2017. *Reef Encounter* **32**, 33–38 (2017); https://www.ncei.noaa.gov/data/oceans/coris/library/NOAA/CRCP/NESDIS/STAR/Project/915/Eakin2017_2014-2017_Coral_Bleaching.pdf.
134. T. P. Hughes *et al*., Global warming and recurrent mass bleaching of corals. *Nature* **543**, 373–377 (2017). doi: 10.1038/nature21707; pmid: 28300113
135. R. D. Stuart-Smith, C. J. Brown, D. M. Ceccarelli, G. J. Edgar, Ecosystem restructuring along the Great Barrier Reef following mass coral bleaching. *Nature* **560**, 92–96 (2018). doi: 10.1038/s41586-018-0359-9; pmid: 30046108
136. L. J. McKenzie *et al*., "Marine Monitoring Program: Annual report for inshore seagrass monitoring 2016-2017 (Great Barrier Reef Marine Park Authority, 2018); https://elibrary.gbrmpa.gov.au/jspui/handle/11017/3398.
137. A. Thompson *et al*., Australian Institute of Marine Science, "Annual Report for inshore coral reef monitoring 2016-2017" (Great Barrier Reef Marine Park Authority, 2018); https://elibrary.gbrmpa.gov.au/jspui/handle/11017/3397.
138. A. Thompson, P. Costello, J. Davidson, M. Logan, G. Coleman, Australian Institute of Marine Science, "Annual Report for inshore coral reef monitoring 2017-2018" (Great Barrier Reef Marine Park Authority, 2019); https://elibrary.gbrmpa.gov.au/jspui/handle/11017/3491.
139. M. J. Salinger *et al*., The unprecedented coupled ocean-atmosphere summer heatwave in the New Zealand region 2017/18: Drivers, mechanisms and impacts. *Environ. Res. Lett.* **14**, 044023 (2019). doi: 10.1088/1748-9326/ab012a
140. G. Gao *et al*., Drivers of marine heatwaves in the East China Sea and the South Yellow Sea in three consecutive summers during 2016–2018. *J. Geophys. Res. Oceans* **125**, 1–19 (2020). doi: 10.1029/2020JC016518
141. T. Singh, M. Iijima, K. Yasumoto, K. Sakai, Effects of moderate thermal anomalies on Acropora corals around Sesoko Island, Okinawa. *PLOS ONE* **14**, e0210795 (2019). doi: 10.1371/journal.pone.0210795; pmid: 30699163
142. Fish Information & Services, "Strong heatwave causes mortality in aquaculture farms," *Seafood Media Group*, 10 August 2017; https://seafood.media/fis/worldnews/worldnews.asp?l=e&id=93266&ndb=1.
143. N. Phongsuwan *et al*., Status and changing patterns on coral reefs in Thailand during the last two decades. *Deep. Res. Part II Top. Stud. Oceanogr.* **96**, 19–24 (2013). doi: 10.1016/j.dsr2.2013.02.015
144. G. Manta, S. de Mello, R. Trinchin, J. Badagian, M. Barreiro, The 2017 record marine heatwave in the Southwestern Atlantic shelf. *Geophys. Res. Lett.* **45**, 12449–12456 (2018). doi: 10.1029/2018GL081070
145. L. J. Raymundo *et al*., Successive bleaching events cause mass coral mortality in Guam, Micronesia. *Coral Reefs* **38**, 677–700 (2019). doi: 10.1007/s00338-019-01836-2
146. J. A. Burt, F. Paparella, N. Al-Mansoori, A. Al-Mansoori, H. Al-Jailani, Causes and consequences of the 2017 coral bleaching event in the southern Persian/Arabian Gulf. *Coral Reefs* **38**, 567–589 (2019). doi: 10.1007/s00338-019-01767-y
147. D. Y. Ige, S. D. Case, "11/5/19 – Coral bleaching not as severe as predicted but still widespread; extensive surveys show bleaching event now abating" (Department of Land and Natural Resources, 2019); https://dlnr.hawaii.gov/blog/2019/11/05/nr19-186/.
148. E. Rardon, "Rare tropical sea turtle rescued from Alberni Inlet," *Alberni Valley News*, 3 October 2019; https://www.albernivalleynews.com/news/tropical-sea-turtle-rescued-from-alberni-inlet/.
149. D. Hammock, "Commercial Dungeness crab season opener delayed through at least mid-December," *The Daily World*, 30 November 2020; https://www.thedailyworld.com/news/commercial-dungeness-crab-season-opener-delayed-through-at-least-mid-december/.
150. J. L. Boldt, A. Javorski, P. C. Chandler, *State of the Physical, Biological and Selected Fishery Resources of Pacific Canadian Marine Ecosystems in 2019: Canadian Technical Report of Fisheries and Aquatic Sciences 377* (Fisheries and Oceans Canada, 2020); https://publications.gc.ca/collections/collection_2020/mpo-dfo/Fs97-6-3377-eng.pdf.

ACKNOWLEDGMENTS

We thank the Coastal Observation and Seabird Survey Team (https://COASST.org) for providing the seabird mortality data presented in Fig. 2. **Funding:** D.A.S. was supported by a UK Research and Innovation (UKRI) Future Leaders Fellowship (MR/S032827/1). T.W. was supported by the Australian Research Council (DP200100201). M.T. was supported by The Brian Mason Trust, New Zealand. P.J.M. and M.T.B. were supported by NERC-Newton Fund grant NE/S011692/2. **Author contributions:** A.S.G., A.J.H., M.T.B., M.T., P.J.M., T.W., and D.A.S conceived the research. K.E.S. conducted the systematic analysis of the literature. K.E.S. and D.A.S. wrote the original draft of the manuscript. All authors contributed to reviewing and editing the manuscript. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

10.1126/science.abj3593

1848 SOCIETY
AAAS



An estate gift to AAAS

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

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

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

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

—Prof. Elisabeth Ervin-Blankenheim, 1848 Society member

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

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

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

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

cut here

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

1848 SOCIETY
AAAS

PLEASE CONTACT ME AT:

Name: _____

Address: _____

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

Email: _____ Phone: _____

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

RESEARCH ARTICLE SUMMARY

CORONAVIRUS

Low-dose mRNA-1273 COVID-19 vaccine generates durable memory enhanced by cross-reactive T cells

Jose Mateus, Jennifer M. Dan†, Zeli Zhang†, Carolyn Rydyznski Moderbacher, Marshall Lammers, Benjamin Goodwin, Alessandro Sette*, Shane Crotty*, Daniela Weiskopf*

INTRODUCTION: Understanding human immune responses to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) RNA vaccines is of interest for a panoply of reasons. mRNA vaccines have demonstrated impressive protection against COVID-19, but the durability of immunity has been a major unknown. Moreover, a better understanding of age-associated differences and mRNA vaccine dose response curves for dose sparing considerations is needed. Additionally, the impact of preexisting cross-reactive memory on immune responses to SARS-CoV-2 proteins remains an open question. Cross-reactive memory $CD4^+$ T cells recognizing SARS-CoV-2 have been found in ~50% of individuals. A vaccine trial is a controlled context for testing the relevance of such cross-reactive T cells. Each of these topics was addressed in this study using blood samples from a National Institutes of Health clinical trial of 25-µg mRNA-1273 COVID-19 vaccinees as well as from 100-µg mRNA-1273 COVID-19 vaccinees and SARS-CoV-2–infected individuals.

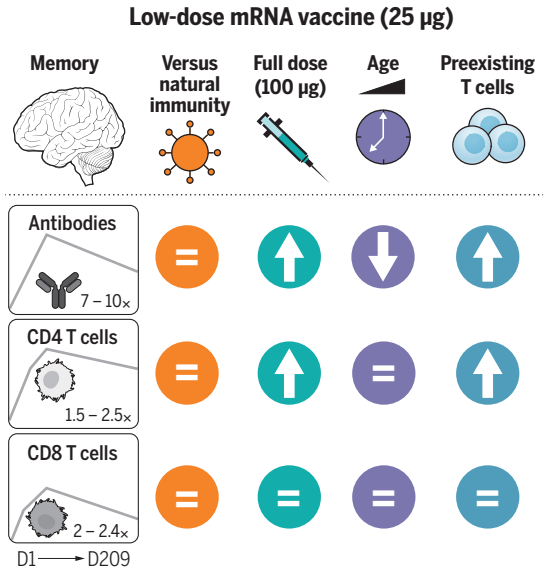
RATIONALE: Vaccination and infection are two different paths to immunity. Comparison of vaccine-generated and infection-generated immune memory is of value. Given evidence that antibodies, $CD4^+$ T cells, and $CD8^+$ T cells can each participate in protective immunity against COVID-19, we measured acute and memory SARS-CoV-2 spike–specific antibodies, $CD4^+$ T cells, and $CD8^+$ T cells in the blood of subjects who received a low-dose (25 µg) or standard-dose (100 µg) mRNA-1273 COVID-19 vaccine. Immunological measurements were used to address the four issues described above: namely, the durability of immune memory over 7 months after vaccination, mRNA vaccine dose responses, age differences, and the impact of preexisting cross-reactive T cells.

RESULTS: Longitudinal samples from 35 volunteers immunized with 25 µg of mRNA-1273 on days 1 and 29 were used to measure SARS-CoV-2 spike–binding antibodies, receptor binding domain (RBD)–binding antibodies, SARS-CoV-2 pseudovirus (PSV) neutralizing antibodies, spike-specific $CD4^+$ T cells, and spike-specific $CD8^+$ T cells. Overall, substantial anti-spike, anti-RBD, and PSV neutralizing antibodies were induced in response to two 25-µg mRNA-1273 vaccinations, were maintained in 88 to 100% of vaccinees for at least 6 months after the second immunization, and were comparable in magnitude and quality to those observed 6 to 7 months after infection with SARS-CoV-2.

Spike-specific $CD4^+$ T cells were generated by low-dose mRNA-1273 and were maintained as memory $CD4^+$ T cells. We observed strong T follicular helper (T_{FH}) and type 1 T helper cell polarization of these cells, which is advantageous for antiviral immunity. Spike-specific $CD8^+$ T cells were detectable in 88% of vaccinees and maintained for at least 6 months in 67% of vaccinees. Spike-specific $CD4^+$ or $CD8^+$ T cell frequencies were not lower in older vaccinee groups than in 18- to 55-year-olds, either in the acute or memory phase. Thus, 25-µg mRNA-1273 vaccination induced spike antibody levels and memory T cell frequencies at 7 months after vaccination similar to those observed for COVID-19 cases 7 months after symptom onset.

Next, to assess the impact of mRNA dosing, we compared immune responses between 25-µg and 100-µg doses of mRNA-1273 vaccine. Peak anti-spike, anti-RBD, and PSV neutralizing antibody levels were about twofold higher in 100-µg vaccinees than in 25-µg vaccinees. Spike-specific $CD4^+$ T cells responses were ~1.4-to-2.0-fold higher in 100-µg vaccinees, whereas peak $CD8^+$ T cell responses were comparable between 25-µg and 100-µg dose regimens.

Finally, to address potential positive or negative effects of preexisting cross-reactive memory T cells, we compared 25-µg mRNA-1273 COVID-19 vaccine responses between subjects with or without measurable preexisting SARS-CoV-2 spike–reactive memory $CD4^+$ T cells. Preexisting immunity enhanced vaccine antibody responses after a single vaccine dose, which was associated with higher spike-specific T_{FH} cells and total spike-specific $CD4^+$ T cell responses. Individuals with preexisting cross-reactive memory T cells also sustained higher SARS-CoV-2–neutralizing antibodies 6 months after vaccination.



Response to low-dose mRNA-1273 vaccination over 7 months. Immunological memory of antibodies, $CD4^+$ T cells, and $CD8^+$ T cells was examined after low-dose mRNA vaccination. Levels of spike-specific immune memory were then compared to immune memory observed after natural infection with SARS-CoV-2 or after full-dose vaccination. Robust immune memory comparable to natural infection but lower than after full-dose vaccination was observed. Increased vaccinee age correlated with reduced antibody levels but had no effect on cellular immune memory. Immune memory was enhanced by preexisting cross-reactive T cells. D1, day 1.

CONCLUSION: The 25-µg dose of mRNA-1273 vaccine induces durable and functional T cell and antibody memory at comparable magnitude to natural infection. This work expands our understanding of immune memory to mRNA vaccine in humans, vaccine dose sparing, and possible timing of boosters. Finally, these data provide evidence that cross-reactive memory $CD4^+$ T cells are biologically relevant and can exert a considerable positive influence on immunity generated by vaccination, with potential implications for vaccines and SARS-CoV-2 infections. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: alex@lji.org (A.S.); shane@lji.org (S.C.); daniela@lji.org (D.W.)
†These authors contributed equally to this work.
Cite this article as J. Mateus *et al.*, *Science* **374**, eabj9853 (2021). DOI: 10.1126/science.abj9853
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.

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abj9853

RESEARCH ARTICLE

CORONAVIRUS

Low-dose mRNA-1273 COVID-19 vaccine generates durable memory enhanced by cross-reactive T cells

Jose Mateus[1], Jennifer M. Dan[1,2]†, Zeli Zhang[1]†, Carolyn Rydyznski Moderbacher[1], Marshall Lammers[1], Benjamin Goodwin[1], Alessandro Sette[1,2]*, Shane Crotty[1,2]*, Daniela Weiskopf[1]*

Vaccine-specific CD4$^+$ T cell, CD8$^+$ T cell, binding antibody, and neutralizing antibody responses to the 25-μg Moderna messenger RNA (mRNA)–1273 vaccine were examined over the course of 7 months after immunization, including in multiple age groups, with a particular interest in assessing whether preexisting cross-reactive T cell memory affects vaccine-generated immunity. Vaccine-generated spike-specific memory CD4$^+$ T cells 6 months after the second dose of the vaccine were comparable in quantity and quality to COVID-19 cases, including the presence of T follicular helper cells and interferon-γ–expressing cells. Spike-specific CD8$^+$ T cells were generated in 88% of subjects, with equivalent memory at 6 months post-boost compared with COVID-19 cases. Lastly, subjects with preexisting cross-reactive CD4$^+$ T cell memory exhibited stronger CD4$^+$ T cell and antibody responses to the vaccine, demonstrating the biological relevance of severe acute respiratory syndrome coronavirus 2–cross-reactive CD4$^+$ T cells.

Understanding human immune responses to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) RNA vaccines is of interest for a panoply of reasons. mRNA vaccines have demonstrated impressive protection against COVID-19 (*1*–*7*). The COVID-19 vaccine mRNA-1273, developed as a collaboration between Moderna and the National Institutes of Health Vaccine Research Center, encodes a stabilized SARS-CoV-2 full-length spike (*8*, *9*). Durability of immunity has been, and remains, a major unknown for mRNA vaccines in humans. Encouraging reports from both Pfizer-BioNTech and Moderna indicate protective immunity of 91 and 93%, respectively, over the 6-month period after the second immunization (7 months after the first immunization) (*10*, *11*), down modestly from the 95% maximal protection observed for each of those two vaccines within 1 to 2 months after the second immunization (*1*, *2*). Although neutralizing antibodies are a clear correlate of immunity after two immunizations (*12*), the underlying mechanisms remain unclear. Moreover, those mechanisms of immunity may change as the immune response develops (e.g., after a single immunization) or as immune memory changes composition (*13*–*15*). Direct measurements of immune memory compartments in humans are necessary to provide insights into these important topics.

Infection and vaccination are two different paths to immunity. Comparison of vaccine-generated immune memory with immune memory of persons infected with SARS-CoV-2 is of value, as studies have indicated that natural immunity is 93 to 100% protective against symptomatic reinfection for 7 to 8 months (*16*–*19*), although natural immunity protection against certain variants of concern (VOCs) is likely to be lower (*20*). After SARS-CoV-2 infection, immunological memory has been observed for ≥8 months for CD4$^+$ T cells, CD8$^+$ T cells, memory B cells, and antibodies (*21*, *22*). The immune memory in response to SARS-CoV-2 infection exhibits a relatively gradual decline that partially stabilizes within 1 year (*23*–*26*). The 100-μg mRNA-1273 vaccination has been shown to induce durable antibody responses (*27*), but it is unknown whether immune memory to the mRNA-1273 vaccine months after immunization is similar to or different than memory generated by SARS-CoV-2 infection. Additionally, both 25-μg- and 100-μg-dose mRNA-1273 vaccinations have been tested in clinical trials (*9*, *28*), with 100-μg mRNA-1273 proceeding toward licensure (*2*, *29*).

Preexisting cross-reactive memory CD4$^+$ T cells that recognize SARS-CoV-2 have been found in ~50% of individuals pre-pandemic (*30*–*37*). There has been intense interest in understanding whether these preexisting cross-reactive memory CD4$^+$ T cells, identified in vitro, are biologically relevant in vivo (*33*, *38*, *39*). One approach to test the relevance of such T cells in a controlled fashion is in the context of a vaccine trial, as individuals in a clinical trial are all exposed to a well-defined dose of antigen at a specific time. Additionally, exposure to a low antigen dose may be more sensitive to influence by cross-reactive memory. Thus, we examined immune responses to the 25-μg dose of the mRNA-1273 COVID-19 vaccine.

[1]Center for Infectious Disease and Vaccine Research, La Jolla Institute for Immunology (LJI), La Jolla, CA 92037, USA. [2]Department of Medicine, Division of Infectious Diseases and Global Public Health, University of California, San Diego (UCSD), La Jolla, CA 92037, USA.
*Corresponding author. Email: alex@lji.org (A.S.); shane@lji.org (S.C.); daniela@lji.org (D.W.)
†These authors contributed equally to this work.

Spike-specific antibody elicited by the 25-μg mRNA-1273 vaccine dose over time

An open-label, age de-escalation phase 1 trial used the mRNA-1273 vaccine with 25-μg immunizations on days 1 and 29 (*9*, *28*), with blood samples collected on study days 1, 15, 43, and 209. SARS-CoV-2 spike–binding antibodies, receptor binding domain (RBD)–binding antibodies, and SARS-CoV-2 pseudovirus (PSV) neutralization titers were determined (Fig. 1). Anti-spike and anti-RBD immunoglobulin G (IgG) were maintained at detectable levels for at least 7 months after the first vaccination in 100% (33/33) of subjects (Fig. 1, A and B). RBD IgG was induced by one immunization in 94% (33/35) of subjects. This response rate increased to 100% (33/33) after the second immunization and was maintained for at least 6 months after the second vaccination. SARS-CoV-2 PSV neutralization titers were detected in 29% (10/35) of subjects after one vaccination and 100% of subjects after two vaccinations (33/33), and 88% (29/33) of subjects maintained detectable neutralizing antibodies for at least 6 months after the second vaccination (Fig. 1C). All three antibody measurements demonstrated similar kinetics (Fig. 1, A to C) and were highly correlated (correlation coefficient, r = 0.89 to 0.90, fig. S1). Peak spike IgG, RBD IgG, and PSV titers were 6.8-, 9.5-, and 9.5-fold higher, respectively, than at 7 months (study day 209; 181 days after the second immunization). Similar fold changes were reported for 100-μg mRNA-1273 vaccination, indicating similar memory quality and durability (*40*). The 25-μg mRNA-1273 vaccine–generated antibodies were comparable to antibodies from SARS-CoV-2–infected subjects collected at a similar time after exposure [7 months post–symptom onset (PSO), 170 to 195 days] (Fig. 1D). Thus, increased anti-spike IgG, anti-RBD IgG, and PSV neutralizing antibodies were induced in response to two 25-μg mRNA-1273 vaccinations. These levels were maintained in 88 to 100% of vaccinees for at least 6 months after the second immunization and were comparable to those observed after infection with SARS-CoV-2.

Spike-specific CD4$^+$ T cells elicited by the 25-μg mRNA-1273 vaccine dose over time

SARS-CoV-2 spike–specific CD4$^+$ T cell responses were first measured using a flow cytometry activation-induced marker (AIM) assay (Fig. 2A and fig. S2). On day 1, before vaccination, spike-specific CD4$^+$ T cells with a predominantly memory phenotype were detected in 49% of clinical trial subjects (17/35), demonstrating the presence of preexisting SARS-CoV-2 spike–cross-reactive memory CD4$^+$ T cells, as discussed in the latter part of this Report. Spike-specific CD4$^+$ T cell responses were observed after the first vaccination in 97% of subjects (34/35) (Fig. 2A). Cytomegalovirus (CMV)–specific CD4$^+$ T cells were unchanged, as

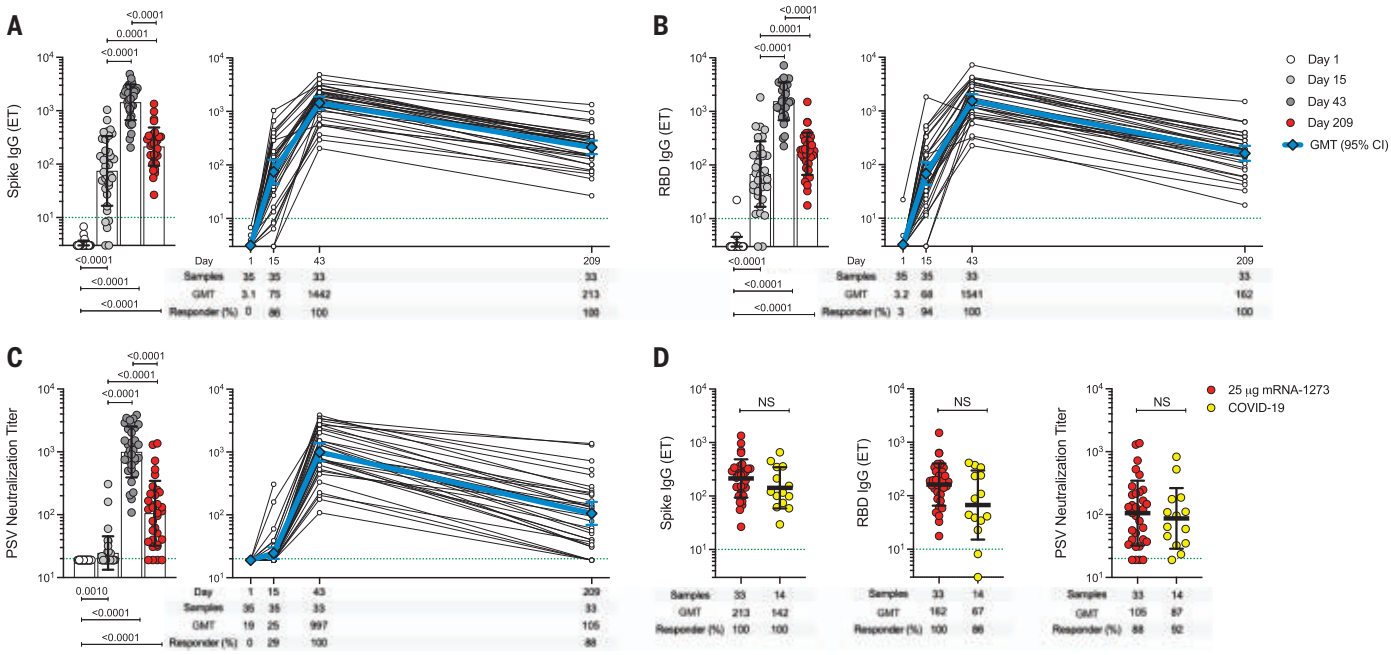


Fig. 1. Spike antibodies induced by 25-µg mRNA-1273 vaccination. Participants received two injections of the 25-µg mRNA-1273 vaccine, 28 days apart. PBMC samples were collected on day 1, day 15 ± 2 (2 weeks after first dose), day 43 ± 2 (2 weeks after second dose), and day 209 ± 7 days (6 months after second dose). (**A**) Longitudinal anti-SARS-CoV-2 spike IgG binding titers, (**B**) longitudinal anti-SARS-CoV-2 RBD IgG binding titers, and (**C**) longitudinal SARS-CoV-2 spike pseudovirus neutralizing titers (PSV). (**D**) Comparison of anti-spike IgG, anti-RBD IgG, and PSV neutralizing titers induced by two doses of 25-µg mRNA-1273 vaccine at day 209 ± 7 (n = 33) and COVID-19 convalescent donors at 170 to 195 days PSO (n = 14). Dotted green lines indicate the limit of quantification (LOQ). The bars in (A), (B), (C), and (D) indicate the geometric mean titers (GMTs) and geometric SD for anti-spike IgG (endpoint titer, ET), anti-RBD IgG (ET), and PSV neutralizing titers, respectively. Data were analyzed for statistical significance using Wilcoxon signed-rank test [(A), (B), and (C)] and Mann-Whitney U test (D). NS, nonsignificant. Background-subtracted and log data analyzed in all cases.

expected, indicating no bystander influence of the mRNA-1273 vaccination (fig. S3). The SARS-CoV-2 spike–specific $CD4^+$ T cell response rate increased to 100% (32/32) after the second vaccination and was maintained for at least 6 more months. Spike-specific memory $CD4^+$ T cell frequencies at 7 months were similar to those observed for COVID-19 cases (COVID-19 samples collected 170 to 195 days PSO) (Fig. 2B). Median mRNA-1273 vaccine–generated spike-specific $CD4^+$ T cell frequencies at all time points after vaccination also exceeded CMV-specific $CD4^+$ T cell frequencies (Fig. 2A and fig. S3A).

T follicular helper (T_{FH}) cell differentiation and cytokine production by vaccine-generated spike-specific $CD4^+$ T cells were then assessed (Fig. 2, C to F). T_{FH} cells are the specialized subset of $CD4^+$ T cells required for B cell help and are critical for the generation of neutralizing antibodies in most conditions (*41*). Spike-specific circulating T_{FH} (cT_{FH}) cells were detected in 71% (25/35) and 75% (24/32) of subjects after the first and second vaccination, respectively (Fig. 2C, right panel). Spike-specific cT_{FH} cells were detectable in 94% of subjects overall (33/35). Different response kinetics were observed at the level of individual subjects (Fig. 2C, right panel). Spike-specific memory cT_{FH} cells were still detected in 63% of vaccinees 6 months after the second vaccination (20/32) (Fig. 2C). Vaccine-specific $CD4^+$ T cell cytokine profiles were determined by intracellular cytokine staining (ICS) (Fig. 2, D to F, and fig. S4). Interferon-γ positive ($IFN\gamma^+$) spike-specific $CD4^+$ T cells were detected in 85% (28/33), tumor necrosis factor–α positive ($TNF\alpha^+$) in 97% (32/33), interleukin-2 positive ($IL\text{-}2^+$) in 100% (33/33), and granzyme B positive (GzB^+) in 76% (25/33) of subjects at day 43 (Fig. 2D). Little-to-no IL-4, IL-17A, or IL-10 was detected (fig. S5). Cytokine-producing spike-specific $CD4^+$ T cells ($CD40L^+$ cells producing IFNγ, TNFα, IL-2, and/or GzB) (Fig. 2E and fig. S6) were observed in 94% (33/35) and 100% (33/33) of subjects after the first and second vaccination, respectively, and were maintained for at least 6 months after the second vaccination [97% (32/33)] (Fig. 2E). Spike-specific $CD4^+$ T cells generated by the 25-µg mRNA-1273 vaccine exhibited multifunctionality comparable with that of CMV-specific cells (Fig. 2F, fig. S3C, and table S1). Thus, robust spike-specific $CD4^+$ T cells and T cell memory were generated by the low-dose mRNA-1723 vaccine, with strong T_{FH} and T helper 1 cell polarization advantageous for antiviral immunity.

Spike-specific $CD8^+$ T cells elicited by the 25-µg mRNA-1273 vaccine dose over time

SARS-CoV-2 spike–specific $CD8^+$ T cells were measured by AIM ($CD69^+$ and $CD137^+$) (fig. S2) and were observed in 34% (12/35) and 53% (17/32) of subjects after the first and second 25-µg mRNA-1273 vaccination, respectively (Fig. 3A). Spike-specific $CD8^+$ T cells were detectable for >6 months after the second vaccination, with a response rate comparable to that observed for COVID-19 cases (COVID-19 samples collected 170 to 195 days PSO) (Fig. 3B). Next, SARS-CoV-2 spike–specific $CD8^+$ T cells were measured by ICS (IFNγ, TNFα, IL-2, or GzB) (Fig. 3C and fig. S4). The first immunization induced significant spike-specific $CD8^+$ T cell cytokine responses in 51% (18/35) of subjects (Fig. 3D and fig. S7), increasing to 70% (23/33) of subjects after the second vaccination (Fig. 3D). $IFN\gamma^+$ spike-specific $CD8^+$ T cells were detected in 70% (23/33), $TNF\alpha$ in 39% (13/33), and $IL\text{-}2^+$ in 12% (4/33) of vaccinees at day 43 (Fig. 3C). Multiple positive- and negative-control samples and experimental conditions were used to demonstrate the specificity of the spike $CD8^+$ T cells (Fig. 3 and figs. S2 and S4, F and G). Correlation between AIM and ICS methods was highly significant for both spike-specific $CD4^+$ and $CD8^+$ T cells ($P < 0.0001$) (fig. S8). The fraction of multifunctional spike-specific $CD8^+$ T cells increased between first and second vaccination (three or more effector molecules expressed) (Fig. 3E and table S2). The most prevalent profile of $CD8^+$ T cells with

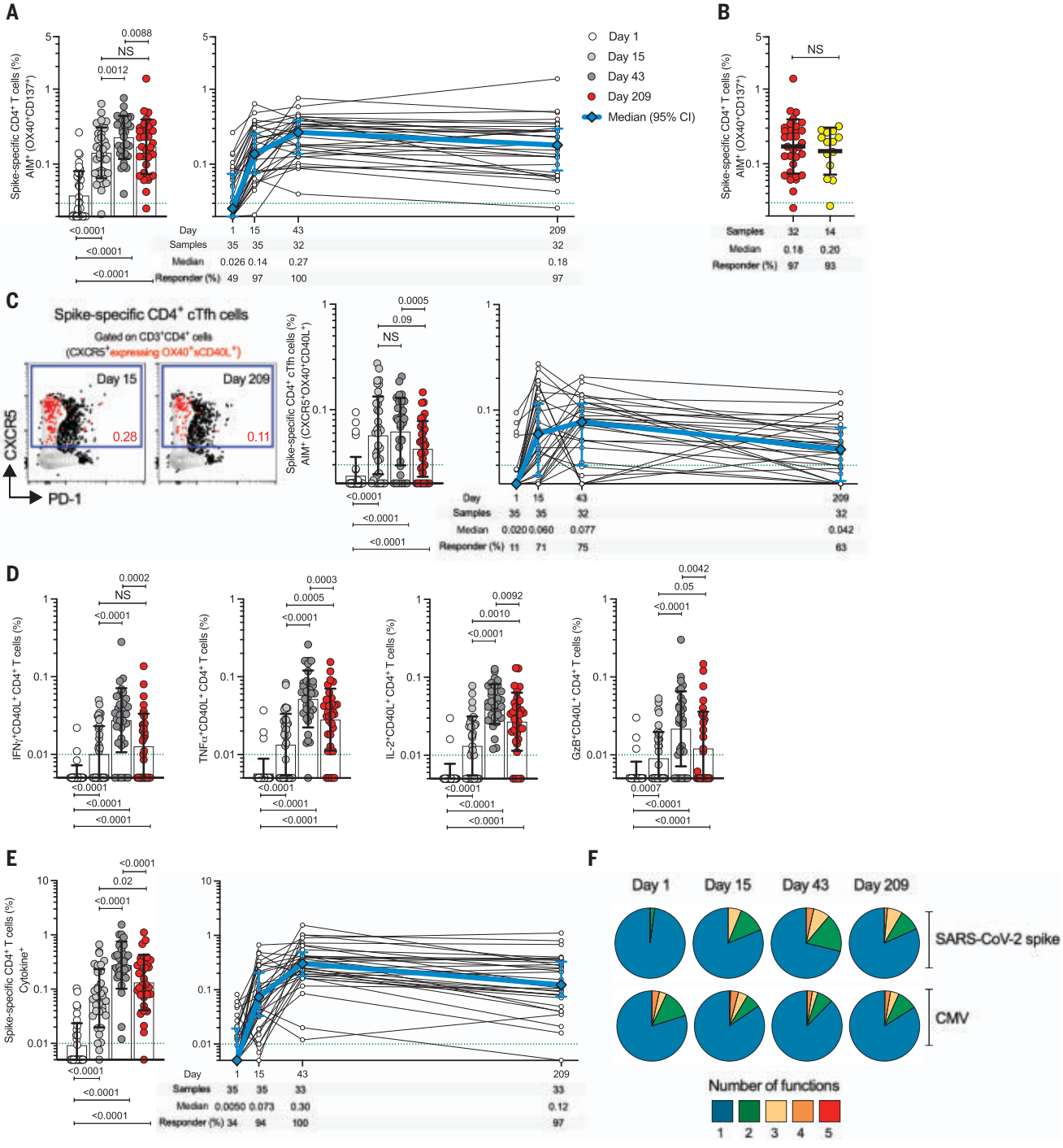


Fig. 2. mRNA-1273 vaccination induces durable and multifunctional spike-specific CD4$^+$ T cell responses. (**A**) Longitudinal spike-specific CD4$^+$ T cells in mRNA-1273 vaccinees measured by AIM. Spike-specific CD4$^+$ T cells quantified by AIM (surface OX40$^+$CD137$^+$) after stimulation with spike megapool (MP) in mRNA-1273 vaccinees (see fig. S2 for gating strategy). (**B**) Comparison of spike-specific AIM$^+$ CD4$^+$ T cell frequencies between 25-μg mRNA-1273 vaccine at day 209 ± 7 (red circles, n = 32) and COVID-19 convalescent donors at 170 to 195 days PSO (yellow circles, n = 14). (**C**) Quantitation of spike-specific circulating T follicular helper (cT_{FH}) cells (CXCR5$^+$OX40$^+$surface CD40L$^+$, as percentage of CD4$^+$ T cells) after stimulation with spike MP. Representative examples of spike-specific cT_{FH} cells (red), overlaid on total CD4$^+$ T cells, at days 15 ± 2 and 209 ± 7. (**D**) Spike-specific CD4$^+$ T cells expressing intracellular CD40L (iCD40L) and producing IFNγ, TNFα, IL-2, or granzyme B (GzB) in mRNA-1273 vaccinees. (**E**) Longitudinal spike-specific CD4$^+$ cytokine$^+$ T cells expressing iCD40L or producing IFNγ, TNFα, IL-2, or GzB in 25-μg mRNA-1273 vaccinees (see fig. S4 for gating strategy). Dotted green lines indicate the limit of quantification (LOQ). White, day 1; light gray, day 15 ± 2; dark gray, day 43 ± 2; red, day 209 ± 7. The bars in (A) to (E) indicate the geometric mean and geometric SD in the analysis of the spike-specific CD4$^+$ T cell frequencies. (**F**) Longitudinal multifunctional spike-specific CD4$^+$ T cells in mRNA-1273 vaccinees. Proportions of multifunctional activity profiles of the spike-specific CD4$^+$ T cells from mRNA-1273 vaccinees evaluated on days 1, 15 ± 2, 43 ± 2, and 209 ± 7. The blue, green, yellow, orange, and red colors in the pie charts depict the production of one, two, three, four, or five functions, respectively (see figs. S4 and S6 for details). Data were analyzed for statistical significance using Wilcoxon signed-rank test [(A), (C), (D), and (E)] and Mann-Whitney U test (B). Background-subtracted and log data analyzed in all cases.

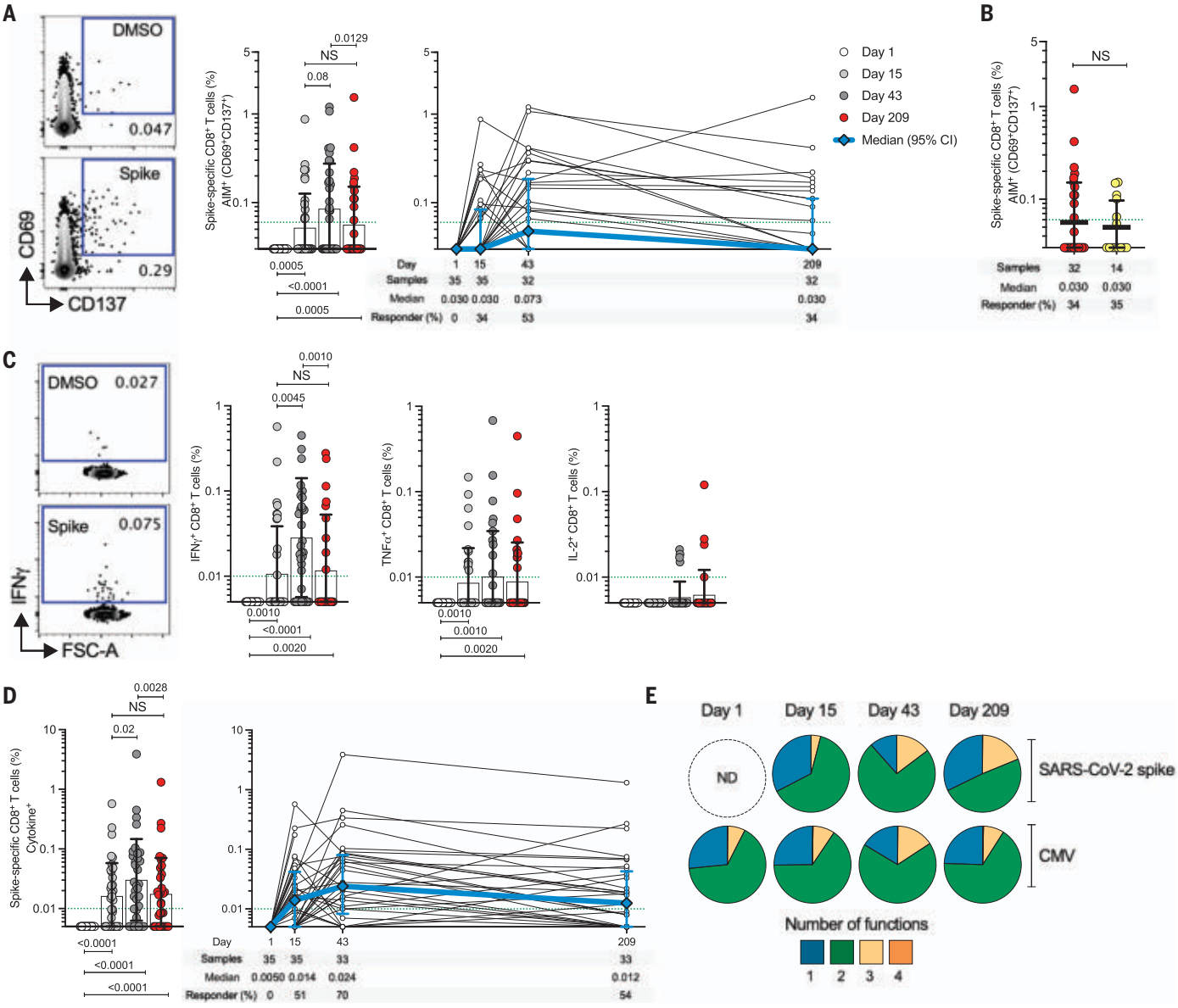


Fig. 3. mRNA-1273 vaccination induces multifunctional spike-specific CD8⁺ T cells. (**A**) Longitudinal spike-specific CD8⁺ T cells in mRNA-1273 vaccinees measured by AIM (surface CD69⁺CD137⁺). (Left panel) Representative examples of flow cytometry plots of spike-specific CD8⁺ T cells compared with DMSO control (see fig. S2 for gating strategy). (Right panel) Spike-specific CD8⁺ T cells quantified. (**B**) Comparison of spike-specific AIM⁺ CD8⁺ T cell frequencies between 25-μg mRNA-1273 vaccinees at day 209 ± 7 (n = 32) and COVID-19 convalescent donors at 170 to 195 days PSO (n = 14). (**C**) Spike-specific CD8⁺ T cells producing IFNγ, TNFα, or IL-2 by intracellular cytokine staining (ICS) in 25-μg mRNA-1273 vaccinees. (**D**) Longitudinal spike-specific CD8⁺ cytokine⁺ T cells producing IFNγ, TNFα, IL-2, or GzB in 25-μg mRNA-1273 vaccinees (see fig. S4 for gating strategy). Dotted green lines indicate the LOQ. The bars in (A) to (D) indicate the geometric mean and geometric SD. White, day 1; light gray, day 15 ± 2; dark gray, day 43 ± 2; red, day 209 ± 7. (**E**) Multifunctional activity profiles of spike-specific CD8⁺ T cells from 25-μg mRNA-1273 vaccinees, evaluated for IFNγ, TNFα, IL-2, or GzB (see figs. S4 and S8 for details). The blue, green, yellow, and orange colors in the pie charts depict the production of one, two, three, or four functions, respectively. ND, nondetectable. Data were analyzed for statistical significance using Wilcoxon signed-rank test [(A), (C), and (D)] and Mann-Whitney U test (B). Background-subtracted and log data analyzed in all cases.

three functions was GzB⁺IFNγ⁺TNFα⁺ (fig. S7), similar to the profile seen in CMV-specific CD8⁺ T cells (Fig. 3D and fig. S9). Thus, 25-μg mRNA-1273 vaccination induces multifunctional spike-specific memory CD8⁺ T cells.

Anti-spike antibody and CD4⁺ and CD8⁺ T cell responses generated by 25-μg mRNA-1273 vaccination were multifunctional, durable, and comparable in magnitude to those induced by natural infection (Table 1) (*22*). A concern has been raised that vaccination may not induce adequate immune memory in the elderly (*42*). Our vaccination cohort consisted of volunteers from three different age groups (*9*, *28*). Spike IgG and RBD IgG in the older groups (56–70 and >70 years) were reduced about twofold on day 209 (Fig. 4, A and B), similar to what was reported for 100-μg mRNA-1273 vaccination (*40*). Spike-specific CD4⁺ or CD8⁺ T cells were not reduced in the older vaccinee groups compared with the 18- to 55-year-old age group. Memory CD4⁺ and CD8⁺ T cell frequencies at day 209 were at least as strong in the older age groups as in younger adults (Fig. 4, D and H). Thus, although the study size is underpowered for in-depth examination of the three age groups, a small reduction in antibody but not T cell

Table 1. Spike-specific immune responses detected in 25-μg mRNA-1273 vaccinees. ELISA, enzyme-linked immunosorbent assay; AIM, activation-induced markers; ICS, intracellular cytokine staining.

Component	Assay	Days after vaccination			
		1	15 ± 2	43 ± 2	209 ± 7
Antibodies					
Anti-spike IgG	ELISA	0	86%	100%	100%
Anti-RBD IgG	ELISA	3%	94%	100%	100%
Neutralizing	Neutralization	0	29%	100%	88%
T cells					
Spike-specific $CD4^+$ T cells	AIM*	49%	97%	100%	97%
	ICS†	34%	94%	100%	97%
	Total‡	49%	97%	100%	97%
Spike-specific $CD8^+$ T cells	AIM*	0	34%	53%	34%
	ICS†	0	51%	70%	54%
	Total‡	0	69%	88%	67%

*Antigen-specific T cells using AIM were measured as a percentage $CD4^+$ T expressing $OX40^+CD137^+$ and $CD8^+$ T cells expressing $CD69^+CD137^+$ after stimulation of PBMC with spike overlapping peptides spanning the entire protein. †Antigen-specific T cells using ICS were measured as a percentage $CD4^+$ T expressing CD40L or producing IFNγ, TNFα, IL-2, or GzB; and $CD8^+$ T cells producing IFNγ, TNFα, IL-2, or GzB after stimulation of PBMC with spike overlapping peptides spanning the entire protein. ‡The overall spike-specific T cell response was calculated on the basis of the AIM and ICS results per donor and time point.

memory was observed in older adults compared with younger adults.

Spike-specific immune responses elicited by the 100-μg mRNA-1273 vaccine dose

More than 100 million doses of the 100-μg mRNA-1273 vaccine have been administered in the United States to date. We compared immune responses to the 25-μg and 100-μg doses of mRNA-1273 (Fig. 4, I to K). Anti-spike IgG, anti-RBD IgG, and PSV neutralizing titers were about twofold higher in 100-μg vaccinees than in those who received the 25-μg dose (Fig. 4I), which is consistent with earlier reports (*9*, *28*, *40*). Spike-specific $CD4^+$ T cells responses were ~1.4- to 2.0-fold higher in 100-μg vaccinees than in 25-μg vaccinees (Fig. 4J). Furthermore, the spike-specific $CD8^+$ T cell responses were comparable between the doses (Fig. 4K).

Preexisting cross-reactive memory

Preexisting cross-reactive memory T cells recognizing SARS-CoV-2 in vitro are found in many individuals (*30*–*37*). It was hypothesized that the existence of preexisting spike-recognizing immune memory may modulate immune responses to infection or vaccination (*43*). To address this question, we separated our cohort as a function of whether each subject had preexisting cross-reactive memory $CD4^+$ T cells reactive against SARS-CoV-2 spike (Fig. 5, A and B). As noted above (Fig. 2A), preexisting SARS-CoV-2 spike–specific $CD4^+$ T cells were present in 49% (17/35) of the 25-μg mRNA-1273 vaccinees. After vaccination, spike-specific $CD4^+$ T cells were significantly higher at day 15 in subjects with cross-reactive memory than in subjects with no cross-reactive memory (2.3-fold, $P < 0.0001$) (Fig. 5C). Spike-specific memory $CD4^+$ T cell frequencies were also higher after the second vaccination in subjects with cross-reactive memory than in those without ($P = 0.02$) (Fig. 5) and remained higher for ≥6 months ($P = 0.01$) (Fig. 5C). The impact of preexisting cross-reactive spike-specific $CD4^+$ T cell memory was more than additive (day 15, $P = 0.018$) (fig. S13), demonstrating that the spike-specific $CD4^+$ T cell response to the vaccine was enhanced by cross-reactive memory. Higher frequencies of cytokine-positive spike-specific $CD4^+$ T cells ($P = 0.0051$) (Fig. 5D) and multifunctional cells ($P = 0.02$) (Fig. 5E and fig. S10) were also observed after the first vaccination in individuals with cross-reactive memory. Preexisting cross-reactive $CD4^+$ T cells were observed in all three age groups (Fig. 4D). We did not detect preexisting cross-reactive spike-specific $CD8^+$ T cell memory (Figs. 3, A and D, and 4, G and H), and we observed no modulation of vaccine $CD8^+$ T cell responses by preexisting $CD4^+$ T cell memory (fig. S11). Cross-reactive memory $CD4^+$ T cells recognizing SARS-CoV-2 nonspike epitopes were present (fig. S12), as expected (*31*, *34*). The nonspike cross-reactive memory $CD4^+$ T cell frequencies remained unchanged over 7 months and were not modulated by mRNA-1273 vaccination (fig. S12), consistent with the vaccine containing only spike antigen and not causing bystander activation. Thus, preexisting cross-reactive $CD4^+$ T cell memory can influence mRNA-1723 vaccine–generated $CD4^+$ T cell responses.

T_{FH} cells in subjects with and without cross-reactive memory were of particular interest, because of their relevance in antibody responses. Frequencies of spike-specific cT_{FH} cells 2.6-fold higher were observed on day 15 in vaccinees with preexisting cross-reactive memory ($P = 0.0023$) (Fig. 5F). Likewise, significantly higher levels of anti-spike IgG ($P = 0.02$) (Fig. 5G) and anti-RBD IgG ($P = 0.047$) (Fig. 5H) were detected on day 15 in vaccinees with preexisting cross-reactive memory. The group with preexisting cross-reactive $CD4^+$ T cell memory demonstrated higher SARS-CoV-2 neutralizing titers 7 months after vaccination than did the group without preexisting cross-reactive $CD4^+$ T cell memory ($P = 0.04$) (Fig. 5I). Thus, a coordinated increase in spike-specific cT_{FH} responses, anti-spike IgG, and anti-RBD IgG are detected after a single vaccination in subjects with preexisting cross-reactive memory. Furthermore, 6 months after the second vaccination, higher levels of vaccine-specific memory $CD4^+$ T cells and higher titers of SARS-CoV-2 PSV neutralizing antibodies are present in individuals who have preexisting cross-reactive $CD4^+$ T cell memory.

Concluding remarks

The SARS-CoV-2 mRNA vaccines have been extraordinary successes. It is important to better understand the immunology of these vaccines in order to better appreciate (i) the mechanisms of protective immunity provided by the vaccines, (ii) the durability of immune memory generated by the vaccines (and thus infer trajectories of protective immunity), and (iii) the immunological features of these vaccines that may be relevant for vaccine design against other pathogens. Here, studying 35 vaccinated subjects 7 months out from the initial immunization, we found that two-dose 25-μg mRNA-1273 vaccination generated immune memory against spike comparable to that of SARS-CoV-2 infection for antibodies, $CD4^+$ T cells, and $CD8^+$ T cells. Furthermore, immune responses were considerably enhanced by the presence of preexisting cross-reactive $CD4^+$ T cell memory.

We consistently found spike-specific memory $CD4^+$ T cells in vaccinated subjects 6 months after second-dose 25-μg mRNA-1273 immunization. Less than a twofold difference in spike-specific $CD4^+$ T cell frequencies was observed between peak and 6 months after the second vaccination, indicative of durable vaccine T cell memory. Spike-specific memory $CD4^+$ T cell frequencies were also similar between low-dose mRNA-1273 vaccinated persons and COVID-19 cases. The vaccine-generated cells also exhibited an antiviral functional profile, including substantial T_{FH} cell counts and IFNγ expression, and the presence of multi-cytokine-expressing cells in proportions similar to CMV-specific memory $CD4^+$ T cells.

Uncertainty has surrounded whether the mRNA-1273 vaccine elicits effector and memory $CD8^+$ T cells in humans (*9*, *44*, *45*). Here

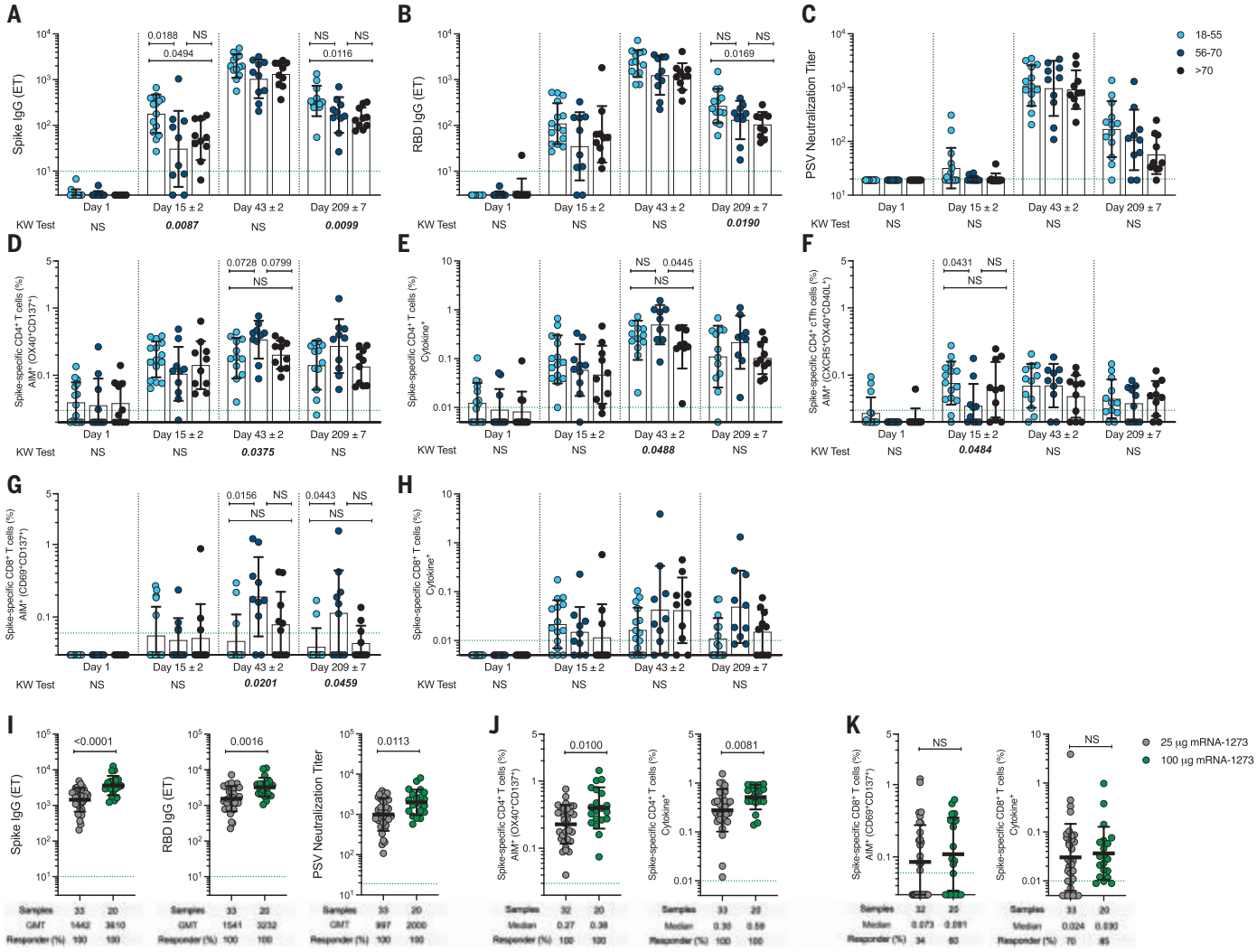


Fig. 4. Spike-specific antibody and T cell responses induced by mRNA-1273 vaccination. (**A** to **H**) Immune responses to 25-µg mRNA-1273 vaccination in three adult age groups: 18–55 (light blue symbols), 56–70 (dark blue), and over 70 years of age (black). [(A) to (C)] Anti-spike IgG, anti-RBD IgG, and PSV neutralizing titers. [(D) and (E)] Spike-specific $CD4^+$ T cells by AIM (D) or ICS (E). (F) Spike-specific cT_{FH} cells. [(G) and (H)] Spike-specific $CD8^+$ T cells by AIM (G) or ICS (H). Data were analyzed for statistical significance using nonparametric analysis of variance Kruskal-Wallis (KW) test and Dunn's post-test for multiple comparisons. The *P* values plotted on the bottom show the KW test results, and the *P* values plotted on the top show the post-test analysis comparing age groups. (**I**) Anti-SARS-CoV-2 spike IgG, RBD IgG, and PSV neutralizing titers in 25-µg and 100-µg mRNA-1273 vaccinees at day 43 (2 weeks after second dose). (**J**) Spike-specific $CD4^+$ AIM^+ (left) and $cytokine^+$ (right) T cells in mRNA-1273 vaccinees. (**K**) Spike-specific $CD8^+$ AIM^+ (left) and $cytokine^+$ (right) T cells in mRNA-1273 vaccinees. Dotted green lines indicate the limit of quantification (LOQ). The bars indicate geometric mean and geomean SD. Data in (I) to (K) were analyzed for statistical significance using Mann-Whitney *U* test. Background-subtracted and log data analyzed in all cases.

we report, at the peak of the immune response, spike-specific $CD8^+$ T cells by AIM or ICS assays were detected in 88% of subjects receiving low-dose mRNA-1273, which is a $CD8^+$ T cell response rate equivalent to that of COVID-19 cases (*21*, *22*, *26*, *31*). We speculate that absence of detection of spike-specific $CD8^+$ T cells in some studies reflects the stringency of the experimental conditions used. Here, allowance for 24 hours of antigen stimulation revealed vaccine-generated $CD8^+$ T cells in most individuals. $CD8^+$ T cells have been observed in response to the Pfizer-BioNTech BNT162b2 vaccine using peptide–major histocompatibility complex multimers (*44*). Moreover, $CD8^+$ T cells have been found 2 months after second immunization with BNT162b2 (*44*). In this study, spike-specific memory $CD8^+$ T cells were detected 6 months after the second immunization with 25-µg mRNA-1273. These mRNA-1273 vaccine–generated memory $CD8^+$ T cells were detected in 67% of subjects and were not dissimilar in magnitude to spike-specific memory $CD8^+$ T cells in COVID-19 cases. Limitations of this study include the relatively small sample size and limited cell availability. Overall, the data show that $CD4^+$ and $CD8^+$ T cell memory are generated by both low-dose and 100-µg dose COVID-19 mRNA-1273 vaccine.

Low-dose RNA vaccines have potential advantages for future needs and applications such as dose sparing. Low-dose immunization is also less reactogenic (*9*, *44*), which may also be appealing in contexts of multidose regimens. It is of interest to consider different vaccine doses across age groups, or high- versus low-risk groups, but a better understanding of immune memory to different doses is key for such considerations. Data reported here are encouraging demonstrations of the potential of RNA vaccines to generate durable T cell and antibody immune memory, including at lower vaccine doses.

Preexisting immunity in the form of cross-reactive memory $CD4^+$ T cells affected immune

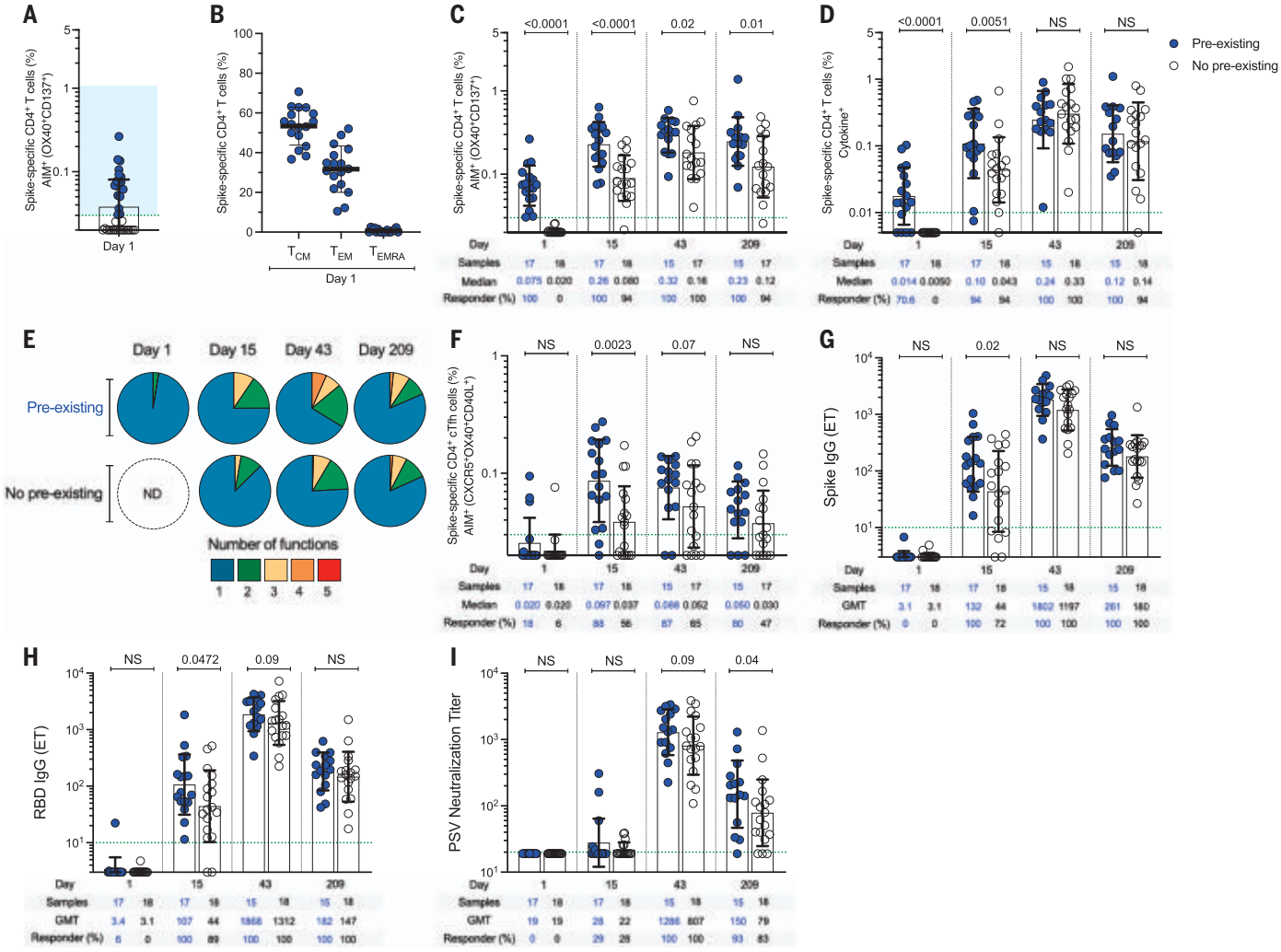


Fig. 5. Preexisting anti-spike immunity modulates T cell and antibody responses. (**A**) Preexisting spike-specific CD4$^+$ AIM$^+$ T cells at day 1 (see Fig. 2 for details). (**B**) Memory phenotype of preexisting spike-specific CD4$^+$ AIM$^+$ T cells from (A). T_{CM}, central memory cells; T_{EM}, effector memory cells; T_{EMRA}, terminally differentiated effector memory cells. (**C**) Spike-specific CD4$^+$ AIM$^+$ T cells in mRNA-1273 vaccinees with ("preexisting," blue) and without ("no preexisting," white) preexisting cross-reactive spike-reactive memory CD4$^+$ T cells evaluated on days 1, 15 ± 2, 43 ± 2, and 209 ± 7 after immunization (see fig. S13 for details). (**D**) Spike-specific CD4$^+$ cytokine$^+$ T cells in mRNA-1273 vaccinees with and without preexisting cross-reactive memory CD4$^+$ T cells. (**E**) Proportions of multifunctional spike-specific CD4$^+$ T cells in mRNA-1273 vaccinees with ("pre-existing") and without ("no preexisting") preexisting cross-reactive spike-reactive memory CD4$^+$ T cells evaluated on days 1, 15 ± 2, 43 ± 2, and 209 ± 7 after immunization (see fig. S10 for details). (**F**) Spike-specific cT$_{FH}$ cells (as percentage of CD4$^+$ T cells), (**G**) anti-spike IgG, (**H**) anti-RBD IgG, and (**I**) SARS-CoV-2 PSV neutralizing titers in mRNA-1273 vaccinees without and with preexisting cross-reactive spike-reactive memory CD4$^+$ T cells. Dotted green lines indicate the LOQ. The bars in (A) and (C) to (I) indicate the geometric mean and geomean SD in the analysis of the antibody levels or spike-specific CD4$^+$ and CD8$^+$ T cells in mRNA-1273 vaccinees with ("preexisting") and without ("no preexisting") pre-existing cross-reactive spike-reactive memory CD4$^+$ T cells evaluated on days 1, 15 ± 2, 43 ± 2, and 209 ± 7 after immunization. The bars in (B) indicate the mean and SD in the analysis of the memory phenotype of spike-specific CD4$^+$ T cells. Data were analyzed for statistical significance using Mann-Whitney *U* test. Background-subtracted and log data analyzed in all cases.

responses to the mRNA COVID-19 vaccine in this cohort. This indicates that cross-reactive memory T$_{FH}$ cells may both accelerate B cell priming and antibody responses to a new antigen and increase robustness of long-term humoral immunity, as evidenced by the higher neutralizing antibody titers. Both total spike-specific CD4$^+$ T cells and spike-specific T$_{FH}$ cells were enhanced after one immunization in persons with cross-reactive memory, suggesting that the spike cross-reactive memory CD4$^+$ T cells are recalled upon vaccination and affect the CD4$^+$ T cell repertoire. By contrast, CD8$^+$ T cell responses were unchanged. Our findings of preexisting immunity enhancing spike-specific CD4$^+$ T cell, T$_{FH}$, and antibody responses after immunization with an RNA vaccine do not represent the exact scenario of SARS-CoV-2 infection. It therefore remains unresolved whether preexisting T cells have a biological function during human SARS-CoV-2 infection (*33*). Nevertheless, these data provide evidence that the cross-reactive CD4$^+$ T cells are biologically relevant in the context of vaccination. Thus, it is plausible that the presence and magnitude of cross-reactive memory T cells could accelerate the speed and magnitude of CD4$^+$ T cell and antibody responses to SARS-CoV-2 infection, as compared with persons who have undetectable levels of cross-reactive memory T cells. Moreover, early T cell responses have been linked to less-severe COVID-19 clinical outcomes (*46*, *47*). These findings show substantial immune responses and immune memory to a low-dose mRNA vaccine and indicate

biological relevance of cross-reactive memory T cells.

Materials and methods

Human subjects

Samples from the phase 1 25-μg mRNA-1273 SARS-CoV-2 trial

A total of 140 peripheral blood samples were obtained from 35 participants who received the 25-μg dose mRNA-1273 SARS-CoV-2 vaccine in a dose-escalation, open-label phase 1 trial (mRN-1273 ClinicalTrials.gov number NCT04283461) (*9*). Participants received two injections of the trial vaccine 28 days apart between March and April of 2020 at one of the two study sites, Kaiser Permanente Washington Health Research Institute in Seattle or at the Emory University School of Medicine in Atlanta. Peripheral blood mononuclear cell (PBMC) samples used in this study were collected at the start of the trial (day 1), 14 days after each vaccination (day 15 ± 2 and 43 ± 2) and 7 months after the first dose (day 209 ± 7).

Participants were divided into three age groups: 18 to 55 (n = 15), 56 to 70 (n = 10), and >70 (n = 10) years of age (table S3). Two participants from the 18–55 group received only one dose of the 25-μg mRNA-1273 vaccine. The day 40 and day 209 samples (both collected after the second dose) were excluded from analysis for these two vaccinees. Both sexes were represented [20:15, male:female (M:F)], and the average age of the participants was 58 years. Participants identified as white (n = 34), Hispanic (n = 1), or mixed race (Asian/Hispanic, n = 1.) An overview of samples analyzed in this study is provided in table S3.

The trial was reviewed and approved by the Advarra institutional review board, as previously published (*9*). All experiments performed at the La Jolla Institute (LJI) were approved by the institutional review board (IRB) of the La Jolla Institute (IRB#: VD-214). Collection and processing of vaccinee PBMC samples was performed at one of the two study location sites, as previously described (*9*).

Samples from convalescent COVID-19 donors

To compare levels of immune memory responses induced by 25-μg mRNA-1273 SARS-CoV-2 vaccination with immune memory responses induced by natural infection with SARS-CoV-2, we collected blood from individuals that experienced natural infection with SARS-CoV-2. We matched the 7-month (209-day) postvaccination samples with samples from convalescent donors collected on average 181 days (range: 170 to 195 days) post–symptom onset (PSO). Assuming an average incubation period of 12 days until symptom onset, the time point after exposure or vaccination is comparable between the cohorts (*48*). To further match ethnicities between the cohorts, we selected 13 samples from Caucasian donors and one sample from a Hispanic donor. Convalescent donors were California residents who were either referred to the study by a health care provider or self-referred. In the overall cohort, both sexes were represented (10:4, M:F), and the average age of the donors was 35 years (±16.53). Details of this COVID-19 convalescent cohort are listed in table S4. All of the convalescent donors experienced mild illness, defined as patients with a SARS-CoV-2 positive test who have never been hospitalized (*46*). Seropositivity against SARS-CoV-2 was confirmed by enzyme-linked immunosorbent assay (ELISA), as describe below. At the time of enrollment, all COVID-19 convalescent donors provided informed consent to participate in the present and future studies.

100-μg mRNA-1273 vaccinees

We also included a cohort of individuals vaccinated locally in San Diego, California, who received the emergency use authorization–approved 100-μg mRNA-1273 SARS-CoV-2 vaccine. Samples from 20 vaccinees collected 42 ± 6 days after first immunization (15 days after second immunization) were compared with samples from the 25-μg mRNA-1273 cohort on day 43 ± 2. To match ethnicities between the cohorts, we selected 12 samples from Caucasian donors, 4 samples from Asian donors, and 4 samples from Hispanic donors. In the overall cohort, both sexes were represented (6:14, M:F) and the average age of the donors was 48 years (±14.45). Details of this cohort are listed in table S5. At the time of enrollment, all 100-μg mRNA-1273 SARS-CoV-2 vaccinees provided informed consent to participate in the present and future studies.

Peripheral blood mononuclear cells (PBMCs) and plasma isolation

Whole blood samples from convalescent COVID-19 donors and 100-μg mRNA-1273 vaccinees were collected at La Jolla Institute in heparin-coated blood bags and centrifuged for 15 min at 803*g* to separate the cellular fraction and plasma. The plasma was then carefully removed from the cell pellet and stored at −20°C. PBMCs were isolated by density-gradient sedimentation using Ficoll-Paque (Lymphoprep, Nycomed Pharma, Oslo, Norway), as previously described (*22*, *31*). Isolated PBMCs were cryopreserved in cell recovery media containing 10% dimethyl sulfoxide (DMSO) (Gibco), supplemented with 10% heat-inactivated fetal bovine serum (FBS; Hyclone Laboratories, Logan UT), and stored in liquid nitrogen until used in the assays.

Antibody measurements

SARS-CoV-2 ELISAs

SARS-CoV-2 ELISA titers in vaccinated and convalescent samples were determined as previously described (*22*, *31*, *46*, *49*). Endpoint titers (ETs) were plotted for each sample, using background subtracted data, with ETs calculated at the dilution giving a reading above the optical density cutoff of 0.1. Negative and positive controls were used to standardize each assay and normalize across experiments. A positive control standard was created by pooling plasma from six convalescent COVID-19 donors to normalize between experiments. The limit of detection was defined as 1:3 for IgG. The limit of quantification (LOQ) for vaccinated individuals was established on the basis of uninfected subjects, using plasma from healthy donors never exposed to or vaccinated against SARS-CoV-2.

Pseudovirus (PSV) neutralization assay

The PSV neutralization assays of vaccinated and convalescent samples were performed as previously described (*22*). Neutralization titers or inhibition dose 50 (ID_{50}) were calculated using the One-Site Fit Log IC50 model in Prism 9.2 (GraphPad). As internal quality control to define the interassay variation, three samples were included across the PSV neutralization assays. Samples that did not reach 50% inhibition at the lowest serum dilution of 1:20 were considered to be non-neutralizing, and the values were set to 19. PSV neutralization titers were performed as two independent experiments on different days with two replicates per experiment. Results were comparable between experiments and results from the first experiment are graphed. We included the World Health Organization (WHO) International Reference Panel for anti-SARS-Cov2 immunoglobulin (20/268) to calibrate our PSV neutralization titers. The ID_{50} of WHO-High and WHO-Mid were measured by four independent experiments with two replicates per experiment. The geometric means (GMTs) of ID_{50} of WHO-High and WHO-Mid were 2658 and 364, respectively, by our PSV neutralization assay. The WHO assigned neutralization activity unitage of 1473 and 210 IU/ml for the WHO-High and WHO-Mid standards. The calibration factor was thus calculated as ((2658/1473) + (364/ 210))/ 2 = 1.766. The GMTs of PSV neutralization ID_{50} in 25-μg and 100-μg mRNA-1273 vaccinees at day 43 were 997 and 2000, respectively, in our figures. The WHO IU calibrated neutralization ID_{50} (cID_{50}) GMTs in 25-μg mRNA-1273 vaccinees at day 43 would be 565 IU (997/1.766), and for 100-μg mRNA-1273 vaccinees at day 43 would be 1133 IU (2000/1.766). The limit of detection was calculated as 10.7 IU (19/1.766).

Peptide megapools (MPs)

We have previously developed the MP approach to allow simultaneous testing of a large number of epitopes (*22*, *31*, *46*, *50*). Here, three MPs to evaluate the antigen-specific T cell

response against SARS-CoV-2 were used, as described below. A cytomegalovirus (CMV) MP was used as a control against a ubiquitous pathogen in the experiments.

SARS-CoV-2 MPs

To characterize SARS-CoV-2–specific T cell response, we used three MPs previously described (*31*, *43*). First, we used a spike MP of 253 overlapping peptides spanning the entire sequence of the spike protein. As this peptide pool consists of peptides with a length of 15 amino acids, both $CD4^+$ and $CD8^+$ T cells have the capacity to recognize this MP (*51*). In addition, to confirm that the spike-specific $CD8^+$ T cell response observed in the mRNA-1273 vaccinees is also induced in the absence of the spike-specific $CD4^+$ T cell response, we performed some experiments with an optimal MP of HLA class I epitopes (CD8-S MP). This CD8-S MP consists of 197 9- and 10-mers derived spike peptides that have previously been described to be recognized by $CD8^+$ T cells in SARS-CoV-2 exposed donors (*22*, *31*, *46*, *52*). Lastly, we used a predicted SARS-CoV-2–specific MP (CD4-R) to evaluate the nonspike response or the remainder of the SARS-CoV-2 genome of 246 HLA class II $CD4^+$ epitopes previously described (*31*). We have previously shown that these MPs are suitable for stimulating T cell responses from either COVID-19–exposed or SARS-CoV-2–unexposed individuals (*31*, *34*).

Cytomegalovirus (CMV) MP

As a control, we used a MP of 313 experimentally defined epitopes. This CMV MP consists of HLA class I and class II epitopes and $CD4^+$ and $CD8^+$ T cells have the capacity to recognize this MP, as has been previously published (*50*).

Flow cytometry assays

Activation-induced markers (AIM) assay

Antigen-specific $CD4^+$ T cells were measured as a percentage of AIM^+ ($OX40^+CD137^+$) $CD4^+$ and ($CD69^+CD137^+$) $CD8^+$ T cells after stimulation of PBMCs from mRNA-1273 vaccinees and COVID-19 convalescent donors with peptide MPs. Antigen-specific circulating T follicular helper (cT_{FH}) cells ($CXCR5^+OX40^+CD40L^+$, as percentage of $CD4^+$ T cells) were defined by the AIM assay.

Before addition of MPs, cells were blocked at 37°C for 15 min with 0.5 μg/ml of anti-CD40 monoclonal antibody (mAb) (Miltenyi Biotec). Then, cells were incubated at 37°C for 24 hours in the presence of fluorescently labeled anti-chemokine receptor antibodies (anti-CCR4, -CCR6, -CCR7, -CXCR3, and -CXCR5) and SARS-CoV-2 MPs (1 μg/ml) or CMV MP (1 μg/ml) in 96-well U-bottom plates, as previously described (*31*, *46*). In addition, PBMCs were incubated with an equimolar amount of DMSO as negative control and with phytohemagglutinin (5 μg/ml) (PHA, Roche) as a positive control.

For the surface stain, 1×10^6 PBMCs were resuspended in phosphate-buffered saline (PBS), incubated with BD human FC block (BD Biosciences, San Diego, CA) and the LIVE/DEAD marker in the dark for 15 min and washed with PBS. Then, an antibody mix containing the rest of the surface antibodies was added directly to cells and incubated for 60 min at 4°C in the dark. After surface staining, cells were washed twice with PBS containing 3% FBS (FACS buffer). All samples were acquired on a Cytek Aurora (Cytek Biosciences, Fremont, CA). A list of antibodies used in this panel can be found in table S6, and a representative gating strategy of spike-specific $CD4^+$ and $CD8^+$ T cells using the AIM assay is shown in fig. S2.

Antigen-specific $CD4^+$ and $CD8^+$ T cells were measured as background (DMSO)–subtracted data, with a minimal DMSO level set to 0.005%. A response of >0.02% and a stimulation index (SI) of >2 for $CD4^+$, and a response of >0.03% and SI of >3 for $CD8^+$ T cells, were considered positive. The LOQ for antigen-specific $CD4^+$ T cell responses (0.03%) and antigen-specific $CD8^+$ T cell responses (0.06%) was calculated using the median twofold standard deviation of all negative controls. As an internal quality control to define interassay variation, the CMV-specific $CD4^+$ and $CD8^+$ T cell responses were evaluated for a SARS-CoV-2–unexposed donor included in each independent experiment. The antigen-specific response against CMV and the response to the positive control was compared across experiments and revealed a coefficient of variation (CV) between 10 and 13% for the antigen-specific stimulation with the CMV-MP and a CV between 6 and 12% for mitogenic stimulation with PHA (fig. S2).

Intracellular cytokine staining (ICS) assay

To optimize spike-specific detection of cytokine-producing $CD4^+$ and $CD8^+$ T cells, we experimented with different incubation times (5, 6, and 24 hours) in the presence of GolgiPlug containing brefeldin A and GolgiStop containing monensin (BD Biosciences, San Diego, CA) for an additional 1, 3, or 4 hours, respectively. To establish optimal conditions for the ICS assay, we evaluated the IFNγ-producing $CD4^+$ and $CD8^+$ T cells in 100-μg mRNA-1273 vaccinees (n = 4) and COVID-19 convalescent donors (n = 4) (table S7). The highest signal of IFNγ-producing $CD4^+$ and $CD8^+$ T cells was detected after 24+4 hours incubation in both vaccinees and convalescent donors. Thus, we chose 24 hours as the best condition to identify spike-specific $CD4^+$ and $CD8^+$ T cells producing intracellular cytokines.

Before the addition of MPs, cells were blocked at 37°C for 15 min with 0.5 μg/ml of anti-CD40 mAb, as previously described (*46*). PBMCs were cultured in the presence of SARS-CoV-2 MPs (1 μg/ml) for 24 hours at 37°C. In addition, PBMCs were incubated with an equimolar amount of DMSO as a negative control and also CMV MP (1 μg/ml) as a positive control. After 24 hours, Golgi-Plug and Golgi-Stop were added to the culture for 4 hours, as described above. Cells were then washed and surface-stained for 30 min at 4°C in the dark and fixed with 1% of paraformaldehyde (Sigma-Aldrich, St. Louis, MO). Antibodies used in the ICS assay are listed in table S8, and a representative gating strategy of spike-specific $CD4^+$ and $CD8^+$ T cells using the ICS assay is shown in fig. S4.

Antigen-specific $CD4^+$ and $CD8^+$ T cells were measured as background (DMSO) subtracted data, with a minimal DMSO level set to 0.001%. Responses of >0.005% and a SI of >2 for $CD4^+$ and $CD8^+$ T cells were considered positive. The limit of quantification for antigen-specific $CD4^+$ and $CD8^+$ T cell responses (0.01%) was calculated using the median twofold standard deviation of all negative controls. As internal quality control to define interassay variation, the CMV-specific $CD4^+$ and $CD8^+$ T cell responses were evaluated for a SARS-CoV-2–unexposed donor included in each independent experiment. The antigen-specific response of $CD4^+$ and $CD8^+$ T cells against CMV were compared across experiments and revealed a CV of 14% for the antigen-specific stimulation with the CMV MP (fig. S4E).

The gates applied for the identification of $CD4^+$ and $CD8^+$ T cells producing cytokines were defined according to the cells cultured with DMSO for each individual as is shown in fig. S4. A Boolean analysis was performed to define the multifunctional profiles on FlowJo 10.7.1. The analysis included CD40L, GzB, IFNγ, IL-2, and TNFα gated on $CD3^+CD4^+$ cells and GzB, IFNγ, IL-2, and TNFα gated on $CD3^+CD8^+$ cells. The overall response to spike was defined as the sum of the background subtracted responses to each combination of individual cytokines. To define the multifunctional profiles of antigen-specific T cells, all positive background-subtracted data (>0.005% and a SI > 2 for $CD4^+$ T cells and >0.002% and a SI > 2 for $CD8^+$ T cells) was aggregated into a combined sum of antigen-specific $CD4^+$ or $CD8^+$ T cells on the basis of the number of functions. Values higher than the LOQ (0.01%) were considered for the analysis of the multifunctional antigen-specific T cell responses. The average of the relative $CD4^+$ and $CD8^+$ T cell response was calculated per group to define the proportion of multifunctional antigen-specific T cell responses (figs. S6 and S7 and tables S1 and S2).

Statistical analysis

Data were analyzed using FlowJo 10.7.1. Statistical analyses were performed in GraphPad Prism 9.2, unless otherwise stated. The statistical details of the experiments are provided in the respective figure legends. Data plotted in linear scale were expressed as means ± standard deviations (SD). Data plotted in logarithmic scales were expressed as geometric means ± geometric standard

deviations (SD). Mann-Whitney *U* or Wilcoxon tests were applied for unpaired or paired comparisons, respectively. Differences among age groups were evaluated using Kruskal-Wallis and Dunn's post-test for multiple comparisons. Details pertaining to significance are also noted in the respective legends.

REFERENCES AND NOTES

1. F. P. Polack *et al*., Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N. Engl. J. Med.* **383**, 2603–2615 (2020). doi: 10.1056/NEJMoa2034577; pmid: 33301246
2. L. R. Baden *et al*., Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *N. Engl. J. Med.* **384**, 403–416 (2021). doi: 10.1056/NEJMoa2035389; pmid: 33378609
3. R. L. Soiza, C. Scicluna, E. C. Thomson, Efficacy and safety of COVID-19 vaccines in older people. *Age Ageing* **50**, 279–283 (2021). doi: 10.1093/ageing/afaa274; pmid: 33320183
4. C. Pawlowski *et al*., FDA-authorized mRNA COVID-19 vaccines are effective per real-world evidence synthesized across a multi-state health system. *Med* **2**, 979–992.e8 (2021). doi: 10.1016/j.medj.2021.06.007; pmid: 34223401
5. S. Amit, G. Regev-Yochay, A. Afek, Y. Kreiss, E. Leshem, Early rate reductions of SARS-CoV-2 infection and COVID-19 in BNT162b2 vaccine recipients. *Lancet* **397**, 875–877 (2021). doi: 10.1016/S0140-6736(21)00448-7; pmid: 33610193
6. R. W. Frenck Jr. *et al*., Safety, immunogenicity, and efficacy of the BNT162b2 Covid-19 vaccine in adolescents. *N. Engl. J. Med.* **385**, 239–250 (2021). doi: 10.1056/NEJMoa2107456; pmid: 34043894
7. V. J. Hall *et al*., COVID-19 vaccine coverage in health-care workers in England and effectiveness of BNT162b2 mRNA vaccine against infection (SIREN): A prospective, multicentre, cohort study. *Lancet* **397**, 1725–1735 (2021). doi: 10.1016/S0140-6736(21)00790-X; pmid: 33901423
8. D. Wrapp *et al*., Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020). doi: 10.1126/science.abb2507; pmid: 32075877
9. L. A. Jackson *et al*., An mRNA vaccine against SARS-CoV-2 — Preliminary report. *N. Engl. J. Med.* **383**, 1920–1931 (2020). doi: 10.1056/NEJMoa2022483; pmid: 32663912
10. S. J. Thomas *et al*., Six month safety and efficacy of the BNT162b2 mRNA COVID-19 vaccine. *medRxiv* 2021.07.28.21261159 [Preprint] (2021). https://doi.org/10.1101/2021.07.28.21261159.
11. Moderna, "Moderna Reports Second Quarter Fiscal Year 2021 Financial Results and Provides Business Updates," press release (5 August 2021); https://investors.modernatx.com/news-releases/news-release-details/moderna-reports-second-quarter-fiscal-year-2021-financial.
12. P. B. Gilbert *et al*., Immune correlates analysis of the mRNA-1273 COVID-19 vaccine efficacy trial. *medRxiv* 2021.08.09.21261290 [Preprint] (2021). https://doi.org/10.1101/2021.08.09.21261290.
13. D. S. Khoury *et al*., Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection. *Nat. Med.* **27**, 1205–1211 (2021). doi: 10.1038/s41591-021-01377-8; pmid: 34002089
14. A. Tauzin *et al*., A single dose of the SARS-CoV-2 vaccine BNT162b2 elicits Fc-mediated antibody effector functions and T cell responses. *Cell Host Microbe* **29**, 1137–1150.e6 (2021). doi: 10.1016/j.chom.2021.06.001; pmid: 34133950
15. A. Sette, S. Crotty, Adaptive immunity to SARS-CoV-2 and COVID-19. *Cell* **184**, 861–880 (2021). doi: 10.1016/j.cell.2021.01.007; pmid: 33497610
16. L. J. Abu-Raddad *et al*., SARS-CoV-2 antibody-positivity protects against reinfection for at least seven months with 95% efficacy. *EClinicalMedicine* **35**, 100861 (2021). doi: 10.1016/j.eclinm.2021.100861; pmid: 33937733
17. S. F. Lumley *et al*., Antibody status and incidence of SARS-CoV-2 infection in health care workers. *N. Engl. J. Med.* **384**, 533–540 (2021). doi: 10.1056/NEJMoa2034545; pmid: 33369366
18. 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
19. A. Leidi *et al*., Risk of reinfection after seroconversion to SARS-CoV-2: A population-based propensity-score matched cohort study. *Clin. Infect. Dis.* ciab495 (2021). doi: 10.1093/cid/ciab495; pmid: 34043763
20. N. R. Faria *et al*., Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* **372**, 815–821 (2021). doi: 10.1126/science.abh2644; pmid: 33853970
21. J. Zuo *et al*., Robust SARS-CoV-2-specific T cell immunity is maintained at 6 months following primary infection. *Nat. Immunol.* **22**, 620–626 (2021). doi: 10.1038/s41590-021-00902-8; pmid: 33674800
22. J. M. Dan *et al*., Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science* **371**, eabf4063 (2021). doi: 10.1126/science.abf4063; pmid: 33408181
23. R. J. Boyton, D. M. Altmann, Risk of SARS-CoV-2 reinfection after natural infection. *Lancet* **397**, 1161–1163 (2021). doi: 10.1016/S0140-6736(21)00662-0; pmid: 33743219
24. C. Li *et al*., Twelve-month specific IgG response to SARS-CoV-2 receptor-binding domain among COVID-19 convalescent plasma donors in Wuhan. *Nat. Commun.* **12**, 4144 (2021). doi: 10.1038/s41467-021-24230-5; pmid: 34230476
25. Z. Wang *et al*., Naturally enhanced neutralizing breadth against SARS-CoV-2 one year after infection. *Nature* **595**, 426–431 (2021). doi: 10.1038/s41586-021-03696-9; pmid: 34126625
26. K. W. Cohen *et al*., Longitudinal analysis shows durable and broad immune memory after SARS-CoV-2 infection with persisting antibody responses and memory B and T cells. *Cell Rep. Med.* **2**, 100354 (2021). doi: 10.1016/j.xcrm.2021.100354; pmid: 34250512
27. A. Pegu *et al*., Durability of mRNA-1273 vaccine-induced antibodies against SARS-CoV-2 variants. *Science* eabj4176 (2021). doi: 10.1126/science.abj4176; pmid: 34385356
28. E. J. Anderson *et al*., Safety and immunogenicity of SARS-CoV-2 mRNA-1273 vaccine in older adults. *N. Engl. J. Med.* **383**, 2427–2438 (2020). doi: 10.1056/NEJMoa2028436; pmid: 32991794
29. US Food and Drug Administration, "Vaccines and Related Biological Products Advisory Committee Meeting Presentation," 22 October 2020.
30. J. Braun *et al*., SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19. *Nature* **587**, 270–274 (2020). doi: 10.1038/s41586-020-2598-9; pmid: 32726801
31. A. Grifoni *et al*., Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. *Cell* **181**, 1489–1501.e15 (2020). doi: 10.1016/j.cell.2020.05.015; pmid: 32473127
32. N. Le Bert *et al*., SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. *Nature* **584**, 457–462 (2020). doi: 10.1038/s41586-020-2550-z; pmid: 32668444
33. M. Lipsitch, Y. H. Grad, A. Sette, S. Crotty, Cross-reactive memory T cells and herd immunity to SARS-CoV-2. *Nat. Rev. Immunol.* **20**, 709–713 (2020). doi: 10.1038/s41577-020-00460-4; pmid: 33024281
34. J. Mateus *et al*., Selective and cross-reactive SARS-CoV-2 T cell epitopes in unexposed humans. *Science* **370**, 89–94 (2020). doi: 10.1126/science.abd3871; pmid: 32753554
35. B. J. Meckiff *et al*., Imbalance of regulatory and cytotoxic SARS-CoV-2-reactive $CD4^+$ T cells in COVID-19. *Cell* **183**, 1340–1353.e16 (2020). doi: 10.1016/j.cell.2020.10.001; pmid: 33096020
36. A. Sette, S. Crotty, Pre-existing immunity to SARS-CoV-2: The knowns and unknowns. *Nat. Rev. Immunol.* **20**, 457–458 (2020). doi: 10.1038/s41577-020-0389-z; pmid: 32636479
37. D. Weiskopf *et al*., Phenotype and kinetics of SARS-CoV-2-specific T cells in COVID-19 patients with acute respiratory distress syndrome. *Sci. Immunol.* **5**, eabd2071 (2020). doi: 10.1126/sciimmunol.abd2071; pmid: 32591408
38. E. M. Anderson *et al*., Seasonal human coronavirus antibodies are boosted upon SARS-CoV-2 infection but not associated with protection. *Cell* **184**, 1858–1864.e10 (2021). doi: 10.1016/j.cell.2021.02.010; pmid: 33631096
39. M. Sagar *et al*., Recent endemic coronavirus infection is associated with less-severe COVID-19. *J. Clin. Invest.* **131**, e143380 (2021). doi: 10.1172/JCI143380; pmid: 32997649
40. N. Doria-Rose *et al*., Antibody persistence through 6 months after the second dose of mRNA-1273 vaccine for Covid-19. *N. Engl. J. Med.* **384**, 2259–2261 (2021). doi: 10.1056/NEJMc2103916; pmid: 33822494
41. S. Crotty, T follicular helper cell biology: A decade of discovery and diseases. *Immunity* **50**, 1132–1148 (2019). doi: 10.1016/j.immuni.2019.04.011; pmid: 31117010
42. C. E. Gustafson, C. Kim, C. M. Weyand, J. J. Goronzy, Influence of immune aging on vaccine responses. *J. Allergy Clin. Immunol.* **145**, 1309–1321 (2020). doi: 10.1016/j.jaci.2020.03.017; pmid: 32386655
43. A. Grifoni *et al*., A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. *Cell Host Microbe* **27**, 671–680.e2 (2020). doi: 10.1016/j.chom.2020.03.002; pmid: 32183941
44. U. Sahin *et al*., BNT162b2 vaccine induces neutralizing antibodies and poly-specific T cells in humans. *Nature* **595**, 572–577 (2021). doi: 10.1038/s41586-021-03653-6; pmid: 34044428
45. K. S. Corbett *et al*., Evaluation of the mRNA-1273 vaccine against SARS-CoV-2 in nonhuman primates. *N. Engl. J. Med.* **383**, 1544–1555 (2020). doi: 10.1056/NEJMoa2024671; pmid: 32722908
46. C. Rydyznski Moderbacher *et al*., Antigen-specific adaptive immunity to SARS-CoV-2 in acute COVID-19 and associations with age and disease severity. *Cell* **183**, 996–1012.e19 (2020). doi: 10.1016/j.cell.2020.09.038; pmid: 33010815
47. A. T. Tan *et al*., Early induction of functional SARS-CoV-2-specific T cells associates with rapid viral clearance and mild disease in COVID-19 patients. *Cell Rep.* **34**, 108728 (2021). doi: 10.1016/j.celrep.2021.108728; pmid: 33516277
48. S. A. Lauer *et al*., The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Ann. Intern. Med.* **172**, 577–582 (2020). doi: 10.7326/M20-0504; pmid: 32150748
49. F. Amanat *et al*., A serological assay to detect SARS-CoV-2 seroconversion in humans. *Nat. Med.* **26**, 1033–1036 (2020). doi: 10.1038/s41591-020-0913-5; pmid: 32398876
50. S. Carrasco Pro *et al*., Automatic generation of validated specific epitope sets. *J. Immunol. Res.* **2015**, 763461 (2015). doi: 10.1155/2015/763461; pmid: 26568965
51. M. F. Kotturi *et al*., The $CD8^+$ T-cell response to lymphocytic choriomeningitis virus involves the L antigen: Uncovering new tricks for an old virus. *J. Virol.* **81**, 4928–4940 (2007). doi: 10.1128/JVI.02632-06; pmid: 17329346
52. A. Grifoni *et al*., SARS-CoV-2 human T cell epitopes: Adaptive immune response against COVID-19. *Cell Host Microbe* **29**, 1076–1092 (2021). doi: 10.1016/j.chom.2021.05.010; pmid: 34237248

ACKNOWLEDGMENTS

This work used samples from the phase 1 mRNA-1273 study (NCT04283461) (*9*, *28*). The study was conducted in collaboration with ModernaTX, and funding for the manufacture of mRNA-1273 phase 1 material was provided by the Coalition for Epidemic Preparedness Innovation. We thank E. O. Saphire for providing the spike plasmids and the LJI Clinical Core for healthy donor enrollment and blood sample procurement. **Funding:** The mRNA-1273 phase 1 study was sponsored and primarily funded by the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), Bethesda, MD; in part with federal funds from the NIAID under grant awards UM1AI148373 to Kaiser Washington; UM1AI148576, UM1AI148684, and NIH P51 OD011132 to Emory University; NIH AID AI149644 and contract award HHSN272201500002C to Emmes. This work was funded by the NIH NIAID under awards AI142742 (Cooperative Centers for Human Immunology) (A.S., S.C.) and NIH contract 75N93019C00065 (D.W., A.S.). This work was also supported in part by LJI Institutional Funds and the NIAID under K08 award AI135078 (J.M.D.) **Author contributions:** Conceptualization: A.S., S.C., and D.W. Methodology: J.M., J.M.D., Z.Z., C.R.M., M.L., B.G., and D.W. Formal analysis: J.M., J.M.D., Z.Z., A.S., S.C., and D.W. Investigation: J.M., A.S., S.C., and D.W. Funding acquisition: A.S., S.C., and D.W. Writing: J.M., J.M.D., Z.Z., A.S., S.C., and D.W. Supervision: A.S., S.C., and D.W. **Competing interests:** A.S. is a consultant for Gritstone Bio, Flow Pharma, Arcturus Therapeutics, ImmunoScape, CellCarta, Oxford Immunotec, and Avalia Immunotherapies. S.C. has consulted for Avalia Immunotherapies, Roche, and GSK. LJI has filed for patent protection for various aspects of T cell epitope and vaccine design work. All other authors declare no conflicts of interest. **Data and materials availability:** Epitope pools used in this paper are available to the scientific community upon request and execution of a material transfer agreement (MTA). All other data are available in the main text or the supplementary materials. 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.abj9853
Figs. S1 to S13
Tables S1 to S8
References
MDAR Reproducibility Checklist
Data S1

16 June 2021; accepted 7 September 2021
Published online 14 September 2021
10.1126/science.abj9853

RESEARCH ARTICLE SUMMARY

SOCIAL NEUROSCIENCE

Social agent identity cells in the prefrontal cortex of interacting groups of primates

Raymundo Báez-Mendoza*, Emma P. Mastrobattista, Amy J. Wang, Ziv M. Williams*

INTRODUCTION: Social groups play a foundational role in the behavior of many animal species. During group interactions, an individual's behavior can affect not only the well-being of others but also how others respond in return. Therefore, representing both the identities of others as well as their actions and outcomes is necessary for the ability to interact successfully within groups. Without these combined representations, it would not be possible to engage effectively with others or to understand how one's actions affect particular group members. It would also not be possible to form mutually beneficial affiliations, avoid exploitation by others, or understand how the behaviors of specific individuals within a group interrelate.

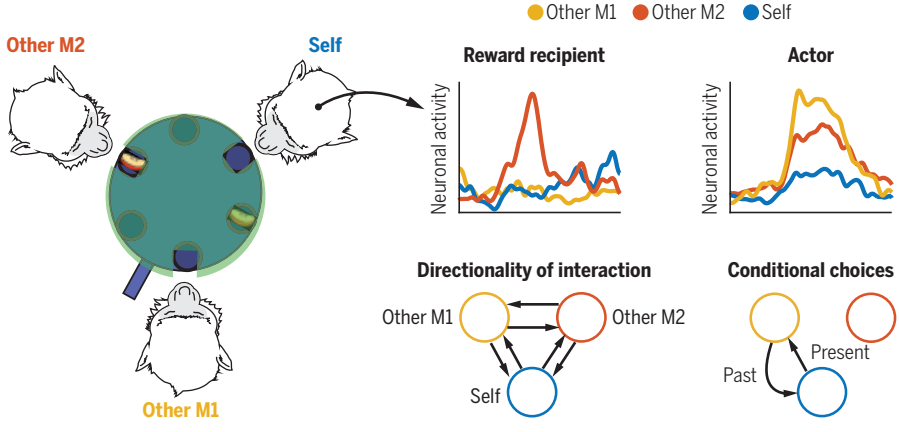
RATIONALE: Most primates, including humans, live in social groups wherein an individual's success relies on the ability to interact effectively with others. Rhesus macaques, in particular, are known to form nonkin interactions and long-lasting alliances. They also engage in mutually beneficial behavior based on reciprocity between individual group members. Yet how neurons in the primate brain precisely represent the interactive behavior of groups or resolve the basic problem of coding for multiple agents remains largely unknown. In this study, we developed a three-agent paradigm using small groups of interacting Rhesus macaques in combination with single-neuronal recording and stimulation techniques to begin addressing these questions.

RESULTS: Groups of adult male rhesus macaques performed a three-agent task in which each individual could offer, in sequence, a food reward to one of the other group members using a rotary table. During these group interactions, we found that the primates reciprocated past offers of reward and retaliated when they did not receive a reward from another. We also observed that the primates formed mutually beneficial affiliations that reflected the reputation and past behavior of other group members. These results, therefore, together suggested that the primates kept track of their interactions with specific individuals within their groups and responded adaptively to the behavior of others.

As the primates performed the task, we obtained single-neuronal recordings from the dorsomedial prefrontal cortex (dmPFC; Brodmann's area 24) in an area proposed to play a role in social behavior and its dysfunction. Collectively, these neurons encoded information about the choices and reward outcomes of different group members, providing, in combination, a highly detailed account of the group's interactions. These neurons tracked the others' behaviors across spatial locations, reflecting the social context of the animals' interactions and the agents involved. They also reliably represented interactions between other individuals, even when those interactions did not involve the recorded animals themselves, thus explicitly linking the identities of others with their specific behaviors.

Finally, by following the group's behavior over time, we find that neurons in the dmPFC reliably tracked prior interactions and integrated information about the past behavior of other agents to influence the animals' own actions. Modeled together, these neurons reliably predicted the animals' own upcoming decisions to reciprocate or retaliate in response to the behavior of individual group members. Conversely, disrupting neuronal activity in the dmPFC diminished the animals' ability to engage in agent-specific behavior and, therefore, their capacity to form mutually favorable interactions.

CONCLUSION: By following the social behavior of small groups of three rhesus macaques, we find neurons in the dmPFC that code for the agency identity of others and that can therefore link the identities of others with their specific behaviors. By encoding both the current and past actions and outcomes of specific group members, these neuronal ensembles may allow animals to track the behavior of other individuals. In combination with other core areas involved in social cognition, these neurons could form a detailed representation of multiple interactions and together construct a comprehensive "map" of the group's dynamics. By further integrating information about these dynamics to inform upcoming decisions, these populations could therefore crucially enable the formation of agent-specific interactions necessary for the effective social behavior of groups. ■



A three-agent task in rhesus macaques reveals a rich representation of group behavior and agent identity in the dmPFC. (Left) Three monkeys sat around a rotary table apparatus that allowed them to allocate food to each other. Simultaneously, neuronal activity was recorded from their dmPFC. (Top right) Individual neurons represented the specific group members offering (actor) and receiving a reward (recipient). (Bottom right) Neurons in the dmPFC collectively encoded detailed information about specific interactions within the group, the directionality of those interactions, the past behavior of other group members, and their influence on the animals' own decisions.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: zwilliams@mgh.harvard.edu (Z.M.W.); rbaez-mendoza@mgh.harvard.edu (R.B.-M.)
Cite this article as R. Báez-Mendoza *et al.*, *Science* 374, eabb4149 (2021). DOI: 10.1126/science.abb4149

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abb4149

RESEARCH ARTICLE

SOCIAL NEUROSCIENCE

Social agent identity cells in the prefrontal cortex of interacting groups of primates

Raymundo Báez-Mendoza[1]*, Emma P. Mastrobattista[1], Amy J. Wang[1], Ziv M. Williams[1,2,3]*

The ability to interact effectively within social groups is essential to primate and human behavior. Yet understanding the neural processes that underlie the interactive behavior of groups or by which neurons solve the basic problem of coding for multiple agents has remained a challenge. By tracking the interindividual dynamics of groups of three interacting rhesus macaques, we discover detailed representations of the groups' behavior by neurons in the dorsomedial prefrontal cortex, reflecting not only the other agents' identities but also their specific interactions, social context, actions, and outcomes. We show how these cells collectively represent the interaction between specific group members and their reciprocation, retaliation, and past behaviors. We also show how they influence the animals' own upcoming decisions and their ability to form beneficial agent-specific interactions. Together, these findings reveal prefrontal neurons that code for the agency identity of others and a cellular mechanism that could support the interactive behavior of social groups.

Social groups play a foundational role in the behavior of most animal species. To interact effectively within social groups, individuals must be able to represent not only the identities of other group members but also their specific behaviors (*1*, *2*). Without such representations, it would not be possible to understand how the actions and outcomes of specific individuals relate or how one's actions affect specific group members (*3*). It would also not be possible to develop mutually beneficial affiliations and avoid exploitation by others (*1*, *4*, *5*). Understanding how neurons in the brain represent the behavior of specific individuals or their interaction within groups, however, has remained a challenge. Although prior investigations have revealed neurons in temporal regions that respond to the specific identities or facial features of others (*6*–*9*), they do not reveal how neurons encode another's behavior or their interaction. Other studies, by comparison, have identified neurons in associative brain regions that respond to another's behavior (*10*–*15*) but do not reveal how neurons represent their specific identities or group interactions.

Most primates live within social groups in which an individual's success relies on the ability to interact effectively with conspecifics (*1*, *4*, *16*). Rhesus macaques, in particular, can recognize different individuals (*6*, *17*) and form nonkin interactions and long-lasting alliances (*18*–*20*). They also engage in mutually beneficial behavior based on reciprocation between specific individuals and keep track of others' behavior (*19*, *20*). By studying the group behavior of rhesus macaques, we can therefore begin to characterize how neurons in the primate brain represent interactions within small social groups and to explore how neurons code for the specific identities and behaviors of others.

[1]Department of Neurosurgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA. [2]Harvard-MIT Division of Health Sciences and Technology, Boston, MA 02115, USA. [3]Program in Neuroscience, Harvard Medical School, Boston, MA 02115, USA.
*Corresponding author. Email: zwilliams@mgh.harvard.edu (Z.M.W.); rbaez-mendoza@mgh.harvard.edu (R.B.-M.)

Three-agent group interaction task in rhesus macaques

We devised a three-agent task in which three adult male rhesus macaques sit at a turntable and each of which could offer a food reward to either of the other two monkeys over successive trials (Fig. 1, A and B). For each successive trial, one of the primates was assigned to be the "actor" and could use a handle on the turntable apparatus to offer a food reward to one of the other two agents ("recipient"). Further, the primate assigned to be the actor would alternate in a pseudo-random fashion from trial to trial (Fig. 1C). Thus, for example, monkey 1 could be the actor in one trial and may offer a reward to monkey 2. On the subsequent trial, monkey 2 could be the actor and may reciprocate that same offer of a reward to monkey 1 or, instead, offer a reward to monkey 3.

Next, to further dissociate the actor's movements from the specific individual receiving a reward, we set the apparatus such that either a clockwise or counterclockwise handle movement allowed the actor to offer a reward to the same animal on different trials (Fig. 1D, left, and fig. S1). To further limit the possibility that animals used simple conditioned responses, the trials also alternated such that the monkey receiving a reward may be different from the monkey offering a reward as the actor in the next trial (Fig. 1D, middle). Finally, to dissociate the location of reward from the specific monkey receiving it, we alternated the physical locations of the primates in relation to one another halfway through the session (Fig. 1D, right).

Therefore, taken together, the nature of this task aimed to mimic some of the basic ethological features that define interactions between primates within groups (e.g., offering and receiving reward or grooming and being groomed) but in a way that could be studied in a neurophysiologic setting. More importantly, it allowed us to examine interactions between specific individuals within the group (e.g., did monkey 1 or 2 receive a reward, and, if so, was a reward given to them by monkey 2 or 3). All trial conditions were controlled in an automated fashion, and all events were recorded and analyzed offline at millisecond resolution (*21*). For each new session and day, a different triad of monkeys was selected from a possible four communally housed adult male macaques. The group performed an average of 105 ± 8.7 (mean ± SEM) trials per session for a total of 22 sessions.

Tracking agent-specific interactions within the primate groups

Behaviorally, the primates reciprocated past offers of reward, suggesting that they kept track of their interaction with specific individuals in their group. Because the actors had to choose between two possible agents, we could examine the interaction between specific group members. Here, we find that the animals were significantly more likely than chance to reciprocate an offer of reward from another animal (9.2 ± 4.0% above chance; signed-rank test, $Z = 2.3$, $P = 0.01$; Fig. 1E, left, and fig. S2A). Therefore, if monkey 1 gave a reward to monkey 2 on a particular trial, for example, monkey 2 would be more likely to offer a reward to monkey 1 on a subsequent trial (*16*, *22*).

The animal's behavior also reflected the specific type of interaction with the other group members. Whereas reciprocation of reward reflects a mutually positive interaction, retaliation reflects a negative one. For example, if monkey 1 gave a reward to monkey 2 in the previous trial, monkey 3 retaliated against monkey 1 by offering a reward to monkey 2 in the next trial (Fig. 1E, middle). Here, we find that the primates were significantly more likely to retaliate than expected by chance (10.2 ± 4.2%; signed-rank test, $Z = 2.30$, $P = 0.01$; Fig. 1E, middle, and fig. S2A). Moreover, when considering both positive and negative interactions together (i.e., "tit-for-tat" strategy in which the current actor gives back to the previous actor if it received reward and withholds reward if they had not received reward) (*16*), we find that the

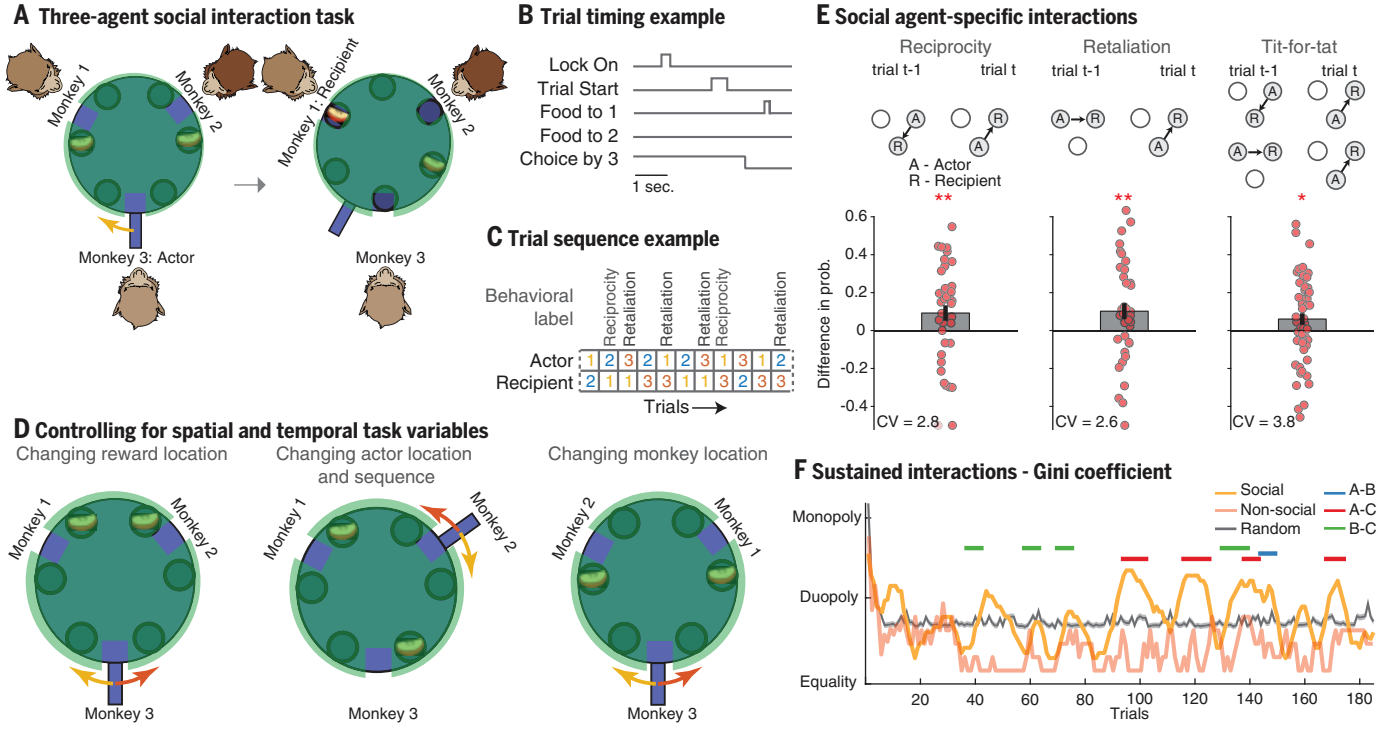


Fig. 1. Three-agent task for testing partner-specific interactions in rhesus macaques. (**A**) Groups of three monkeys sat around a custom-made turntable apparatus that allowed them to interact with each other through food allocation. All monkeys could observe the initial food location through a transparent cover (shown in green), the actor's choice, and the reward recipient. Turntable movement, together with food location, determined the reward recipient (fig. S1). (**B**) Example timing of events for the trial shown in (A). In this trial, monkey 3 (referred to here as "choice by 3") offered a reward to monkey 2 (referred to here as "food to 2"). (**C**) The animals interacted with each other over multiple trials, with the actor on each trial being selected in a pseudo-random fashion. The actor could engage in reciprocity or retaliation based on what the previous actor chose. (**D**) Control measures were used to dissociate the identities of the different agents from variables such as the direction of movement (left), the role of each agent (middle), or the monkeys' spatial locations (right). (**E**) Illustration of trial combinations in which the animals displayed reciprocation, retaliation, and tit-for-tat behavior. Arrows show to whom the actor offered a reward. The animals displayed reciprocation, retaliation, and tit-for-tat behavior with specific individuals at probabilities that were significantly higher than expected from chance (*P < 0.05, and **P < 0.01; CV, coefficient of variation). Bars are the probability of reciprocating compared with not reciprocating ± SEM. Each point depicts an individual's probability within a particular session (fig. S2). (**F**) Gini coefficient illustrates the distribution of reward (dark orange) during a representative session. The highlighted horizontal lines illustrate transient duopolies (permutation test, P < 0.05). For comparison, the distributions of reward expected from chance (gray) and in a representative nonsocial session (light orange) are displayed separately.

animals were more likely than chance to display this behavior (6.1 ± 3.0%; signed-rank test, Z = 1.98, P = 0.023; Fig. 1E, right, and fig. S2A). All animals showed similar reciprocity and retribution across groups [all tests: Z > 1.65, P < 0.04; (*21*)]. These results therefore suggest that the animals kept track of not only whom they interacted with but also how.

The monkeys' behavior did not reflect simple conditioned responses. An important feature of the task was that either a clockwise or counterclockwise movement by the actor could deliver the reward to the same individual on different trials [fig. S1; (*21*)]. Moreover, the pseudo-random sequence in which the role of actor was assigned meant that there was no guarantee that the reward recipient would be the actor on the next trial [Fig. 1C; (*21*)]. Here, we find that the animals were significantly more likely to reciprocate past offers of reward to a specific individual when the offer occurred one or two trials back (7.1 ± 3.2%; signed-rank test, Z = 2.04, P = 0.02). By contrast, the monkeys displayed no evidence of "win-stay-lose-switch" behavior (i.e., simply responding to receipt of reward independently of the specific agent offering it; signed-rank test, Z = 0.4, P = 0.68), together suggesting that the animals responded to the past actions of specific individuals within the group rather than simply the last location from which they received a reward.

Social context dependency and specificity of group interactions

To further confirm that the monkeys kept track of their interactions with specific individuals in the group, we switched their physical seating positions in relation to one another halfway through the session. Using this manipulation, we found that none of the primates displayed a systematic reward assignment preference to a particular location either before or after the switch (signed-rank test, Z = 0.32, P = 0.74). More notably, they continued to display reciprocity with specific animals that had offered them reward on past interactions regardless of seating arrangement (signed-rank test, Z = −0.89, P = 0.37; before versus after change).

Next, we examined the influence that the history of past interactions and social context played in the animal's behavior. Prior studies have shown that the past behavior or "reputation" of specific individuals and their social dominance status can markedly influence how group members interact with them (*23*–*25*). Here, we find that difference in the other's reputation based on past interactions (i.e., how likely they were to reciprocate over the past 20 trials) had a significant effect on the animal's choices [odds ratio (OR) = 1.54, t = 9.2, P = 3.5 × 10^{-20}; fig. S2C]. Furthermore, all animals developed transient duopolies (i.e., consistent runs of reciprocation) at probabilities that were significantly higher than expected from chance (permutation test, P < 0.05; Fig. 1F). Although social dominance did not have an independent effect on the animal's choices

(i.e., on the current trial; reciprocity, $Z = 0.41$, $P = 0.68$; retaliation, $Z = -0.061$, $P = 0.95$; tit-for-tat, $Z = 0.68$, $P = 0.49$; Wilcoxon rank-sum test; fig. S2B), it did play a role when considering the animals' past interactions (i.e., whether past interactions were with a more dominant or subordinate animal; OR = 1.14, $t = 2.1$, $P = 0.035$; fig. S2C).

We also confirmed the social-context dependency of the animals' behavior by replacing the other two primates with distinct inanimate totems while yoking trials from past sessions [fig. S2E; (*21*)]. Here, we find that replacing the other group members with totems led to a loss of reciprocation (signed-rank test, $Z = 0.42$, $P = 0.34$; fig. S2F) and tit-for-tat behavior (signed-rank test, $Z = 1.05$, $P = 0.15$). Together, these results suggest that the animals kept track of who they previously interacted with and that their choices were dependent on the social context of their interaction.

Finally, we used two additional ethological metrics to evaluate the animals' interactions (fig. S3). Consistent with prior field studies demonstrating that primates are more likely to look at the individuals they interact with (*25*–*29*), we find that the monkeys look first [58.9 ± 4.4% versus 50% chance; $\chi^2(1) = 5.35$, $P = 0.021$] and longer (42.6 ± 3.4% versus 57.3 ± 3.5%, nonrecipient versus recipient; $t_{17} = 2.12$, $P = 0.049$, paired t test) at the monkey receiving reward (fig. S3, A to C). We also examined whether differences in facial expressions may have affected the animals' choices. Of the trials tested ($n = 450$), we find that the most common facial expression displayed by the animals (when they were potential recipients of reward) before the actor's choice was affiliative (83.8%, $n = 78$; fig. S3D). These expressions, however, did not alter the overall likelihood that the actor would reciprocate with reward to the expressing monkey [$\chi^2(1) = 1.36$, $P = 0.24$] or that the actor would retaliate [$\chi^2(1) = 0.53$, $P = 0.46$] or engage in tit-for-tat strategy [$\chi^2(1) = 0.004$, $P = 0.94$; fig. S3E].

Single neuronal representations of individuals receiving reward

Based on these findings, we next investigated the relationship between neuronal activity and the real-time interaction dynamics between animals in these groups. Together, we recorded from 521 neurons in the primates' dorsomedial prefrontal cortex (dmPFC; Brodmann's area 24) along the dorsal anterior cingulate sulcus (fig. S4A)—an area previously implicated in social cognition in both monkeys (*10*, *30*–*32*) and humans (*33*–*35*). Only units with a high degree of signal-to-noise ratio, adequate refractory period, and stable waveform morphology were used (Fig. 2A, inset, and fig. S4, B and C). Here, for neuronal analysis, we defined the primate from which neuronal activity was recorded as "self" and the other two agents as "other monkey 1" and "other monkey 2" (Fig. 2A).

We first asked whether certain neurons in the population responded to the reward outcome of specific individuals within the group. Because each recorded animal interacted with two other agents, we could importantly examine not only whether another monkey received a reward but also which specific monkey received it. Focusing on the reward period, we found that 19.9% ($n = 104$) of the neurons displayed a change in their activity when any of the other animals received a reward [two-way analysis of variance (ANOVA) with post hoc testing corrected for repeated comparison across the three agents, $P < 0.01$; (*21*)]. More notably, 9.6% ($n = 50$) of neurons displayed a significant change in their activity only when a specific other individual received reward (i.e., the neurons "preferred" the other monkey; Fig. 2D, top, and fig. S5A), a proportion that was significantly higher than expected by chance given the number of neurons recorded (permutation test, $P < 0.0001$; Fig. 2B, bottom). Figure 2, B and C, illustrates representative neurons recorded from the same animal that responded distinctively to reward received by oneself, any other agent, or a specific-other agent as well as their population dynamic.

Neurons that responded to receipt of reward by specific other agents were largely distinct from those that responded to the animal's own receipt of reward. Overall, 26.4% ($n = 138$) of the neurons displayed a change in their firing activity when reward was received by the recorded animal itself. However, most of these displayed little response to the other's reward, with only 14 neurons displaying a change in their activity to both self-reward and specific-other reward [$\chi^2(1) = 9.7$, $P = 0.001$; Fig. 2B, bottom], results that were largely consistent across statistical analyses (table S1). The responses of these neurons to receipt of reward, by comparison, did not reflect more generalized processes such as a negative reward prediction error. Because any of the three agents could function as actors, we could dissociate signals that reflected another agent's observed receipt of reward from its expectancy (i.e., the animals had no expectancy of reward and therefore held no reward prediction error when they were the actor) (*36*). Consistently, we found that neurons responding to another specific agent's reward displayed no difference in response based on whether a putative negative reward prediction error was present ($n = 50$; rank-sum test, $Z = 0.86$, $P = 0.38$; Fig. 2D, bottom), confirming that they responded selectively to the specific agents receiving reward.

Specificity of neuronal responses to social context and identity

The responses of these neurons to receipt of reward by other specific agents were also robust to differences in their physical locations. It could be argued, for example, that neurons that responded to the other agents may have simply encoded the location of reward rather than the specific agents receiving it. Therefore, to control for this possibility, we switched the locations of the other two animals halfway through the sessions as the primates continued to perform the task. However, we find that only four of the neurons that displayed selectivity to the agents receiving reward also displayed selectivity to reward location [$\chi^2(1) = 23.7$, $P = 1.13 \times 10^{-6}$]. More notably, the neurons that displayed selective responses to particular agents before the switch continued to show similar responses to those same agents afterward (rank-sum test, $Z = 2.86$, $P = 0.004$; Fig. 2E, right, and fig. S5B). Figure 2E, left, illustrates the responses of one such neuron before versus after the switch, with the vertex of the triangle representing maximal neuronal activity for a specific monkey.

We also considered the possibility that these neurons may have responded to lower-level sensory features such as the others' faces (*7*, *37*) independently of their social interaction. For example, simply looking at the other agents may have elicited similar responses. Therefore, to test for this, we examined a separate intertrial control period in which no task was performed but in which the primates gazed freely at the other two monkeys (*21*). We find, however, that even when directly viewing the other animals under this control, only 6.1% ($n = 15$ of 244) of the neurons distinguished between which specific agent the primates were looking at (i.e., based on the recorded animal's eye positions; Fig. 2F). More notably, only two of these neurons overlapped with those that responded to the specific agents receiving reward [$\chi^2(1) = 21.6$, $P = 3.4 \times 10^{-6}$], and the degree to which the recorded animal's gaze modulated these neurons' activities was negligible (rank-sum test, $Z = -4.89$, $P = 9.7 \times 10^{-7}$; Fig. 2F). Therefore, unlike interconnected areas such as those in the temporal lobe (*8*, *9*, *38*, *39*), neurons in this area did not reflect information about the others' faces.

Finally, to confirm that neuronal responses to the other agents reflected the social context of their interaction, we recorded from an additional 403 neurons while the recorded primates performed the nonsocial control (*12*). As before, the two other animals were replaced with distinct inanimate totems while we yoked the distribution of reward from a past session (fig. S2E). Unlike the main task, however, we find that 0.6% ($n = 3$) of the neurons changed their activity based on which specific totem was given reward and at a proportion significantly lower from that observed before [$\chi^2(1) = 40.6$, $P = 1.8 \times 10^{-6}$; Fig. 2G, inset]. Moreover, these differences in neuronal response were not associated with

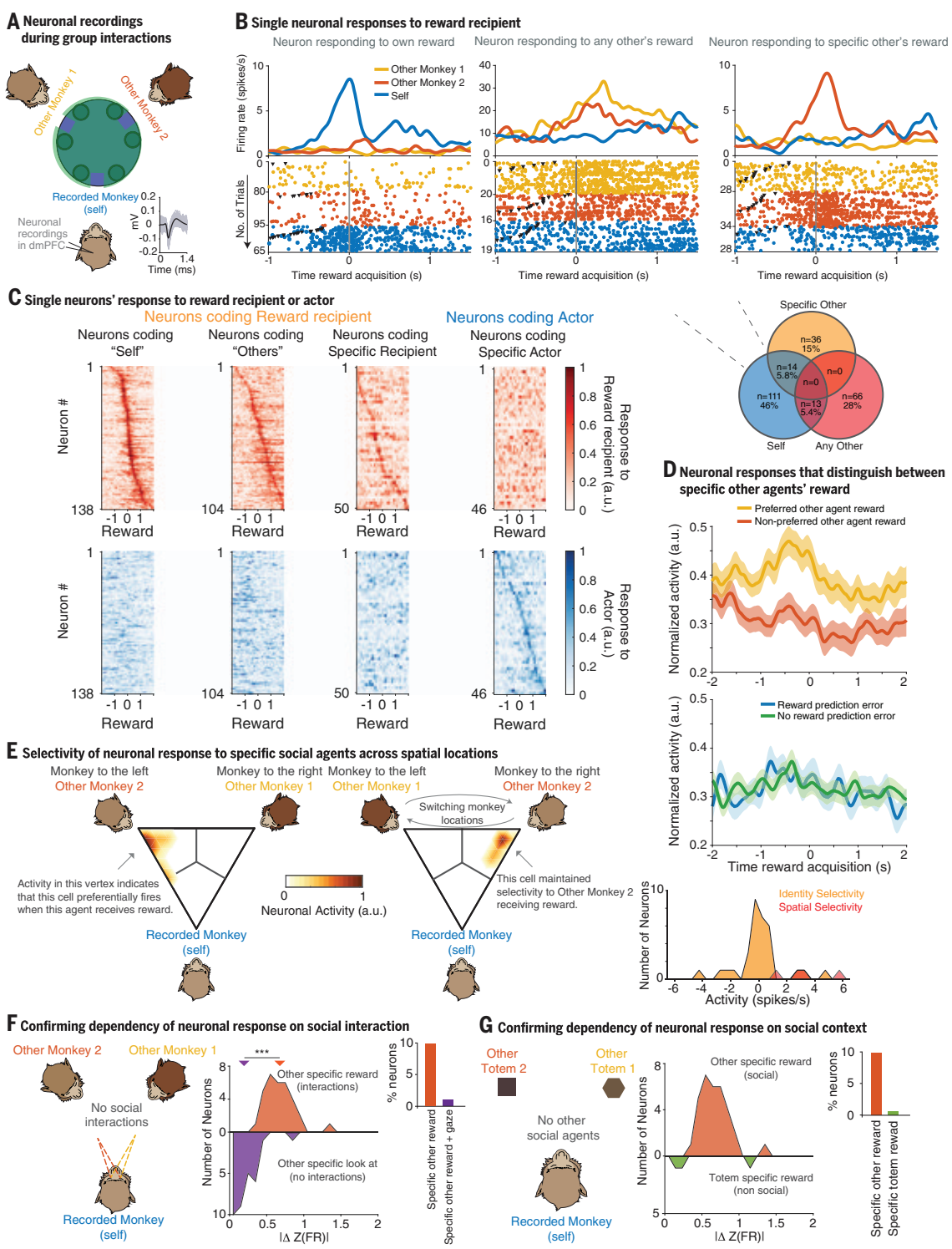


Fig. 2. Selectivity of neurons to specific social agents during group interactions. (**A**) The monkey undergoing neuronal recordings from the dmPFC within each session was referred to as "self" and the two other monkeys as "other monkey 1" and "other monkey 2." Recorded neurons displayed stable waveform morphology (inset; fig. S4). (**B**) Peri-event time histogram and raster examples of neurons that displayed changes in their activities when particular agents within the group received a reward. The inverted black triangles mark when the actor chose. Venn diagram of neurons that displayed response selectivity to reward recipient agency. (**C**) Heatmap of single neurons' responses to reward recipient (top, red) and actor (bottom, blue) aligned to the timing of reward acquisition. Only neurons with significant modulation are shown (ANOVA, $P < 0.01$). (**D**) (Top) Normalized population activity of neurons encoding "specific-other-reward" to the preferred and the nonpreferred other monkey. (Bottom) The same neuronal population as above but parsed by the absence or presence of a possible reward prediction error for "self." (**E**) The locations of other monkey 1 and other monkey 2 were switched halfway in the session to test the selectivity of neuronal responses to specific agents independently of their spatial locations. Heatmap of neuronal activities on a ternary plot before and after the switch of a representative neuron. Here, each vertex represents maximal neuronal activity for a particular monkey. The color code provides the density of activity across trials. The particular neuron displayed here responded almost exclusively to receipt of a reward by other monkey 2 both before and after switching its location relative to the recorded animal. Shown on the right is a histogram of neurons that retained a preferential response to a specific agent ($n = 34$, orange) and neurons ($n = 4$, red) that signaled both reward receipt and location. a.u., arbitrary units. (**F**) To test that neuronal responses were not explained by looking at others' faces, we tracked the recorded animals' eye positions during an intertrial period (left). The middle panel shows the distribution of neurons' activity displaying social agent–specific reward responses based on whether others received reward (top, orange) or whether the recorded animal looked at others during the intertrial period (bottom, blue; normalized to the preferred animal). The proportion of cells is shown on the right (***$P < 0.0001$). IΔ Z(FR)I, absolute difference in Z-scored firing rate. (**G**) The primates performed the same task but in the absence of social agents to test the effect of social context on neuronal responses (left). The middle panel shows the distribution of the absolute normalized difference in firing rates of individual-specific reward neurons based on social agents (orange) or nonsocial agents (green). The proportion of neurons relative to the total number of recorded neurons on each task is shown on the right.

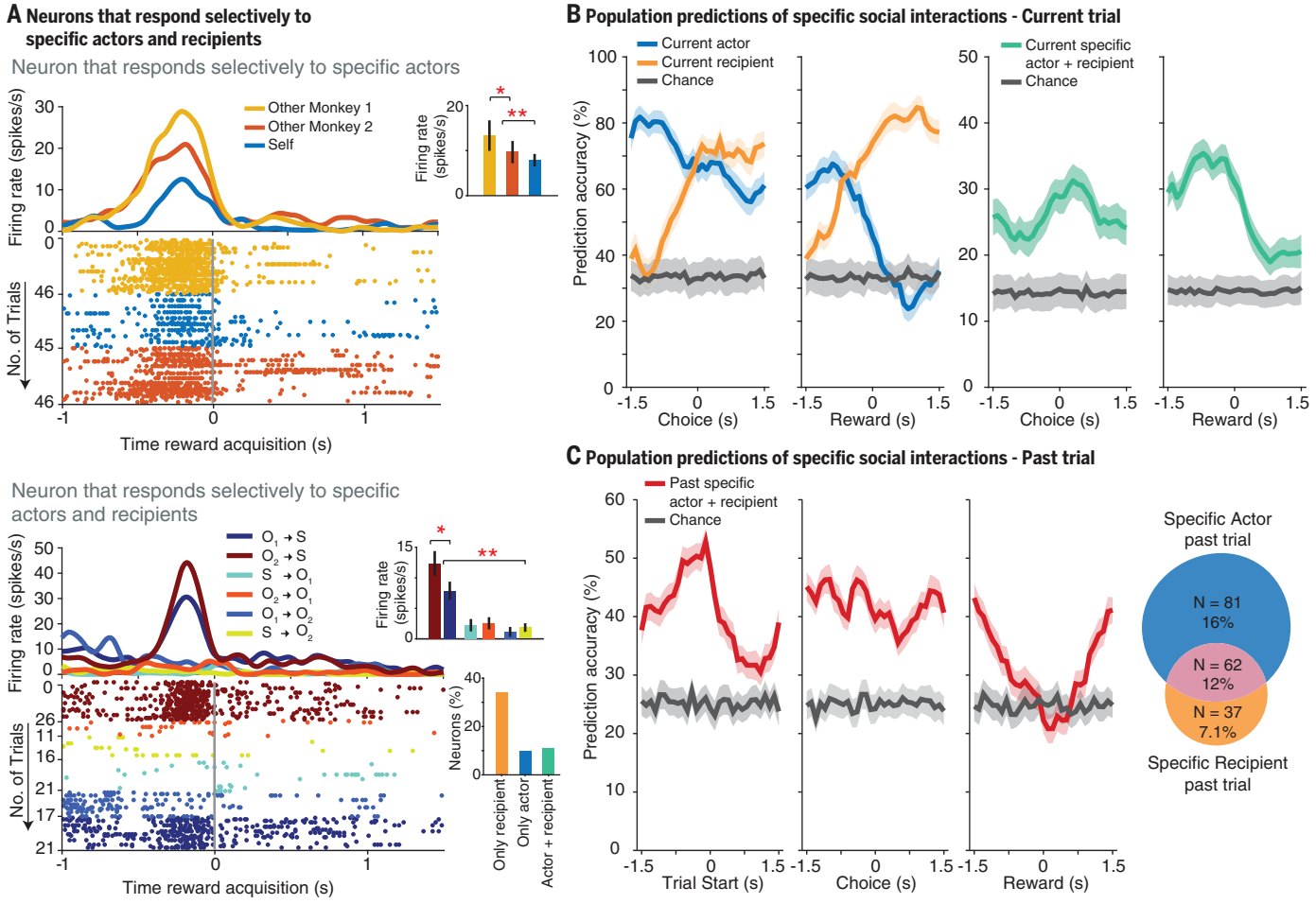


Fig. 3. Neural population predictions of specific interactions within the group. (**A**) (Top) A neuron that displayed a change in its activity based on whether other monkey 1 or other monkey 2 was the actor. (Bottom) A neuron that displayed a change in its activity based on whether other monkey 1 or other monkey 2 was the actor but only when they specifically offered a reward to the recorded. The insets show the average firing rate for each condition during a 1-s time window centered at 0.3 s before the reward was acquired; the error bars represent SEM. The lower inset shows the proportions of neurons encoding specific agent receiving a reward, the specific actor offering a reward, and the combination of the specific actor and recipient across all possible interaction types. *P < 0.05; **P < 0.01. (**B**) (Left) Decoding performance for specific actor and recipient, separately. (Right) Decoding performance for specific interactions in which both the actor and recipient of reward were decoded on a trial-by-trial basis. Multiclass one-versus-all decoders were trained with 80% of trials and tested on the remaining 20% of trials (1-s window advanced in 0.1 s intervals). The colored curves indicate mean prediction accuracy on test trials (±95% confidence interval). (**C**) (Left) Decoding performance for the combination of specific actor and reward recipient in the previous trial when the recorded animal is the actor in the current trial and, therefore, planning its choice. (Right) Venn diagram of the number of neurons displaying selectivity for the specific actor (blue) and the specific recipient (orange) in the past trial.

a change in mean neuronal activity (rank-sum test, $P > 0.5$) and, as noted above, we observed no change in behavioral reaction and movement times to suggest a difference in engagement or attention. Lastly, 0% (n = 0 of 83) of neurons were modulated when a specific other monkey received reward, but when no actor offered it (reward dissociation control; see Materials and methods), together suggesting that the activities of these neurons were indeed dependent on the social context of the animals' interactions.

Neuronal representations of agent-specific actions and interactions

For the primates to effectively interact within these groups, it was necessary for them to know not only who received the reward but also who was the actor that offered it. In our task, any of the three primates could be the actor on a given trial (if they were not the actor on the previous trial) and, in turn, could offer a reward to either of the other two agents. Thus, for example, other monkey 1 may be the actor in one trial and could choose to offer a reward to other monkey 2, or other monkey 2 may be the actor and could choose to offer a reward to other monkey 1. Here, we find that 8.8% (n = 46) of the neurons distinguished between whether other monkey 1 or other monkey 2 was the actor (Figs. 2C and 3A), meaning that they responded differently based on which agent offered reward. Moreover, when considering their group interactions, we find that 11.1% (n = 58) of neurons changed their activity based on which specific animal the actor offered reward to (fig. S5, C and D). Figure 3A, bottom, illustrates such a representative cell, which displays a difference in activity based on whether other monkey 1 or other monkey 2 offered reward to the recorded animal but displays little or no difference in activity for any other interaction.

Next, given these observations, we asked whether and to what degree these neural populations were predictive of interactions within the group and the identities of the specific agents involved on a per-trial basis. Here, we trained multiclass decoders on the

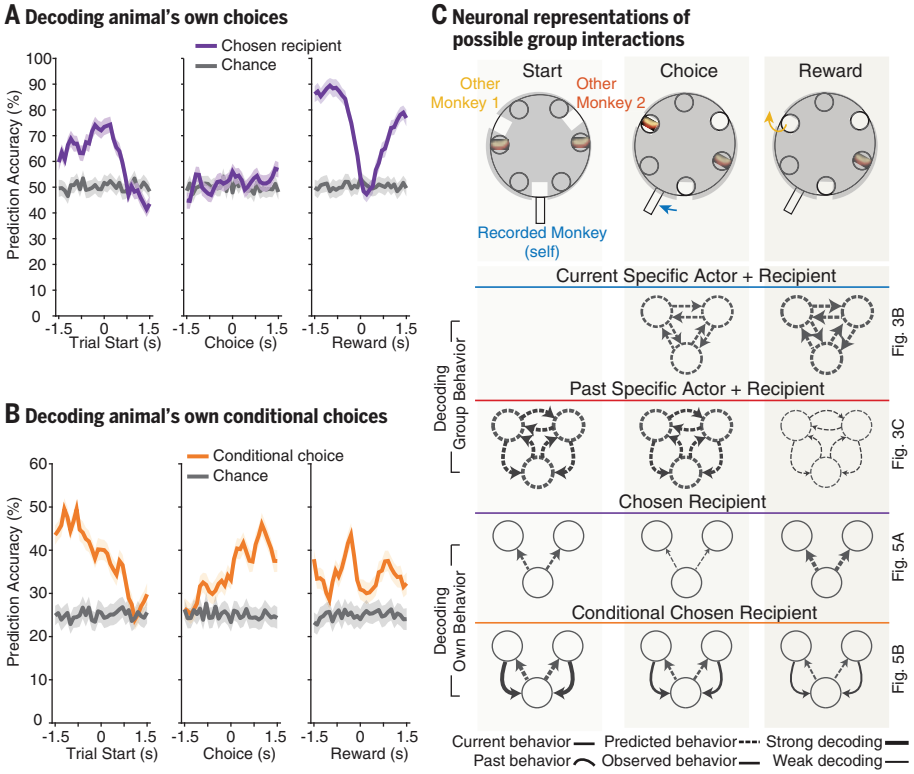


Fig. 4. Dependency between past interactions and predictions of upcoming choices. (**A**) Neuronal responses accurately predicted the animal's own upcoming choices before making its motor selection. (**B**) The animal's past interactions modulated neuronal predictions of the animal's upcoming choice. By considering both the other monkeys' choices and the recorded animals' current choice combinations, the curves here reflect neuronal population predictions contingent on the other's past actions. (**C**) Summary of decoding results. Each column corresponds to one distinct epoch and each row to the relevant information decoded. The arrows reflect the actor (circle) offering a reward to another agent. Each set of arrows reflects the possible combinations of current and past behavior, predicted and observed behavior, and the relative strengths of decoding. Thus, for example, thick arrows indicate that those specific interactions could be highly accurately decoded from the neural population response, whereas thin arrows indicate that decoding accuracy for those interactions was poor when compared with chance. The relevant figures for each panel are shown on the right to allow for ease of comparison.

neuronal responses of 80% of matched trials from all recorded cells and tested the model's performance in the held-out sample [performed in 1-s windows advanced in 0.1-s intervals; (*21*)]. We find that the identity of the specific actor could be decoded with an accuracy of 81.7 ± 2.9% (mean ± 95% CI) before choice selection (Fig. 3B, left), meaning that these neurons could be used to predict which specific agent offered reward. As the trial progressed, however, prediction accuracy increased for the specific identity of the reward recipient, with a decoding accuracy of 70.1 ± 2.8% once the actor made its choice. More notably, both the actor (72.8 ± 3.1%) and the recipient of reward (72.75 ± 3.3%) could be accurately decoded even when confining our analyses to interactions between other monkey 1 and other monkey 2 (i.e., excluding the recorded animal as the agent; fig. S6A).

Collectively, the activities of these neurons held detailed representations about specific interactions within the group. Peak decoding accuracy for agent-specific interactions was 35.4 ± 0.78% and significantly higher than chance shortly before reward was acquired (chance is 16.6%, given the number of possible actor-recipient combinations; $P < 0.01$, permutation test; Fig. 3B, right). The highest decoding performances were for interactions that specifically resulted in reward for the recorded animals (52%), meaning that they were predictive of who specifically offered a reward to them. Decoding accuracy for the agent to whom the recorded animals offered reward was slightly lower at 44%. Similar decoding performances were also observed for "mixed-selectivity" neurons (*40*) that encoded information about both the specific actor and recipient of reward [fig. S6, B and C; (*21*)] as well as when comparing decoding performances across spatial locations [fig. S7A; (*21*)]. Decoding accuracy for more basic sensorimotor variables such as movement direction ($P > 0.2$, permutation test; fig. S7B) or the direction of gaze ($P > 0.2$, permutation test), conversely, was at chance. The activities of these neurons, therefore, appeared to hold detailed information about which specific individuals in the group interacted with whom.

Effect of past interactions on neuronal responses and upcoming decisions

Last, we asked how the neural population responses may relate to the animal's own decisions. To interact effectively, the actor had to take into consideration past interactions with other agents when making decisions. Here, we find that 15.2% ($n = 79$) of the neurons displayed a difference in response on the current trial (t) based on who was the specific actor on the prior trial (t-1; $P < 0.01$), whereas 6.7% ($n = 35$) displayed a difference in response based on the past trials' specific recipient of reward (Fig. 3C, right). From all population neurons, we could decode information about past (t-1) interactions within the group on trials (t) in which the monkey was the current actor with an accuracy of 52.8 ± 1.9% [null hypothesis (H_0) = 25% chance; $P < 0.01$, permutation test; Fig. 3C, left]. When further accounting for the actor's own current decisions, these neural populations could predict the animal's upcoming choices contingent on the other agent's past actions with an accuracy of up to 49.5 ± 1.0% (H_0 = 25% chance, permutation test, $P = 0.005$; Fig. 4B). In other words, the activities of these neurons could be used to accurately predict whether the recorded animal will reciprocate or retaliate in response to the other's past choices (i.e., rather than simply based on any receipt of reward, irrespective of which social agent offered it). Overall, peak decoding accuracy for whom the recorded animals will offer a reward before their motor responses was 74.3 ± 1.4% and significantly higher than chance (H_0 = 50% chance; Fig. 4A). Figure 4C further illustrates these decoding performances across the different group interactions and how they relate the animal's own choices. Taken together, these dmPFC neurons therefore appeared to predict the animal's upcoming decisions based on past interactions with specific agents in its group.

Effect of stimulation on agent-specific interactions

Next, based on these observations, we asked whether and what causal role the dmPFC may have played in the animals' decisions during these group interactions. As noted above, the primates reciprocated past offers of reward from specific individuals to enact mutually beneficial interactions. They also displayed evidence of retaliation against individuals who did not, behaviors that are often naturally seen within primate groups (*41*–*43*). Therefore, to

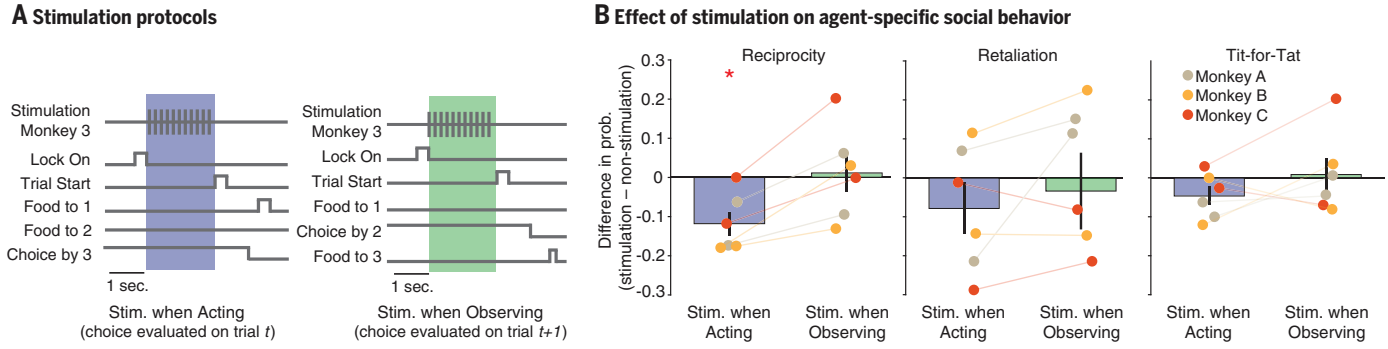


Fig. 5. Effect of stimulation in the dmPFC on group interactions and its selectivity. (**A**) Brief event-triggered electrical stimulation was delivered bilaterally to the dmPFC (200 Hz, 0.1 mA over 2 s, given between locking of the apparatus and trial start) as the primates performed the same task as before. Stimulation was given either when the animal was the actor (blue background) or when they were the observer (green background) for control comparison. (**B**) The bar plot provides the mean difference in probability of reciprocating, retaliating, or using the tit-for-tat strategy on stimulated versus nonstimulated (baseline) trials ± SEM. Each point depicts individual sessions color coded by the animal receiving stimulation. Additional controls used to confirm that stimulation did not affect more basic motoric behavior or cognitive processes such as attention are described in the main text. For specific comparisons, *$P < 0.05$.

further study this question, we used event-triggered stimulation delivered bilaterally to the dmPFC (200 Hz, 0.1 mA over 2 s; from lock on to trial start; Fig. 5A) as the primates performed the same task as before. To allow for control comparison, stimulation was given to the animals on randomly interleaved trials divided equally between those in which the stimulated primate was the actor and observer (*21*).

Before proceeding with the main task, we confirmed that stimulation did not have non-specific effects on the animals' motoric behavior. Overall, we find that the animals displayed similar reaction times (signed-rank test, $Z = 0.07$, $P = 0.94$) and a similar likelihood of selecting one direction over another (signed-rank test, $Z = 0.04$, $P = 0.96$) when comparing stimulated versus nonstimulated trials. We also confirmed that stimulation did not disrupt the monkeys' ability to make appropriate choices. Here, in a separate control task, we allowed the animals to deliver reward to themselves by moving the turntable handle [i.e., without other agents; fig. S1D; (*21*)] but find that stimulation did not affect the animals' performances (100% correct performance for both stimulated, $n = 27$, and nonstimulated, $n = 24$, trials). Next, we considered the possibility that stimulation might affect their memory of past events or more complex strategic behaviors by evaluating the animals' likelihood of enacting win-stay-lose-switch strategies. These canonical strategies represent decisions in which the animal repeated the last choice made on the prior trial if they received a reward irrespective of who offered it (*21*). However, we again find that stimulation did not affect the animals' likelihood of enacting this strategy ($F_{2,15} = 0.67$, $P = 0.52$; fig. S8A). Lastly, we verified that stimulation did not affect the animals' likelihood of looking at particular animals after receiving reward ($F_{2,119} = 0.145$, $P = 0.86$, for stimulation condition; fig. S8B), together confirming that the effect of stimulation was specific.

Finally, based on these findings, we considered the primates' interactions with the other group members. Evaluating the primates' behavior on trials in which they were the actor during the main task, we find that stimulation led to an 11.8% drop in their likelihood of reciprocating past offers of reward from another specific agent ($F_{2,15} = 4.8$, $P = 0.02$). In other words, stimulation diminished their propensity to offer reward to the specific agent from which they received a reward on a previous trial, an effect that was consistent when examined across the different agents (post hoc test, $P = 0.013$; Fig. 5B). By contrast, stimulation had little effect on the animal's likelihood of retaliating in response to past negative interactions or using tit-for-tat strategy ($F_{2,15} = 0.43$, $P = 0.65$; $F_{2,15} = 0.91$, $P = 0.42$; respectively, Fig. 5B) and had no effect on response variability (rank-sum test, $P > 0.2$ for reciprocity, retaliation, and both strategies) to suggest a generalized disruption of behavior. More notably, stimulating when the animal observed the others' choices did not affect the animal's likelihood of reciprocating past offers of reward (post hoc test, $P = 0.80$), together suggesting that stimulation had a temporally selective effect on the primates' ability to enact mutually positive interactions with specific agents in their groups.

Discussion

Most animals, including humans, live within social groups in which they interact with many other group members. The basic cellular processes that precisely underlie group behavior or by which neurons represent specific group interactions, however, have remained poorly understood. Here, we identify neurons in the primate dmPFC that responded selectively to the actions and outcomes of specific group members. By recruiting different subsets of cells to represent the specific actions and outcomes of each individual, these neural populations encoded information not only about the behavior of individuals but also about the directionality of the interactions between them, even when the recorded animals themselves were not involved. Together, these findings identify cells in the primate dmPFC that encode the "agency identity" of others, meaning that they encode information about the behavior of specific individuals. Such computations are essential for effective social behavior.

Another notable finding is that many of the neurons encoded information not only about the actions and outcomes of specific individuals within the group but also about their past behaviors. Moreover, neural predictions of the animal's own upcoming decisions were modulated by the other agents' past actions, suggesting that the animal's upcoming decisions to reciprocate or retaliate were influenced by past interactions with specific group members. Consistent with these observations, stimulation of the dmPFC had a selective effect on the animal's ability to reciprocate past favorable interactions with specific individuals while having little effect on other aspects of its decisions, social viewing preferences, or motoric responses. That stimulation of the dmPFC affected social choices, but not social orienting behaviors, indicates potentially different parallel systems underlying these behaviors. Together, they also suggest that neuronal activity in this area is necessary for mediating mutually

beneficial interactions with specific individuals within these social groups.

Collectively, these findings begin to elucidate the neuronal computations that underlie social group interactions and the role that the dmPFC may play in this process. They also identify neurons capable of encoding the actions, outcomes, and past behavior of specific agents. Given its broad connectivity with temporal regions such as the fusiform gyrus and amygdala that are known to respond to the identities and facial features of others (*6*, *37*, *44*), the dmPFC may be particularly well suited for holding representations of specific group members and mediating mutually favorable interactions. Moreover, the rich representation of agency-specific action and reward recipient in dmPFC could be potentially used for monitoring the consequence of social actions between specific group members, a hypothesized function of the medial prefrontal cortex (*36*, *45*, *46*). Together with other areas proposed to be involved in social cognition (*47*, *48*), the dmPFC may play a core role in orchestrating the interactive social behavior of groups.

Materials and methods

Animals

All procedures were performed under approval by the Massachusetts General Hospital institutional review board and were conducted in accordance with Institutional Animal Care and Use Committee (IACUC) guidelines. Four adult male *Macaca mulatta*, weighing 10 to 13 kg, were included in the study. The animals were singly housed, in the same room, together with other adult male macaques. All animals lived in these settings for more than 2 years and were highly familiar with each other. For each day of testing, three of the four primates were selected for task performance. They were brought individually to a separate neurophysiology room. They were then seated in primate chairs around a custom-made apparatus that allowed them to interact with each other under controlled settings (below). All animals were kept on 80% fluid restriction from ad libitum consumption. The fluid was supplemented through task performance (fruit consumption) based on weight.

Task apparatus

The three monkeys sat in primate chairs that were equally spaced around a custom-made turntable apparatus that allowed them to interact with each other through the allocation of food (Fig. 1, A and B, and fig. S2E). The apparatus consisted of two 35-cm-diameter disks made of transparent acrylic superimposed on each other. The bottom disk had six food wells that were each 2.5 cm in diameter and positioned 12.5 cm away from the center, spaced in 60° intervals. The top disk had three equidistant 3-cm square openings that allowed the animals to retrieve food from the wells. The bottom disk was attached to a 20-cm-diameter bearings turntable, which in turn was attached to a pedestal. The bottom disk had a round 8-cm-long handle that the actor could use to move the turntable either clockwise or counterclockwise. The turntable was made in-house.

Turning the handle allowed the actor to align one of the six food wells with a specific opening such that they could offer a food reward to one of the other two agents (Fig. 1D and fig. S1). Further, by using this apparatus, all three monkeys could determine who the agent was and which specific agent received reward. More importantly, it allowed us to evaluate interactions that were salient to the animals (e.g., offering and receiving reward) and in a way that can be studied under neurophysiologic settings.

The precise timing of the agent's choices and retrieval of food were acquired with infrared-based receivers and emitters (Adafruit). Eye positions were recorded and sampled at 125 Hz via an infrared camera (ISCAN). All behavioral data were acquired at 1-kHz resolution and simultaneously synchronized to the neuronal recordings (MAP, Plexon). Validation of all trial events was also separately made using an overhead video system (Kinect 2, Microsoft).

Behavioral task

For each new session and day, three monkeys were placed in primate chairs (Crist instruments) situated around the turntable. The chairs had an opening that allowed them to perform the task with the right arm. Before starting each trial, two of the wells were baited with a food reward, and the handle was placed in front of the acting animal's arm. The turntable disk was held in place with a linear actuator (Actuonix) under the control of a microcontroller (Arduino). Each trial started with an auditory tone (0.5 s long, Adafruit) coupled with the release of the turntable through disengagement of the actuator. After this, the agent allocating food on a particular trial (actor) moved the turntable using the handle to offer food to one of the other two agents (recipients). The moment the actor moved the handle, an infrared (IR) receiver-emitter (Adafruit) captured the start of choice execution. The recipient monkey then picked up the food from the apparatus, triggering a separate IR event captured by another IR receiver-emitter and concluding a single trial. Therefore, there was a variable delay between the actor's choice and when the recipient collected the food reward.

To ensure that reward allocation was salient to all three monkeys, food delivered to the animals consisted of precut apple cubes measuring about 0.125 cm^3. Individual monkeys consumed on average 150 g of apple per session. To dissociate movement direction from the specific monkey receiving the reward, we set the turntable such that either a clockwise or counterclockwise movement allowed the actor to offer a reward to the same animal on different trials. Thus, on half of the trials, a food reward would be placed on the turntable such that a clockwise movement would provide a reward to monkey 1 sitting on its left (Fig. 1A). On other trials, by comparison, the food would be placed so that the same movement would provide a reward to monkey 2 sitting to its right (Fig. 1D and fig. S1).

To further limit the possibility of simple conditioned responses to receipt of reward, the trials alternated such that the monkey receiving reward would not necessarily be the primate offering reward as the actor in the next trial (Fig. 1D, right; further examples are shown in fig. S1). We pseudo-randomized the turn order for each agent following these two rules: An agent could not be the actor two trials in a row, and all agents had to be actors at least once every five trials. The monkeys were given 20 s to provide a response. If they did not, we proceeded to the next trial with the proviso that the next actor would not be the same as the previous actor. On average, the monkeys did not respond, i.e., move the handle, on only 5.9% of the trials.

Finally, to dissociate the location of reward from the specific monkey receiving it, we swapped the physical locations of the two nonrecorded animals halfway through a session (Fig. 1D, middle). This manipulation, therefore, allowed us to test the role of an animal's physical location in the group's behavior as well as in neuronal responses.

Social context control

To evaluate the effect that social context had on the animal's behavior, in separate sessions, we had the animals perform the same task as before but with no other animals present. Here, the primate being recorded was seated with two different monkey-shaped totems placed on top of the other primate chairs. These totems were monkey-shaped figures (30 cm tall; Gund) and had cartoon-like facial features. The totem sitting to the left of the acting monkey had a red shirt to further distinguish it from the other totem. During these nonsocial controls, the reward was allocated by the totems based on trials that were "yoked" from their conspecifics on previous sessions. Here, movement of the turntable by the totems was controlled using a semiautomated system outside the view of the recorded animal (Arduino). We conducted an average of 145.8 ± 15.3 (mean ± SEM) trials per session in eight sessions of the social context control.

Self-reward control

The self-reward control was used to further confirm that neuronal responses reflected the

social context of the animals' interactions rather than simply the actor's choice selections. In these trials, each monkey could move the turntable to retrieve reward for themselves in the presence of the other animals (fig. S1D). Thus, the recorded animals observed and experienced being both the agent of their own reward and the agent of someone else's reward in a social context.

Reward-dissociation control

A reward-dissociation control was used to further confirm that neuronal responses reflected the agency of the other's actions (i.e., which specific social agent offered reward to whom). In this control, all three monkeys were placed in their chairs but, instead of allowing the actor monkey to reach and perform the action, the turntable was moved automatically. Therefore, instead of the monkey performing an action, rewards occurred probabilistically as if the acting monkey had "acted" (i.e., thus dissociating action from receipt of reward).

Eye calibration and gaze

To calibrate the eyes' position, we aligned their horizontal and vertical positions to four fixed locations along the food wells, each spaced 25 cm apart from each other. For each new session, we then validated these corresponding spatial locations with horizontal and vertical positions derived from the eye tracking system by pseudo-randomly rotating food into each of the four wells, that is, akin to how eye tracking systems are often calibrated using four-point orientation with a horizontal screen (*12*). For the primary analyses, eye fixations were defined as eye velocity below 25% of its statistical SD for longer than 0.1 s (*12*, *49*). Relative looking times were defined as the total time the animal fixated at the reward recipient versus the nonrecipient, starting at the choice time and up to 3 s after the other monkey received reward.

Dominance hierarchy measurements

To assess the animal's relative hierarchical rank, we first performed a food priority test in which the animals sat in front of each other in their chairs and both could reach for a food item (*50*). Dominant monkeys reached for the food first. Supplemented by observations of the veterinary staff (*51*, *52*), we observed a consistent linear transitive hierarchy across the four monkeys, denoted here as A > B > C > D.

Single-neuronal recordings

We implanted microelectrode arrays (FMA-32, MicroProbes for Life Sciences) targeting the dorsal anterior cingulate sulcus of the dmPFC bilaterally (two per hemisphere) under stereotactic guidance (David Kopf Instruments; fig. S4) in three monkeys. These monkeys were selected based on their ability to consistently interact with each other and participate in the task over prolonged durations, as detailed below. The mean location of electrode implantation was +32.2 interaural and +3.5 lateral from the midline (*53*). Each array consisted of 36 microelectrodes, two of which were used for reference and two for ground. A craniectomy followed by a durotomy exposing both hemispheres allowed a dorsal approach to implant the electrode arrays in the dmPFC. Electrode leads and their connectors were secured to the skull with the aid of titanium screws (Synthes) and dental acrylic (Jet dental). Each animal was also implanted with a head-restraining device (Crist Instruments) to enable both eye movement and neuronal recordings. Recordings began 2 weeks after surgical recovery. Neuronal data was amplified (2000 to 10,000 times), band-pass filtered (0.3 to 10 kHz), digitized (40 kHz), and stored using a data acquisition system (MAP, Plexon).

A minimum threshold of three standard deviations was used to differentiate neural signals from background noise (i.e., underlying activity not selected as putative units). Putative units were further isolated based on four primary criteria: (i) the units needed to display spikes with a peak-to-trough 0.3 to 0.5 ms long, (ii) their interspike interval had to be larger than 2 ms, (iii) the unit's waveform morphology had to be consistent throughout the session recording, and (iv) the units needed to be separated based on the first three principal components. Only well-isolated units with $L_{ratio} < 0.2$ and isolation distance > 15 were included in the data analyses (fig. S4C) (*54*, *55*). Single-unit isolation and selection was performed in Offline sorter (Plexon). Overall, we recorded neuronal activity in 44 experimental sessions over a period of 90 days.

Neural stimulation

Stimulation was delivered to the dmPFC using low-impedance electrodes (100 to 500 kilohms) implanted in the same area of the dmPFC. Stimulation consisted of a 2-s-long alternating rectangular positive-to-negative pulse sequences at 200 Hz and 0.1 mA. The pulse duration was 0.002 s with cathodal phase leading (STG4008-16mA, Multichannel Systems). These are similar parameters to those used in clinical settings for deep-brain stimulation (*10*, *56*, *57*). Here, stimulation started when the apparatus was first locked and ended at the start of the trial, that is, the time at which the actor could first make their choice (Fig. 5A). Before proceeding with the main tasks, we performed test stimulations to ensure that stimulation did not elicit a motoric response or apparent sensory percept. Using the self-reward control described above, we also ensured that stimulation did not affect their ability to make objectively appropriate turntable movements.

During stimulation sessions, there were three types of trials: (i) trials in which no stimulation was given, that is, sham stimulation, (ii) trials in which stimulation was given when the animal was the actor, and (iii) trials in which stimulation was given when the animal was not the actor but was observing the others. For each session, we ensured that the animals had at least 60 trials for each of these trial types. The electrical stimulation was performed in a pseudo-random fashion.

Behavioral analysis

Evaluating for agent-specific interactions

Several behavioral metrics were used to evaluate whether the animals kept track of their interaction with other individuals in their group. As described previously (*22*), reciprocity was defined as trials in which the current actor offered reward to the agent that had given reward to them on the previous trial. Thus, for example, if monkey 1 offered reward to monkey 2 on trial *t*-1, then monkey 2 would offer reward to monkey 1 on trial *t*. Retaliation, or spite, was defined as trials in which the current actor chose not to offer reward to the monkey if that monkey had not offered them reward on the previous trial (*22*). Tit-for-tat behavior was defined by the combination of reciprocation and retaliation on trial *t-1*. Thus, the actor would offer reward to an agent if they received reward from them the previous trial but would not offer reward to them if they did not. Finally, win-stay-lose-shift or Pavlov behavior was defined as repeating the animal's last choice if they had received reward and changing their choice if they did not. Like reciprocity, retaliation, and tit-for-tat behavior, win-stay-lose-shift behavior requires them to attend to the task and keep past outcomes in memory. However, it does not require them to attend to the specific agents involved (*5*). The animals' choices in relation to each of these strategies were compared using a Wilcoxon signed-rank test ($P < 0.01$).

Evaluating for interaction between specific group members

The Gini coefficient describes, in a single metric, the distribution of wealth within a group and can therefore efficiently quantify the formation of consistent runs of interaction between their members (*58*). In our task, the Gini coefficient illustrates transient duopolies (i.e., runs of repeated reciprocation between two agents, and, thus, wealth accumulation between a specific monkey pair), monopoly (i.e., only one individual receives reward from both other monkeys), or equality (i.e., all animals receive the same number of rewards). To identify significant runs of reciprocation, we simulated each session with a Bernoulli process (2000 repetitions) and labeled the presence of significant deviations, as tested by a permutation

test, as runs of reciprocation ($P < 0.05$). To illustrate the transient formation of duopolies, we estimated the Gini coefficient with a time horizon of eight trials back.

Evaluating the effects of past history and dominance

To further quantify the degree to which agent-specific features such as past behavior or dominance affected the animal's decisions, we fitted a logit model (GLM 1) that took into account the influence that past interactions and dominance had on the animals' preferences. For past interaction, we considered three complementary metrics: (i) the animal's short-term choice history as defined by the other's choices within the past two trials, (ii) their intermediate-term choice history as defined by the difference in the others' "reputation" over the past 20 trials, and (iii) their long-term choice history as defined by the difference in the others' reputation from the previous session (GLM 1):

$$Y_t = \beta_0 + \beta_1 HiGave_{t-1} + \beta_2 HiGave_{t-2} + \beta_3 \Delta Reputation + \beta_4 \Delta ReputationLastSession$$

Here, the dummy variables $HiGave_{t-1}$ and $HiGave_{t-2}$ were 1 if the higher-ranking monkey gave the actor a reward one or two trials back, respectively. To further take into account the intermediate-term effect of choice history, we defined an individual's reputation as the boxcar moving average of reciprocity choices over the past 20 trials, as described previously (*59*). The $\Delta Reputation$ term was the difference in reputations between the other monkeys. Finally, to consider the long-term effect of the other animal's choice history, the term $\Delta ReputationLastSession$ was defined as the difference in the final reputation of the other animals from the previous session.

Next, to further address the relationship between differences in reputation and the session's time course, we created a second model (GLM 2) with differences in reputation terms covering a time course from four trials back up to 15 trials back. We defined, as for GLM 1, an individual's reputation as the boxcar moving average of reciprocity choices, but here we estimated this over the last four ($\Delta Reputation_{[t-4,\,t-1]}$), five, and up to 15 trials back and tested them simultaneously. The potential recipients' reputation was then compared for each term in GLM 2:

$$Y_t = \beta_0 + \beta_1 \Delta Reputation_{[t-4,t-1]} + \beta_2 \Delta Reputation_{[t-5,t-1]} + \ldots + \beta_{12} \Delta Reputation_{[t-15,t-1]}$$

Social communication and choice

A high-resolution time aligned audio-video setup and custom-programmed graphical user interface was used to quantify and analyze the animals' facial expressions and vocalizations. To examine the potential relation between facial expressions or vocalizations and the animals' choices, the recorded animals were videotaped during performance of the standard task (n = 450 trials). Labeling of the animal's facial expressions started 5 s before the start of each trial and up to 10 s after a choice was made. The behavioral labels were defined as follows:

i. Neutral: Resting or observing without overt expression.

ii. Lip smacking or affiliative: Protrusion of the lips and contraction of the muscles surrounding the mouth.

iii. Threat or agonist behavior: Opening of the mouth accompanied by direct stare, with or without showing teeth.

iv. Vocalization: Opening of the mouth accompanied by an audible call.

v. Eating movement: Chewing or moving food from or to the cheeks.

Next, the choices made by each animal were tracked on a trial-by-trial basis to determine whether the presence of a particular facial expression by the potential recipient led to a higher likelihood of reciprocating reward. Overall, the monkeys displayed overt facial expressions (i.e., a non-neutral expression) before the choice was made in 20.6% of the trials. Of these, the animals lip smacked, which is an affiliative behavior, in 83.8% of the trials (n = 78), whereas they displayed a threatening or agonist-type behavior in the remaining 16.1% (n = 15). They displayed only three overt vocalizations across trials.

Evaluating variability in choices with stimulation

To quantify the degree to which stimulation may have simply disrupted the animal's choice selections (i.e., irrespective of whom they interacted with), we calculated the coefficient of variation on trials in which stimulation was and was not given and evaluated the difference between conditions with a rank-sum test.

Neuronal analysis

Single-neuronal analysis

In the first step of our analyses, we explored the relationship of neuronal activity to reward recipient and actor of each task-related activation using two-way ANOVA ($P < 0.01$). The factors and their levels were as follows: reward recipient (self, other #1, and other #2) and actor (self, other #1, and other #2), and the interaction between both factors. We followed this with a Holm post hoc test, which corrects for multiple comparisons. We classified each neuron's activity depending on the results of the post hoc test, as follows. Neurons reflected self-reward when only the factor reward was significant and mean firing rate during own trials was higher than during any of the conspecific's trials. Likewise, a neuron reflected any other reward when activity was higher when any of the conspecifics received reward compared with own reward. Critically, a neuron reflected a specific other's reward when activity was higher during this conspecific's reward retrieval compared to self-reward and to the other conspecific's reward. We used the same method to classify a neuron as coding action but applied to the actor factor. Neurons coded actor for reward when the statistical interaction actor*reward was significant. These neurons revealed coding of reward for an animal only when either that animal or a conspecific performed the action that led to reward. To classify these neurons, we used the same method described above applied to both factors. Finally, to evaluate whether and to what degree neurons reflected who specifically offered or received reward in the prior trial, we used a similar classification approach based on a two-way ANOVA. The factors, however, were now related to what had occurred in the past trial. We used a chi-square test to compare the number of neurons in different categories ($P < 0.05$). For illustration, a heatmap of single neuronal responses were obtained by normalizing each neuron's activities by the maximum minus minimum of neuronal responses across all task conditions (Fig. 2C). Only neurons that displayed significant modulation are shown (ANOVA, $P < 0.01$).

Next, for statistical validation, we fitted a generalized linear model with a Poisson distribution to the activity of each neuron to assess neuronal encoding of agent-specific reward-related. We defined the following GLM 3:

$$FR_i = \beta_0 + \beta_1 SelfActor + \beta_2 OtherActor + \beta_3 SelfRecipient + \beta_4 OtherRecipient$$

where the variable *Self Recipient* was defined as 1 when the recorded animal received a reward in trial i, and −0.5 if any other monkey received the reward. Thus, the β_3 coefficient captured whether the neuronal activity was related to the reward receipt of the recorded animal or any other agent: it was negative if the activity was higher when any other agent received a reward, and positive if the activity was higher when self received a reward. The variable *OtherRecipient*, on the other hand, was defined as −1 if other monkey 1 was the reward recipient, 1 if other monkey 2 was the reward recipient, and 0 otherwise. The β_4 coefficient, thus, captured if the neuronal activity was related to the identity of the other agent, with the sign indicating for which animal the activity was higher: A negative sign indicates higher activity during other monkey 1 trials, whereas a positive sign indicates higher activity during other monkey 2 trials. The same logic was extended to the variables

SelfActor and *OtherActor*. Neurons reflected *self-reward* when the coefficient $\beta_3 SelfRecipient$ was positive and significant [t test; $P < 0.01$, false discover rate (FDR) corrected], and the $\beta_4 OtherRecipient$ was not significant. Likewise, neurons reflected *any other reward* when only the coefficient $\beta_3 SelfRecipient$ was negative and significant, and the $\beta_4 OtherRecipient$ was not significant. Finally, neurons encoding specific other's rewards showed a significant $\beta_4 OtherRecipient$ coefficient, of which the sign indicated for which other agent the neuronal activity was higher. We used the same method to classify a neuron as coding action but applied it to the actor-related variables: *SelfActor* and *OtherActor*. A neuronal response was classified only if the model fit was significant as determined by the deviance test ($P < 0.01$). To further consider the potential effect that relative rank (between the actor and the recipient) had on neuronal response, we tested GLM 4:

$$FR_i = \beta_0 + \beta_1 SelfActor + \beta_2 OtherActor + \beta_3 SelfRecipient + \beta_4 OtherRecipient + \beta_5 ActorRecipientRelativeRank$$

Given that our original GLM 3 is a reduced model of GLM 4, we determined whether GLM 4 was the better model with the deviance test. Overall, 54.8% of neurons incorporated information about the social rank of the animals receiving reward.

Finally, we considered the potential effects of reward expectancy on neuronal response. Because the recorded animal might have expected to receive reward but did not when other monkey (e.g., other monkey 1) gave reward to other monkey (e.g., other monkey 2), a neuronal response to a specific other's reward could be confounded with a negative reward prediction error-like signal, particularly because the dmPFC has been shown to contain signals related to prediction errors (*36*). To dissociate this alternative explanation, we contrasted the neuronal responses when the specific other received reward when the self was the actor versus when the other agent was the actor with a rank-sum test. Critically, when the recorded animal was the actor, it did not expect to receive reward on that trial; hence, when the other agent received reward there was no reward prediction error. By contrast, when the other agent (e.g., other monkey 2) was the actor, the recorded animal might have expected to receive reward, and if it had not received a reward, it would then generate a negative reward prediction error-like signal.

Population decoding

We used multiclass one-versus-all support vector machine (SVM) classifiers with a linear kernel to quantify the information contained in dmPFC neuronal population activity (*60*). The SVM classifiers were trained on a subset of data to find linear boundaries that provided the best separation between patterns of transformed neuronal population activity defined by a grouping variable (for example, whether reward recipient was self, other monkey 1, or other monkey 2). The classifier was thus trained to map the neuronal signals and the behavioral conditions. SVM classification is biologically plausible: A neuron pool could perform a linear classification by comparing its inputs with a stored vector of synaptic weights (*61*).

We aggregated trial-by-trial firing rates of independently recorded dmPFC neurons, that is, by creating pseudo-populations. The dataset was conformed of 1-s time windows that were advanced in 0.1-s intervals starting 2 s before and ending 2 s after an event. We used all recorded neurons that met inclusion criteria for a minimum trial number, without preselection, with the exception of the analyses of pure- and mixed-selectivity neurons. We divided the dataset into 80% for training the multiclass SVM algorithm and 20% for testing the model's predictions. To construct the training and testing datasets, we randomly selected the trials that would constitute each dataset and averaged the corresponding neuronal activity in 25 equally populated bins for each condition to obtain a balanced dataset. Importantly, the training and testing datasets were balanced for each condition. This procedure was repeated 200 times and averaged to obtain a parametric estimate of the accuracy of each decoder and to perform a permutation test. The performance of the classifier was evaluated by comparing the classifier predictions on the test data with the labels for the test data.

We report the proportion of test repetitions that were labeled correctly as prediction accuracy. In all classification cases, chance performance was determined by randomly shuffling the training labels and reporting the proportion of test repetitions that were labeled correctly. Average chance prediction accuracy was similar to the theoretical chance prediction accuracy in all cases given that we used balanced and equally populated bins for each tested condition. The plotted prediction accuracy is aligned to the center of the analyses window, whereas the reported timing of changes in accuracy including peak (or maximum) accuracy reflect the center of this time window.

Eye position analysis

To control for the possible effects of gaze or "lower-level" sensory features related to the identities of the other social agents, we also tracked the recorded animal's eye position. First, we detected the animal's eye fixations during an intertrial control interval in which no task was performed. During this period, the animals did not perform a task and were free to look at their conspecifics. We defined a fixation when eye velocity was below 25% of its statistical SD for longer than 0.4 s as described previously (*49*). Second, we contrasted single neuronal activity while gazing at other monkey 1 or other monkey 2 in the first 0.4 s of eye fixation using a Wilcoxon rank-sum test. We defined a neuron as distinguishing between specific agents if the test was significant, and a chi-square test was used to compare the number of neurons that displayed a response.

Spatial location analysis

To evaluate the effect that spatial location had on the neuronal response, we swapped the physical locations of the two nonrecorded animals halfway through the trials (Fig. 1D, middle). We then contrasted the neuronal responses to the location of the other monkeys, either to the monkey to its left or its right independent of identity, before and after switching the animal's positions with a two-way ANOVA ($P < 0.05$). We provide a ternary plot to visualize the relation between neuronal activity related to the agents' identity and their spatial location. Any given location in the triangle is determined by the magnitude of response when each of the three monkeys received a reward. Each side of the triangle corresponds to one of these magnitudes, and each vertex represents a maximal response when a specific animal received a reward in relation to the other two monkeys. Thus, for example, activity localized to the left-upper vertex before the switch in Fig. 2E indicates that the neuron only responded when a reward was given to the monkey in that location. The color code provides the density of activity across trials.

REFERENCES AND NOTES

1. T. Clutton-Brock, Cooperation between non-kin in animal societies. *Nature* **462**, 51–57 (2009). doi: 10.1038/nature08366; pmid: 19890322
2. J. Sliwa, J.-R. Duhamel, O. Pascalis, S. Wirth, Spontaneous voice-face identity matching by rhesus monkeys for familiar conspecifics and humans. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1735–1740 (2011). doi: 10.1073/pnas.1008169108; pmid: 21220340
3. W. D. Hamilton, The evolution of altruistic behavior. *Am. Nat.* **97**, 354–356 (1963). doi: 10.1086/497114
4. M. A. Nowak, K. Sigmund, Tit for tat in heterogeneous populations. *Nature* **355**, 250–253 (1992). doi: 10.1038/355250a0
5. M. A. Nowak, Five rules for the evolution of cooperation. *Science* **314**, 1560–1563 (2006). doi: 10.1126/science.1133755; pmid: 17158317
6. E. M. Meyers, M. Borzello, W. A. Freiwald, D. Tsao, Intelligent information loss: The coding of facial identity, head pose, and non-face information in the macaque face patch system. *J. Neurosci.* **35**, 7069–7081 (2015). doi: 10.1523/JNEUROSCI.3086-14.2015; pmid: 25948258
7. L. Chang, D. Y. Tsao, The code for facial identity in the primate brain. *Cell* **169**, 1013–1028.e14 (2017). doi: 10.1016/j.cell.2017.05.011; pmid: 28575666
8. S. M. Landi, W. A. Freiwald, Two areas for familiar face recognition in the primate brain. *Science* **357**, 591–595 (2017). doi: 10.1126/science.aan1139; pmid: 28798130
9. S. M. Landi, P. Viswanathan, S. Serene, W. A. Freiwald, A fast link between face perception and memory in the temporal pole.

Science **373**, 581–585 (2021). doi: 10.1126/science.abi6671; pmid: 34210891
10. K. Haroush, Z. M. Williams, Neuronal prediction of opponent's behavior during cooperative social interchange in primates. *Cell* **160**, 1233–1245 (2015). doi: 10.1016/j.cell.2015.01.045; pmid: 25728667
11. P. H. Rudebeck, M. J. Buckley, M. E. Walton, M. F. Rushworth, A role for the macaque anterior cingulate gyrus in social valuation. *Science* **313**, 1310–1312 (2006). doi: 10.1126/science.1128197; pmid: 16946075
12. R. Báez-Mendoza, C. J. Harris, W. Schultz, Activity of striatal neurons reflects social action and own reward. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 16634–16639 (2013). doi: 10.1073/pnas.1211342110; pmid: 24062436
13. U. Livneh, J. Resnik, Y. Shohat, R. Paz, Self-monitoring of social facial expressions in the primate amygdala and cingulate cortex. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 18956–18961 (2012). doi: 10.1073/pnas.1207662109; pmid: 23112157
14. F. Grabenhorst, R. Báez-Mendoza, W. Genest, G. Deco, W. Schultz, Primate amygdala neurons simulate decision processes of social partners. *Cell* **177**, 986–998.e15 (2019). doi: 10.1016/j.cell.2019.02.042; pmid: 30982599
15. J. Sliwa, W. A. Freiwald, A dedicated network for social interaction processing in the primate brain. *Science* **356**, 745–749 (2017). doi: 10.1126/science.aam6383; pmid: 28522533
16. R. Axelrod, W. D. Hamilton, The evolution of cooperation. *Science* **211**, 1390–1396 (1981). doi: 10.1126/science.7466396; pmid: 7466396
17. C. Perrodin, C. Kayser, N. K. Logothetis, C. I. Petkov, Auditory and visual modulation of temporal lobe neurons in voice-sensitive and association cortices. *J. Neurosci.* **34**, 2524–2537 (2014). doi: 10.1523/JNEUROSCI.2805-13.2014; pmid: 24523543
18. D. Maestripieri, *Macachiavellian Intelligence: How Rhesus Macaques and Humans Have Conquered the World* (Univ. Chicago Press, 2008).
19. C. Testard *et al*., Rhesus macaques build new social connections after a natural disaster. *Curr. Biol.* **31**, 2299–2309.e7 (2021). doi: 10.1016/j.cub.2021.03.029; pmid: 33836140
20. C. M. Berman, E. Kapsalis, Models of affiliative relationships among free-ranging rhesus monkeys (*Macaca mulatta*). *Behaviour* **133**, 1235–1263 (1996). doi: 10.1163/156853996X00387
21. See Materials and methods section.
22. S. A. West, A. S. Griffin, A. Gardner, Social semantics: Altruism, cooperation, mutualism, strong reciprocity and group selection. *J. Evol. Biol.* **20**, 415–432 (2007). doi: 10.1111/j.1420-9101.2006.01258.x; pmid: 17305808
23. R. O. Deaner, A. V. Khera, M. L. Platt, Monkeys pay per view: Adaptive valuation of social images by rhesus macaques. *Curr. Biol.* **15**, 543–548 (2005). doi: 10.1016/j.cub.2005.01.044; pmid: 15797023
24. J. T. Klein, R. O. Deaner, M. L. Platt, Neural correlates of social target value in macaque parietal cortex. *Curr. Biol.* **18**, 419–424 (2008). doi: 10.1016/j.cub.2008.02.047; pmid: 18356054
25. S. Ballesta, J.-R. Duhamel, Rudimentary empathy in macaques' social decision-making. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15516–15521 (2015). doi: 10.1073/pnas.1504454112; pmid: 26621711
26. J. I. Flombaum, L. R. Santos, Rhesus monkeys attribute perceptions to others. *Curr. Biol.* **15**, 447–452 (2005). doi: 10.1016/j.cub.2004.12.076; pmid: 15753039
27. T. Hayashi *et al*., Macaques exhibit implicit gaze bias anticipating others' false-belief-driven actions via medial prefrontal cortex. *Cell Rep.* **30**, 4433–4444.e5 (2020). doi: 10.1016/j.celrep.2020.03.013; pmid: 32234478
28. O. Dal Monte, P. L. Noble, V. D. Costa, B. B. Averbeck, Oxytocin enhances attention to the eye region in rhesus monkeys. *Front. Neurosci.* **8**, 41 (2014). doi: 10.3389/fnins.2014.00041; pmid: 24624055
29. C. Krupenye, F. Kano, S. Hirata, J. Call, M. Tomasello, Great apes anticipate that other individuals will act according to false beliefs. *Science* **354**, 110–114 (2016). doi: 10.1126/science.aaf8110; pmid: 27846501
30. K. Yoshida, N. Saito, A. Iriki, M. Isoda, Representation of others' action by neurons in monkey medial frontal cortex. *Curr. Biol.* **21**, 249–253 (2011). doi: 10.1016/j.cub.2011.01.004; pmid: 21256015
31. K. Yoshida *et al*., Single-neuron and genetic correlates of autistic behavior in macaque. *Sci. Adv.* **2**, e1600558 (2016). doi: 10.1126/sciadv.1600558; pmid: 27679817
32. S. W. Chang, J. F. Gariépy, M. L. Platt, Neuronal reference frames for social decisions in primate frontal cortex. *Nat. Neurosci.* **16**, 243–250 (2013). doi: 10.1038/nn.3287; pmid: 23263442
33. J. Rilling *et al*., A neural basis for social cooperation. *Neuron* **35**, 395–405 (2002). doi: 10.1016/S0896-6273(02)00755-9; pmid: 12160756
34. T. E. Behrens, L. T. Hunt, M. W. Woolrich, M. F. Rushworth, Associative learning of social value. *Nature* **456**, 245–249 (2008). doi: 10.1038/nature07538; pmid: 19005555
35. D. Tomlin *et al*., Agent-specific responses in the cingulate cortex during economic exchanges. *Science* **312**, 1047–1050 (2006). doi: 10.1126/science.1125596; pmid: 16709783
36. M. Matsumoto, K. Matsumoto, H. Abe, K. Tanaka, Medial prefrontal cell activity signaling prediction errors of action values. *Nat. Neurosci.* **10**, 647–656 (2007). doi: 10.1038/nn1890; pmid: 17450137
37. D. I. Perrett *et al*., Neurones responsive to faces in the temporal cortex: Studies of functional organization, sensitivity to identity and relation to perception. *Hum. Neurobiol.* **3**, 197–208 (1984). pmid: 6526706
38. S. T. Carmichael, J. L. Price, Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. *J. Comp. Neurol.* **363**, 615–641 (1995). doi: 10.1002/cne.903630408; pmid: 8847421
39. J. Sallet *et al*., The organization of dorsal frontal cortex in humans and macaques. *J. Neurosci.* **33**, 12255–12274 (2013). doi: 10.1523/JNEUROSCI.5108-12.2013; pmid: 23884933
40. M. Rigotti *et al*., The importance of mixed selectivity in complex cognitive tasks. *Nature* **497**, 585–590 (2013). doi: 10.1038/nature12160; pmid: 23685452
41. F. B. M. de Waal, Coping with social tension: Sex differences in the effect of food provision to small rhesus monkey groups. *Anim. Behav.* **32**, 765–773 (1984). doi: 10.1016/S0003-3472(84)80152-9
42. K. Jensen, J. Call, M. Tomasello, Chimpanzees are vengeful but not spiteful. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13046–13050 (2007). doi: 10.1073/pnas.0705555104; pmid: 17644612
43. K. Jensen, Punishment and spite, the dark side of cooperation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**, 2635–2650 (2010). doi: 10.1098/rstb.2010.0146; pmid: 20679108
44. F. L. Hitti, S. A. Siegelbaum, The hippocampal CA2 region is essential for social memory. *Nature* **508**, 88–92 (2014). doi: 10.1038/nature13028; pmid: 24572357
45. 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
46. K. Yoshida, N. Saito, A. Iriki, M. Isoda, Social error monitoring in macaque frontal cortex. *Nat. Neurosci.* **15**, 1307–1312 (2012). doi: 10.1038/nn.3180; pmid: 22864610
47. P. Gangopadhyay, M. Chawla, O. Dal Monte, S. W. C. Chang, Prefrontal-amygdala circuits in social decision-making. *Nat. Neurosci.* **24**, 5–18 (2021). doi: 10.1038/s41593-020-00738-9; pmid: 33169032
48. M. K. Wittmann, P. L. Lockwood, M. F. S. Rushworth, Neural mechanisms of social cognition in primates. *Annu. Rev. Neurosci.* **41**, 99–118 (2018). doi: 10.1146/annurev-neuro-080317-061450; pmid: 29561702
49. C. D. Dahl, N. K. Logothetis, K. L. Hoffman, Individuation and holistic processing of faces in rhesus monkeys. *Proc. Biol. Sci.* **274**, 2069–2076 (2007). doi: 10.1098/rspb.2007.0477; pmid: 17609192
50. N. Fujii, S. Hihara, A. Iriki, Dynamic social adaptation of motion-related neurons in primate parietal cortex. *PLOS ONE* **2**, e397 (2007). doi: 10.1371/journal.pone.0000397; pmid: 17460764
51. D. Zumpe, R. P. Michael, Dominance index: A simple measure of relative dominance status in primates. *Am. J. Primatol.* **10**, 291–300 (1986). doi: 10.1002/ajp.1350100402; pmid: 31979473
52. J. C. Azzi, A. Sirigu, J. R. Duhamel, Modulation of value representation by social context in the primate orbitofrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 2126–2131 (2012). doi: 10.1073/pnas.1111715109; pmid: 22308343
53. G. Paxinos, X.-F. Huang, A. W. Toga, *The Rhesus Monkey Brain in Stereotaxic Coordinates* (Academic Press, 2000).
54. N. Schmitzer-Torbert, J. Jackson, D. Henze, K. Harris, A. D. Redish, Quantitative measures of cluster quality for use in extracellular recordings. *Neuroscience* **131**, 1–11 (2005). doi: 10.1016/j.neuroscience.2004.09.066; pmid: 15680687
55. S. A. Neymotin, W. W. Lytton, A. V. Olypher, A. A. Fenton, Measuring the quality of neuronal identification in ensemble recordings. *J. Neurosci.* **31**, 16398–16409 (2011). doi: 10.1523/JNEUROSCI.4053-11.2011; pmid: 22072690
56. N. P. Bichot, M. T. Heard, R. Desimone, Stimulation of the nucleus accumbens as behavioral reward in awake behaving monkeys. *J. Neurosci. Methods* **199**, 265–272 (2011). doi: 10.1016/j.jneumeth.2011.05.025; pmid: 21704383
57. W.-H. Zhang, Z. M. Williams, Frontal neurons modulate memory retrieval across widely varying temporal scales. *Learn. Mem.* **22**, 299–306 (2015). doi: 10.1101/lm.036806.114; pmid: 25979992
58. A. Nishi, H. Shirado, D. G. Rand, N. A. Christakis, Inequality and visibility of wealth in experimental social networks. *Nature* **526**, 426–429 (2015). doi: 10.1038/nature15392; pmid: 26352469
59. T. Watanabe *et al*., Two distinct neural mechanisms underlying indirect reciprocity. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3990–3995 (2014). doi: 10.1073/pnas.1318570111; pmid: 24591599
60. C. P. Hung, G. Kreiman, T. Poggio, J. J. DiCarlo, Fast readout of object identity from macaque inferior temporal cortex. *Science* **310**, 863–866 (2005). doi: 10.1126/science.1117593; pmid: 16272124
61. K. I. Tsutsui, F. Grabenhorst, S. Kobayashi, W. Schultz, A dynamic code for economic object valuation in prefrontal cortex neurons. *Nat. Commun.* **7**, 12554 (2016). doi: 10.1038/ncomms12554; pmid: 27618960
62. R. Báez-Mendoza, E. P. Mastrobattista, A. J. Wang, Z. Williams, Social agent identity cells in the prefrontal cortex of interacting groups of primates. figshare (2021); https://doi.org/10.6084/m9.figshare.15102027.

ACKNOWLEDGMENTS

We thank M. Jamali, Y. Kfir, A. Khanna, M. Rose, M. Yartsev, F. Grabenhorst, and Y. Vazquez for comments on the manuscript and K. Hu and J. Grondin for animal training. **Funding:** R.B.-M. is funded by NARSAD YI award 25066 and the MGH Fund for Medical Discovery. A.J.W. is an Howard Hughes Medical Institute Medical Research Fellow. Z.M.W. is funded by NIH R01HD059852, NIH R01NS091390, and NIH U01NS123130. **Author contributions:** R.B.-M. and Z.M.W. conceived and designed the study and wrote the manuscript. R.B.-M., E.P.M., and A.J.W. performed the experiments. R.B.-M. analyzed the data. Z.M.W. supervised all aspects of the research. **Competing interests:** The authors declare no conflicts of interest. **Data and materials availability:** Data are deposited on the figshare repository (*62*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abb4149
Figs. S1 to S8
Tables S1 and S2
MDAR Reproducibility Checklist

21 February 2020; resubmitted 12 February 2021
Accepted 24 September 2021
10.1126/science.abb4149

RESEARCH ARTICLE SUMMARY

SOCIAL NEUROBIOLOGY

Cortical representation of group social communication in bats

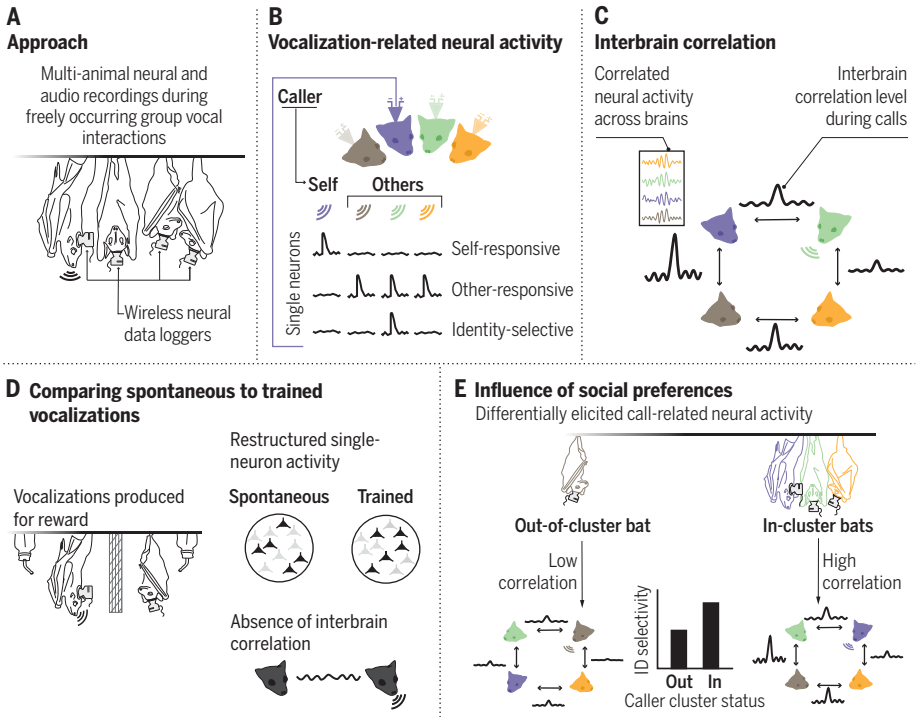
Maimon C. Rose†, Boaz Styr†, Tobias A. Schmid, Julie E. Elie, Michael M. Yartsev*

INTRODUCTION: Many animals live and interact as part of social groups. Social groups present unique advantages to their members, but also necessitate viable modes of communication between individuals, such as vocalization. Yet the neural mechanisms that support group social-vocal communication remain largely unknown. This is because most previous studies have focused on only one side of communication at a time, have measured neural activity in only one brain at a time, and were conducted outside the group social setting. This limited studies of key aspects of group social communication, such as the neural representation of self versus others, individual identity, and the influence of social relationships and context on both inter- and intrabrain activity patterns. Therefore, a major knowledge gap exists between the ethology of group social communication and the study of its neurophysiological basis.

RATIONALE: Egyptian fruit bats live in large groups where individuals form long-lasting social relationships. Communication within these groups occurs through vocalizations that carry socially relevant information relating to individual identity, context, and social relationships. Using simultaneous multi-animal wireless neurophysiological recordings and novel behavioral monitoring techniques, we set out to study the neural basis of social-vocal communication in a group. We focused on the frontal cortex, an area shown to be involved in aspects of vocal and social behaviors across a variety of species.

RESULTS: We allowed groups of bats to interact freely and vocalize as they naturally would when clustered together in the wild. During freely occurring social-vocal interactions, we observed two main patterns of neural activity, one within individual brains and one shared across the brains of group members. Within individual brains, we observed that single-neuron activity could be used to distinguish between vocalizations made by oneself and others. Moreover, the identity of the vocalizing individual could be decoded from the activity of single neurons, beyond what would be expected on the basis of acoustic differences between the vocalizing bats. Across brains, we found that vocalizations elicited synchrony in the power of the high-frequency local field potential between group members. These synchronization patterns were bidirectional and remained stable over weeks of recording. To further dissociate the sensorimotor aspects of the vocal interactions from the social-communicative aspects, we trained bats to vocalize for reward in a setting where vocalizations would not otherwise occur. We found that during trained vocalizations, single-neuron activity was restructured and interbrain shared activity patterns were abolished. Finally, we found that both the intra- and interbrain activity patterns were modulated by the social preferences of individual group members, such that bats preferring to spend more time with others elicited stronger neural identity representation and higher levels of interbrain synchrony.

CONCLUSION: By studying cortical activity in groups of highly social mammals, we uncovered a dedicated neural repertoire for group social communication. We also found that both intra- and interbrain activity patterns were modulated by social context and individual social preferences. Our findings further underscore the benefits of studying neural circuits in specialized model systems while considering ethologically relevant aspects of their behavior. ■



A neural repertoire for communication in social groups. (**A**) We wirelessly recorded single-neuron activity and local field potential (LFP) from the frontal cortices of multiple bats in a group as they freely engaged in social-vocal interactions. (**B**) Single-neuron activity could be used to distinguish between vocalizations made by oneself (self-responsive) and other bats (other-responsive). The activity of identity-selective neurons could be used to further distinguish between the vocalizations of individual bats. (**C**) High-frequency LFP power became transiently correlated across the brains of group members during vocal interactions. (**D**) By comparing spontaneously occurring vocalizations to vocalizations made for reward, we found that single-neuron activity is restructured across contexts and that interbrain correlation is absent during trained vocalizations. (**E**) Vocalizations produced by bats preferring to spend more time with other group members elicited more accurate single-neuron identity selectivity and higher levels of interbrain correlation.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: myartsev@berkeley.edu
†These authors contributed equally to this work.
Cite this article as M. C. Rose *et al.*, *Science* 374, eaba9584 (2021). DOI: 10.1126/science.aba9584

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.aba9584

RESEARCH ARTICLE

SOCIAL NEUROBIOLOGY

Cortical representation of group social communication in bats

Maimon C. Rose[1]†, Boaz Styr[1,2]†, Tobias A. Schmid[1], Julie E. Elie[1,2], Michael M. Yartsev[1,2]*

Social interactions occur in group settings and are mediated by communication signals that are exchanged between individuals, often using vocalizations. The neural representation of group social communication remains largely unexplored. We conducted simultaneous wireless electrophysiological recordings from the frontal cortices of groups of Egyptian fruit bats engaged in both spontaneous and task-induced vocal interactions. We found that the activity of single neurons distinguished between vocalizations produced by self and by others, as well as among specific individuals. Coordinated neural activity among group members exhibited stable bidirectional interbrain correlation patterns specific to spontaneous communicative interactions. Tracking social and spatial arrangements within a group revealed a relationship between social preferences and intra- and interbrain activity patterns. Combined, these findings reveal a dedicated neural repertoire for group social communication within and across the brains of freely communicating groups of bats.

For many animals, including humans, social interactions often occur in group settings. These interactions are commonly mediated by vocal communication signals that convey social information, such as participant identity, context, and social preferences (*1*–*3*). Substantial progress has been made toward investigating the neural representation of sensory, motor, and social aspects of vocalizations separately (*4*–*8*), but rarely have all aspects been examined together (*9*) or in a group setting where behavioral and neural activity were recorded from more than two animals simultaneously. As a result, our ability to connect behavior and neural activity has been limited to individuals or pairs, rather than groups; consequently, it has been difficult to explore the specific social and neural relationships that exist within a group (*10*). A combined approach will allow us to better understand how communication signals are represented within and across the brains of freely interacting group members.

We studied social-vocal communication in groups of Egyptian fruit bats (*Rousettus aegyptiacus*). Similar to other bats (*11*), this species lives in large colonies where individuals form long-term and persistent relationships (*12*–*14*) and communicate using vocalizations that contain socially relevant information, including individual identity and behavioral context (*3*). These vocalizations exclusively occur as part of close-range, direct interactions with conspecifics (*3*, *15*, *16*), providing a reliable indicator of social communication between group members. The social-vocal behavior of the Egyptian fruit bat thus presents an opportunity to investigate neural computations related to key aspects of group communication, including (i) representations of self versus others and individual identities, (ii) shared activity patterns across the brains of group members, (iii) influence of social context, and (iv) impact of social relationships between individuals.

To study social-vocal communication under group conditions, we allowed multiple bats to interact freely while monitoring their behavior in an enclosure in the dark ("free communication session"; Fig. 1A, group sizes of $n = 4$ and $n = 5$ male bats). Under these conditions, the entire group is in close proximity and vocalizations only occur during social interactions (*3*, *15*, *16*). We therefore considered all interactions that included a spontaneously occurring vocalization to be a social-vocal interaction involving all bats in the group. These interactions occurred hundreds of times per session and were nonstereotyped (see supplementary materials for a list of defined behavior types). Because video recordings alone are not sufficient to accurately and consistently assign the identity of the vocalizer (*3*, *15*, *16*), we developed an on-animal wireless vibration sensor that allows for unambiguous identity detection (Fig. 1B and fig. S1). We found that vocalizations varied within and across bats in their spectral and temporal features (fig. S2 and table S1) and that just 1% of vocalizations overlapped in time across bats, indicating that only one bat vocalized at a time during social interactions (Fig. 1B and fig. S1).

Single-neuron activity distinguishes between vocalizations produced by self and others

We wirelessly recorded the activity of 1153 single neurons and local field potential (LFP) activity from the frontal cortices of multiple bats simultaneously in each free communication session (table S2 and figs. S3 and S4; electrophysiological data from $n = 7$ total bats from two separate groups). Activity in this area has previously been shown to relate to social behaviors across a range of mammalian species (*17*–*21*), including bats (*22*). We began by looking for modulation of neural activity when an individual bat vocalized ("self") as well as when it listened to calls produced by other group members ("others"), in accordance with previous studies demonstrating vocalization-related frontal cortical activity in multiple species (*23*–*26*). We found that single neurons modulated their firing rates around the time of vocalizations, both in the vocalizing bat ("self-responsive neurons," 20.7% of neurons; see tables S2 to S4 for subpopulations of neurons) and in the other bats in the group ("other-responsive neurons," 11.7% of neurons) (Fig. 1, C and D, and figs. S4 and S5). These neuronal populations were primarily nonoverlapping (fig. S5D and table S2), indicating that the same area of frontal cortex contains distinct representations of self- and other-generated vocalizations.

A complementary decoding analysis indicated that 26% of all neurons exhibited firing rates containing sufficient information to differentiate between calls made by self and others (table S2). Firing rate modulation occurred at short latencies, specifically during vocalizations, and followed the fast temporal dynamics of vocalization sequences (fig. S6). Furthermore, manual annotation of individual bats' behavior around vocalizations showed that bats engaged in interactive and variable behavior before and after the time of the vocalizations, but that the type of behavior did not significantly affect firing rates for most call responsive neurons (fig. S7).

Self- and other-responsive neurons are not modulated by specific acoustic features, playback, or echolocation production

The neuronal responses we observed could potentially be related primarily to the sensorimotor rather than social aspects of the social-vocal interactions. However, we found that neither self- nor other-responsive neurons modulated their firing rates in relation to the acoustic features of the vocalizations that were produced or heard (figs. S8 and S9). To further dissociate the sensorimotor and social aspects of vocalizations, we conducted an additional set of experiments where bats heard or produced acoustic signals in isolation (fig. S10). In contrast to group social communication, we did not observe any neurons that responded when an isolated bat was exposed to playback of prerecorded vocalizations (Fig. 1E, fig. S10, A and B, and table S2). Similarly, we found that solitary bats flying

[1]Helen Wills Neuroscience Institute, University of California, Berkeley, CA 94720, USA. [2]Department of Bioengineering, University of California, Berkeley, CA 94720, USA.
*Corresponding author. Email: myartsev@berkeley.edu
†These authors contributed equally to this work.

freely and producing tongue-click echolocation pulses (fig. S10, C and D) that recruit many of the same orofacial muscles as vocalizations (*27*, *28*) elicited few echolocation-responsive cells (5.9% of neurons). Notably, there was no overlap between echolocation and self-responsive neurons (Fig. 1E, fig. S10E, and table S2). Collectively, the low rates of responsivity to nonsocial acoustic production and auditory stimulation, as well as the lack of neuronal modulation in relation to acoustic details of vocalizations, indicate that the observed neuronal responses to vocalizations produced by self or others were specific to vocal behavior occurring in a social context.

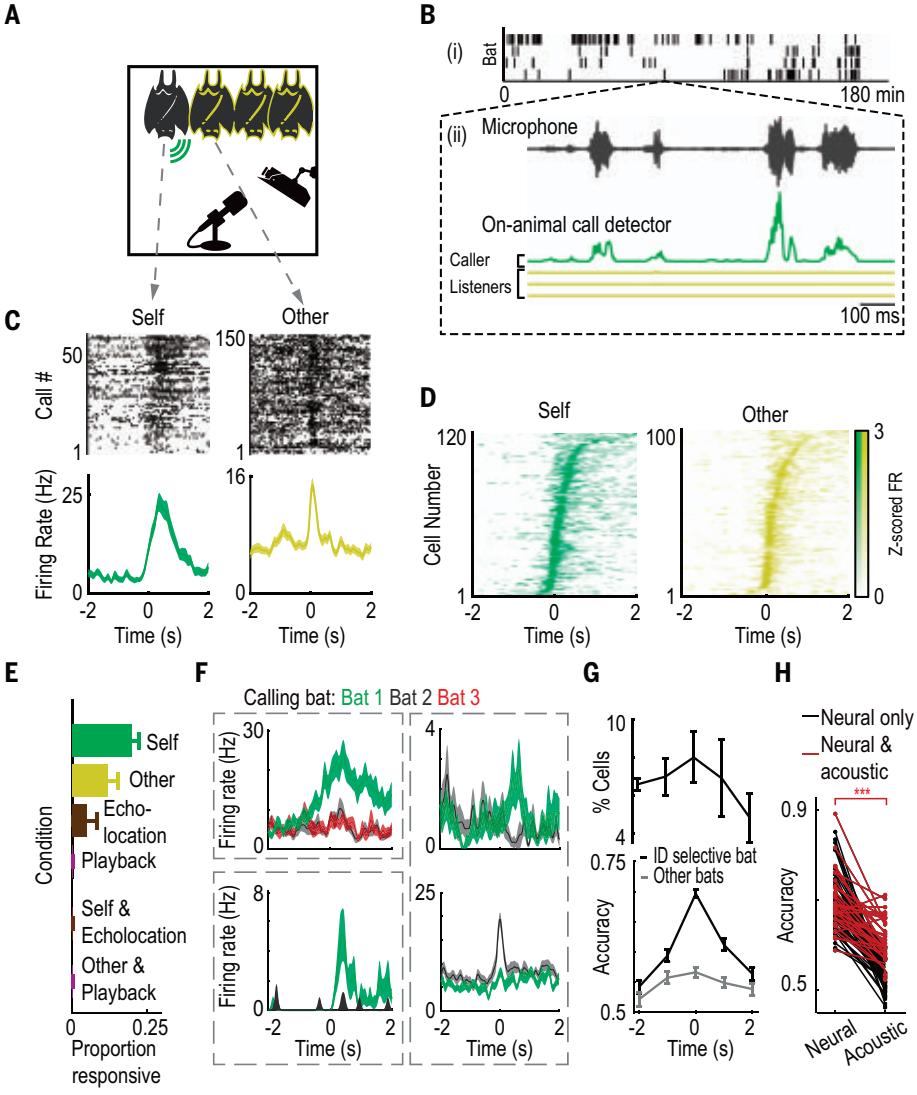


Fig. 1. Neural correlates of social-vocal interactions during free group vocal communication. (**A**) Experimental setup. The calling bat ("self") is indicated by the green symbol; other bats are indicated by a yellow outline. Bats were monitored using cameras and a microphone. (**B**) (i) Ticks indicate vocalizations from different bats on each row (n = 4) during an example session. (ii) Zoom-in on an example train of calls. Top: Microphone recording of example calls. Bottom: Corresponding filtered recordings from on-animal call detectors, colored according to (A). (**C**) Example call-aligned rasters (top) and peri-event time histograms (PETHs, bottom) of self- and other-responsive neurons (here and below, calls start at 0 s). (**D**) Maximum normalized average firing rates of all responsive cells, sorted by time of maximum firing rate, for self- and other-responsive neurons. (**E**) Average fraction of responsive neurons per condition. Error bars denote SEM (n = 7 bats). (**F**) Example identity-selective neurons modulating firing rates in response to one select bat's calls, but not for others (colored traces are one neuron's firing rates when listening to vocalizations from different bats). The two examples at right are identity-selective neurons recorded simultaneously from the same bat, each one selective for a different bat. Shaded areas indicate SEM. (**G**) Top: Percent of all cells that are significantly identity-selective, averaged across bats. Bottom: Identity-decoding accuracy over time averaged across all neurons that were identity-selective at the moment of vocalizations. Plotted are decoding accuracies for the individual bat that drove identity selectivity (black) and for all other bats (gray). Error bars denote SEM. (**H**) Decoding accuracy using neural data (left) and acoustic data (right). Shown are data only from neurons that exhibit significant neural identity selectivity. Instances with both significant neural and acoustic selectivity are in red; instances with only significant neural selectivity are in black. ***$P < 10^{-20}$ (paired t test).

Single-neuron activity exhibits selectivity for calls produced by specific individuals

In addition to discriminating between self and others, the group setting of our experiments enabled us to test whether neural activity in the frontal cortex contains a representation of individual identity—an important aspect of social interactions that has been shown to be behaviorally relevant during vocal communication in this species of bat (*3*, *13*, *14*). We found that a subset of neurons in listening bats responded selectively to vocalizations produced by specific individuals within the group (Fig. 1F). Using the firing rates of these "identity-selective" neurons, we could decode the identity of one bat versus other bats present in the group specifically around the time of vocal interactions (9.3% of neurons; cross-validated and false discovery rate–corrected $P < 0.05$, bootstrap test; Fig. 1, F and G, fig. S11A, and tables S3 and S4). We found that for the majority of identity-selective neurons, acoustic features failed to provide significant classification of caller identity (Fig. 1H, 56% of identity-selective neurons), indicating that these results could not be accounted for solely by acoustic differences across individual bats. Similarly, we found that physical contact and participation in social-vocal interactions did not drive the identity selectivity of single neurons. In particular, single-neuron identity decoding accuracy generally did not significantly decrease, either when restricting our analyses to include only interactions where all bats were in close physical contact (fig. S11B) or when only including interactions not involving the bat from which the identity-selective neuron was recorded (fig. S11C). These analyses do not rule out contributions of unconsidered sensory, social, or multimodal factors that have been shown to modulate frontal cortical activity (*29*) but do indicate that identity selectivity was not solely dependent on proximity, participation in the vocal interaction, or acoustic details. Combined, these results indicate that frontal cortical activity contains sufficient information to distinguish vocalizations produced by self versus others as well as between individual group members.

Vocal interactions elicit stable correlated neural activity across the brains of group members

Having considered the relationship between social-vocal communication and frontal cortical activity in the brains of individual bats,

we next assessed the relationship in neural activity across the brains of group members during vocal interactions. Correlation in brain activity between pairs of individuals has been observed in studies in humans (*30*), duetting birds (*9*), and dyads of socially but not vocally interacting rodents (*20*) and bats (*22*). Interbrain correlation has further been shown to indicate, and possibly facilitate, successful vocal communication in humans (*31*–*34*). Previous findings in bats indicate that LFP activity was modulated on a long time scale (seconds to minutes) by nonvocal social interactions and was correlated across brains (*22*). We observed a strong, temporally precise modulation of LFP power on short time scales (milliseconds) around vocalizations in both low (<20 Hz) and high (>70 Hz) frequency bands (Fig. 2, A and B, and fig. S12). Moreover, we observed increased interbrain correlation precisely around the time of vocalizations between pairs of calling and listening bats as well as between pairs of listening bats (Fig. 2C and fig. S13; n = 8962 bat pairs and calls across two groups). Consistent with previous findings (*22*), pairwise interbrain correlation was found to be most pronounced in high-frequency LFP power (70 to 150 Hz; fig. S13). We therefore focused our subsequent analysis of interbrain relationships on power in the high-frequency LFP range.

The group setting of the experiment allowed us to measure directed interbrain relationships simultaneously between multiple pairs of bats while also accounting for potential spurious correlations driven by common input from other group members using conditional Granger causality (GC) (Fig. 2D). We found that GC magnitude sharply increased around the time of vocal interactions across all calls (Fig. 2E, n = 995 calls and n = 4 bats; fig. S14, A to C, n = 827 calls and n = 3 bats), and also when separately considering the caller-to-listener and listener-to-caller directions (Fig. 2F; n = 7 bats). This suggests a bidirectional interbrain relationship during spontaneous social-vocal interactions between group members, which may reflect vocal as well as other associated behavioral aspects of the interactions between the bats. However, these results do not imply a causal relationship between the behaviors of individuals. The increase in GC

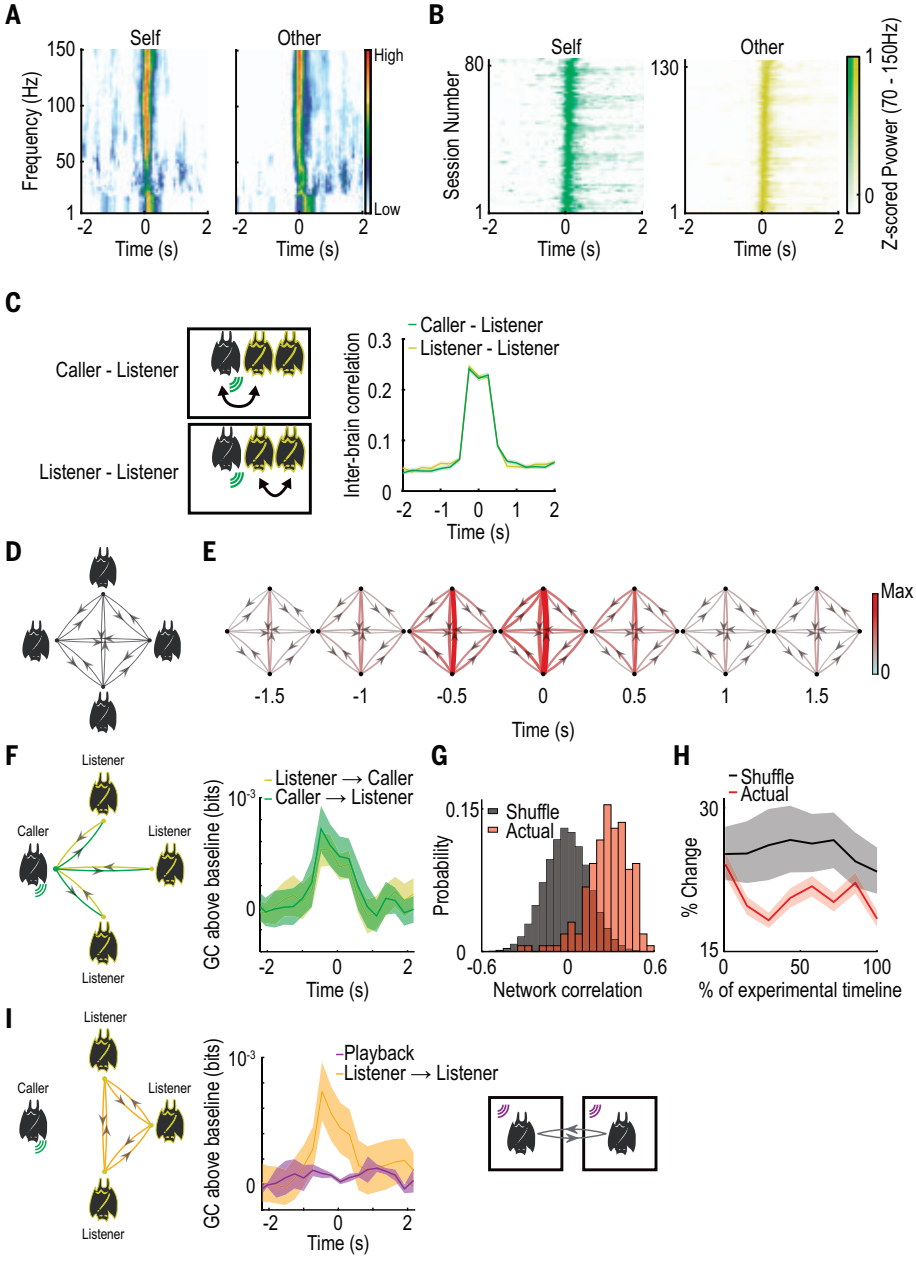


Fig. 2. Group interbrain activity patterns around social vocalizations produced during free group vocal interactions. (**A**) Example average LFP spectrograms (normalized at each frequency bin) from one bat during one session. (**B**) Average power of high-frequency LFP (70 to 150 Hz) for all sessions, displayed as in (A). (**C**) Left: Schematics of interbrain correlation between bat pairs involving the calling bat (top) and involving only listening bats (bottom). Right: Call-aligned interbrain Pearson correlation of high-frequency (70 to 150 Hz) LFP power averaged across all vocalizations and bat pairs. Throughout this figure, shaded areas denote SEM. (**D**) Schematic illustrating all possible connections in a group of four bats. Nodes indicate bats, edges indicate interbrain relationships, and arrows indicate directionality. (**E**) Group interbrain GC graphs using all vocalizations (both calling and listening), calculated using a sliding window of 1-s duration around call onset. GC magnitudes are represented by both color and line width. (**F**) Left: Illustration of the different directional relationships that exist when a given bat is vocalizing. Right: Baseline-subtracted GC values aligned to call onset, calculated separately for vocalizations from each bat and averaged across bats. Data are shown according to the relationships delineated in the diagram at left (colors). (**G**) Distribution of average correlation values between GC magnitudes calculated using vocalizations equally binned across all experimental days (red) compared to shuffled data (gray). High values indicate similarity over time. Actual values are significantly higher than shuffled (P = 0.03, Mantel test). (**H**) Bin-to-bin percent change in GC values over the duration of the experiment, presented as percentage of experimental timeline and compared to shuffled values for the same bins. (**I**) Left: Illustration of listener-to-listener relationships in the group setting. Right: Simultaneous playback experiment. Center: Average baseline-subtracted GC values during playback compared to listener-to-listener GC values during free communication.

was significantly greater than that observed in trial shuffled data, indicating a precise coordination in activity rather than simple coactivation (fig. S14D). Next, we assessed whether the variability in GC we observed across bats, and by listener and caller direction (fig. S14, E and F), reflected random association or a stable pattern of interbrain activity. We found that the correlation in GC values throughout the experimental timeline was significantly higher than in data shuffled across bats (Fig. 2G) and that GC values changed less than what would be expected by chance throughout the experimental timeline (Fig. 2H and fig. S14G). Lastly, we found that providing shared auditory input to two bats in separate, isolated chambers (Fig. 2I) did not elicit an increase in GC magnitude, indicating that social context was necessary for the existence of coordinated interbrain activity between group members (Fig. 2I; n = 2 bats). To verify that these findings were not a result of our choice of parameters or the parametric assumptions of GC, we repeated the above GC analysis using the same data at different values of time lag and also calculated interbrain relationships using the transfer entropy between pairs of bats. Using these measures, we observed similar results (fig. S15).

Intra- and interbrain activity patterns are restructured outside the group social context between spontaneous and trained vocal behavior

The results presented thus far show that social-vocal communication strongly modulates intra- and interbrain neural activity in groups of freely interacting bats. We next sought to dissociate vocalizations from their typical social context in order to assess the impact of engaging in spontaneous, communication-driven vocal interactions. To that end, we developed an operant conditioning task wherein pairs of bats learned to produce vocalizations for reward in a context where they would not normally vocalize. This permitted testing the influence of social context on neural activity by comparing trained, reward-driven vocal interactions to spontaneous group vocal interactions.

In the operant conditioning task, a pair of bats was placed in a divided cage such that they were physically separated but could still see, hear, and smell each other and even interact in a limited way across the mesh divider (n = 4 bats; two pairs of trained bats; Fig. 3A and table S3). Over the extensive course of training (fig. S16A), bats went from the naïve state of never calling in this context (Fig. 3B, Pre-training) to reliably producing hundreds of calls within a session (Fig. 3B, Trained). We found that under these conditions one bat in each pair spontaneously became the sole caller while the other became the listener, in clear contrast to the vocal behavior observed in the free-communication session (Fig. 3C and fig. S16B). Once conditioned, bats produced calls drawn from a similar vocal repertoire as in the free-communication session, but with less variability (Fig. 3D and fig. S16, C and D). To enable a direct comparison of neural activity across contexts, we performed the operant session immediately prior to the free-communication session and recorded the same neural activity throughout (tables S2 and S4). This comparison revealed marked differences between the spontaneous and reward-driven contexts at both the intra- and interbrain activity levels.

At the level of the single brain, certain neurons exhibited "self" or "other"-related response profiles during the operant session, as in the free-communications session (compare fig. S17 and S18A with fig. S4, A to C). However, we also saw substantial differences between the two sessions (fig. S18). Specifically, during the operant session, we observed an increased variability in call-related response latencies (fig. S18E), decreased firing rate consistency when firing rates were aligned to vocalizations rather than to reward (fig. S18F), and an increase in neural activity modulation occurring during the reward period (fig. S18, A and C) relative to the free-communications session. Most strikingly, however, we observed a reorganization of the responsivity of single neurons between the free-communication and operant sessions. We found that 78% (50/64) of neurons responsive in either session were not similarly responsive in both sessions (Fig. 3E, fig. S18, G and H, and table S2), indicating that their responsivity was contextually dependent. This reorganization was evident across the population of responsive neurons, which showed a systematic lack of firing rate consistency across sessions compared to within sessions (Fig. 3F). These differences were not driven primarily by acoustic differences between the calls produced in either session, as indicated by comparing acoustically similar calls across sessions and acoustically different calls within sessions (fig. S18I).

At the level of neural activity across brains, we compared interbrain GC values across session types. We found that the bidirectional interbrain relationship that we observed during communicative behavior breaks down during conditioned vocalizations. Instead of both caller and listener influencing one another (Fig. 2F), we observed no vocalization-related change in GC in the listener-to-caller direction during the operant session (Fig. 3G). As in the single-neuron activity, the observed differences in interbrain correlation were not driven primarily by acoustic differences across sessions (fig. S18J). Combined, these results suggest that the neural activity shifted away from the communication-related representation seen during spontaneous group social communication to a task-related representation in the reward-driven context.

Interbrain correlation and identity selectivity vary with social preferences of group members

Our findings thus far indicate that the neural activity patterns we identified during group communication depend on the social context at the group level. Next, we wanted to understand how these neural activity patterns vary according to social factors at the individual level. To do so, we used the fact that bats demonstrate robust spatial behavioral patterns that reflect social preferences of individuals within the group (*11*, *35*). We developed a system to precisely track the individual positions of a group of eight bats moving freely in a large enclosure [1.2 × 1.2 × 0.6 m (width, length, height); Fig. 4, A and B, and fig. S19] as part of a complementary experimental session (termed the "social-space" session) performed prior to each free-communication session. This approach revealed stable social preferences of group members, which were reflected in the neural activity during vocal interactions in the free-communication session.

During the social-space session, most bats spent the majority of the time, and exclusively vocalized, near other bats (Fig. 4B and figs. S19B and S20). However, individual bats differed in their proximity preferences relative to other group members, resulting in a bimodal distribution of inter-bat distances (Fig. 4C and fig. S21). On the basis of this bimodal distribution, we derived a boundary to define distances between a pair of bats as either close together or far apart (Fig. 4C) and defined that pair's "social dwell time" as the percentage of a session spent close together. Using this metric, we found that the way individuals positioned themselves relative to each other was highly stable over weeks (Fig. 4D and fig. S22). We observed two distinct social-spatial positioning preferences exhibited by different individuals: Some bats consistently spent time far apart from other group members, while others spent the majority of time close together (example shown in Fig. 4E). We quantified these preferences for each individual (Fig. 4, F and G, and fig. S23) and defined the "cluster status" of each bat as either "in-cluster" or "out-of-cluster," reflecting their social-spatial preferences with respect to other group members.

During the corresponding free-communication sessions where all eight bats were close together, we recorded neural activity from the frontal cortex of four of the bats simultaneously. We found that the same neural activity patterns described in the previous smaller groups (n = 4 and n = 5 bats) were replicated in the larger group, including the representation of self versus others, individual identity, and interbrain relationships (n = 8 bats; fig. S24, A and B, and table S3). We additionally

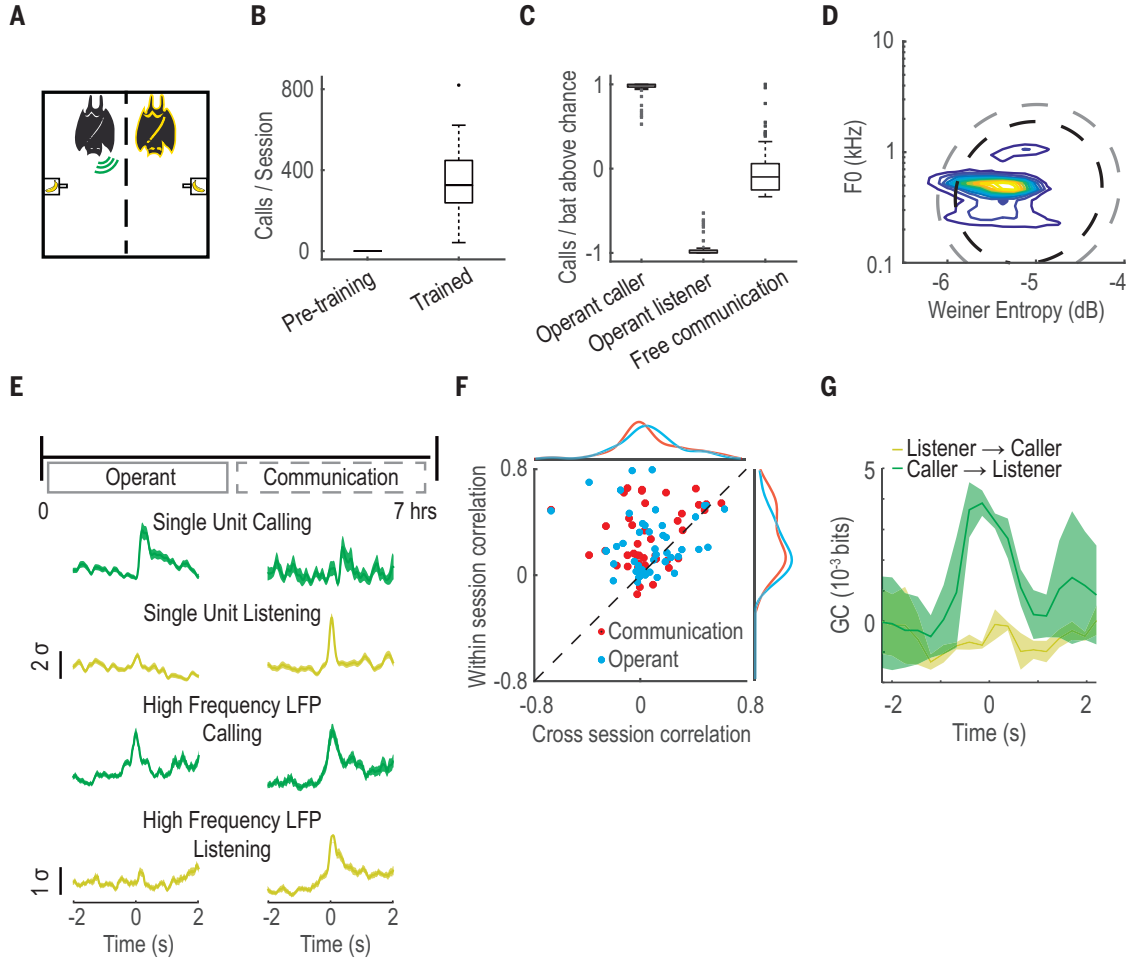


Fig. 3. Restructuring of neural activity during task-induced vocal behavior. (**A**) Operant session schematic. As above, green and yellow represent the calling and listening bat, respectively. (**B**) Number of vocalizations per session before (17 sessions, n = 4 bats) and after training (28 sessions, n = 4 bats). Box plots display median, interquartile range, and ±1.5 SD. (**C**) Average number of calls compared to chance (assuming each bat is equally likely to call), produced by each bat and normalized by the number of bats in a given session. Calling and listening bats in the operant task are separated to illustrate caller-listener dichotomy and to contrast with free communication where all bats vocalized at similar rates, on average. Data are displayed in box plots as in (B). (**D**) Contour plots representing the distribution of operant session vocalization by two acoustic features. Black and gray dashed lines correspond to areas encompassing 95% and 99% of a Gaussian distribution, respectively, fit to free communication session vocalizations. (**E**) Top: Experimental timeline over the course of a day. Operant and free communication sessions are performed in direct succession over the course of ~7 hours. Bottom: Normalized average firing rates from two example single units and normalized average high-frequency LFP power from two example tetrode channels. Shown is activity aligned to vocalizations in the operant (left) and free communication (right) sessions. Shaded areas denote SEM. (**F**) Cross- versus within-session correlation of the PETHs of individual responsive neurons, colored according to session exhibiting responsivity (n = 64 neurons). Marginal distributions of cross- and within-session correlations are also shown. Note the deviation from unity due to high within- but not cross-session correlations. (**G**) Comparison of interbrain GC values averaged across bats during the operant session, shown for both possible directional relationships. Shaded areas denote SEM.

performed a group playback session with all eight bats present to test whether acoustic stimulation in a social context would drive other-responsive or identity-selective neurons. Consistent with our previous findings, we found that during group playback, none of the other-responsive or identity-selective neurons significantly modulated their activity (table S3), nor did we observe any increased interbrain correlation (fig. S24C). Furthermore, using position tracking during the free-communication session, we found that individual bats' positions during calls could not account for the observed identity selectivity in single neurons (fig. S25). Next, we asked whether the cluster status of individual bats inferred from the social-space session was related to the neural activity during vocal interactions in the free-communication session. We focused our analysis on neural measures that intrinsically involve relationships between bats, namely the single-neuron identity selectivity and interbrain correlations.

We first examined whether the accuracy with which a calling bat could be decoded from the activity of identity-selective single neurons was related to that bat's cluster status. We found that neurons selective for in-cluster bats had 34% higher decoding accuracy, on average, than neurons selective for out-of-cluster bats (Fig. 4H; n = 68 neuron and bat pairs; P = 0.006, likelihood ratio test; fig. S26). In- and out-of-cluster bats vocalized at similar rates across sessions (n = 14 sessions, P = 0.48, paired t test) and produced calls that could not be discriminated using acoustic features (n = 15,061 calls, P = 0.15, bootstrap test), which suggests that differences in acoustic content between the groups alone did not account for the neural differences we observed.

Next, we asked whether vocalization-related interbrain correlation in the free-communication

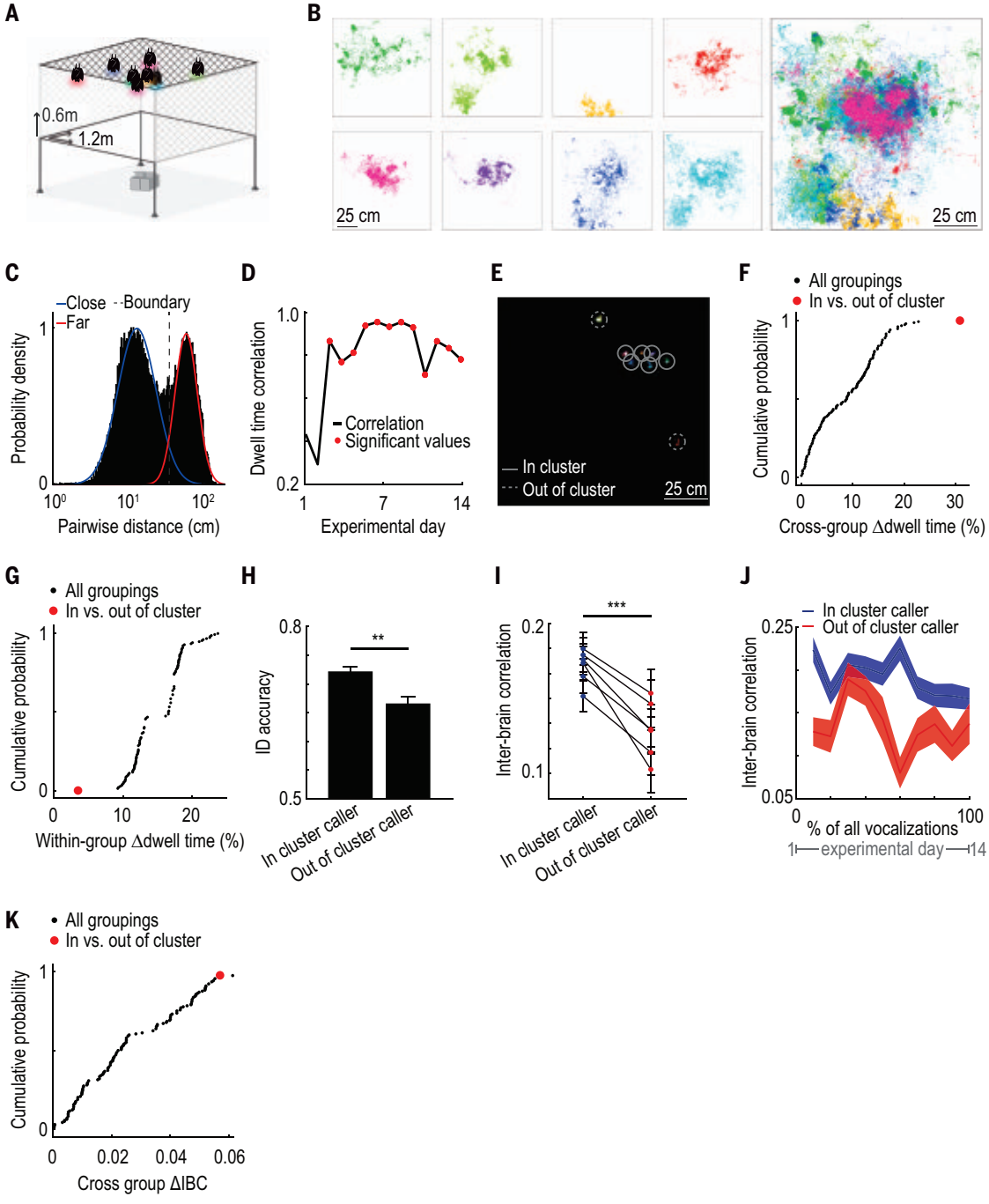


Fig. 4. Social-spatial patterns are related to both intra- and inter-brain neural activity during group communication. (**A**) Schematic of the "social space" enclosure: a large environment (1.2 × 1.2 × 0.6 m) where bats could move freely (bats and cameras not to scale). Transparent plexiglass on two sides and at the cage bottom allowed unobstructed video monitoring of all individuals (n = 8 bats). (**B**) Left: All tracked positions in an example session of each bat colored separately. Positions are shown on the xy plane corresponding to the enclosure ceiling where bats spent the majority of the time. Right: All bats' positions overlaid. (**C**) Histogram of all measured instantaneous distances between pairs of bats displayed on a logarithmic scale. Blue and red lines indicate a two-component Gaussian mixture model fit to these data. Dashed line indicates the threshold for distances that were considered "close together" or "far apart." (**D**) Correlation between dwell times for all bat pairs on a given day and the median pairwise dwell time across all other days. Statistically significant correlation values are indicated with red dots (P < 0.05, Mantel test). (**E**) Still image from the color video camera positioned below the cage. Each bat has an LED of a different color. Individual bat positions are indicated by circles. Lines are solid or dashed for in-cluster and out-of-cluster bats, respectively. (**F**) Cumulative probability distribution of the difference in mean dwell times between subgroups for all possible subgroupings of bats into two groups (n = 146 subgroupings). The in- and out-of-cluster subgrouping maximizes this value and is highlighted in red. (**G**) As in (F), showing the difference in mean dwell time within subgroups instead, which is minimized by the in- and out-of-cluster subgrouping. (**H**) Identity-decoding accuracy of identity-selective neurons when representing in-cluster versus out-of-cluster bats (**P = 0.006, likelihood ratio test; n = 68 neurons and bat pairs). Bars indicate mean decoding accuracy across neurons; error bars denote SEM. (**I**) Average interbrain correlation between bat pairs during vocalizations produced by in-cluster versus out-of-cluster bats (***$P = 1.0 \times 10^{-10}$, likelihood ratio test; n = 5758 calls and bat pairs). Each line indicates average interbrain correlation values for one bat pair; error bars indicate SEM. (**J**) Average interbrain correlation values separated as in (I) during all vocalizations across all free communication sessions, separated into equal-sized bins over time. Shaded areas denote SEM. Note that interbrain correlation values during calls from in- versus out-of-cluster bats are higher at every time point considered, indicating persistence of effect across days. (**K**) Cumulative probability of absolute differences in mean interbrain correlation between subgroups for all possible subgroupings of bats into two groups. Red dot indicates difference in interbrain correlation between the in- and out-of-cluster bat groups. Note that only one other subgrouping has a larger difference in interbrain correlation.

session was related to the cluster status of the caller. Recent studies in humans have suggested that interbrain synchrony can be modulated by differences across individuals in verbal communication (*32*, *34*, *36*) as well as social group affiliation (*37*, *38*). Consistent with these studies, we found that across all calls, interbrain correlation was 43% higher between all pairs of bats when in-cluster bats vocalized than when out-of-cluster bats vocalized (Fig. 4I; n = 5758 bat pairs and calls; $P = 1.0 \times 10^{-10}$, likelihood ratio test, controlling for bat pair and date of recording). This effect was persistent and stable over weeks and was not driven by day-to-day variability (Fig. 4J). Furthermore, differences in pairwise interbrain correlation according to the cluster status of the caller could not be accounted for by (i) the cluster status of the pair, (ii) whether the pair included the caller, or (iii) the physical distance between the bats in the free-communication session (fig. S27). Lastly, the difference in interbrain correlation during calls produced by in- versus out-of-cluster bats was greater than the difference in interbrain correlation for 98.6% of all other possible subgroupings of this group of eight bats (n = 146 subgroups; Fig. 4K), supporting the notion that social-spatial preferences influenced interbrain neural activity. Combined, these results indicate that intra- and interbrain neural activity patterns during group social communication were modulated by the social-spatial preferences of individual group members.

Discussion

In this study, we recorded neural activity from the frontal cortex of groups of bats engaged in spontaneous social communication and studied neural processes both within and across the brains of individual group members. At the level of the individual brain, we found that frontal cortical neural activity distinguished not only between calls produced by self versus others, but also between individual identities. Furthermore, examining neural activity across the brains of group members revealed robust, bidirectional, stable interbrain neural activity patterns that emerged only during spontaneous group vocal interactions.

Previous studies have demonstrated auditory responses in the frontal auditory field, an anatomically restricted region of the frontal cortex in microbats (*39*–*41*), even under anesthesia (*42*), whereas we did not observe neural responses to vocalization playback. This apparent discrepancy may be explained by our recordings being in an area outside of the frontal auditory fields, or possibly the lack of a frontal auditory field region in Egyptian fruit bats, especially considering the substantial neuroanatomical (*43*–*45*) and genomic (*46*) differences between micro- and megabats, as well as auditory functional specializations of the microbats (*47*–*49*). Future comparative studies will be crucial to further uncover the shared, as well as species-specific, neural circuits involved in vocal communicative behaviors. Yet collectively, our results are in line with findings in multiple species demonstrating the role of the frontal cortex in processing cognitive or associative aspects of acoustic information in a contextually dependent manner (*50*–*53*).

By comparing neural activity during a highly trained, task-based vocal behavior to activity during a more ethologically relevant social-vocal behavior, we found that single-neuron activity was highly contextually dependent and that interbrain correlation was entirely absent in the task-based context. This breakdown in interbrain correlation supports the hypotheses that bidirectional interbrain activity patterns are a feature of socially interactive behaviors (*54*, *55*) and that shared interbrain activity patterns may play an important role in social communication between group members (*19*, *31*, *33*, *34*, *54*). These findings also underscore the importance of incorporating components of ethologically relevant spontaneous behavior to complement the prevailing approach of less ethologically relevant, and often highly trained, experimental designs (*56*, *57*).

Consistent with the hypothesis that the repertoire of intra- and interbrain activity patterns we observed reflects (or possibly serves as the basis for) social interactions, we found that bats that chose to spend more time together elicited a more accurate representation of identity at the single-neuron level as well as higher levels of interbrain correlation between the group members. It is possible that the reduced interbrain correlation and identity accuracy elicited by out-of-cluster bats reflects differential communicative efficacy, as suggested by studies in humans (*34*, *36*). This, in turn, may relate to the abilities of individuals to negotiate a place within a group. Therefore, an important avenue for future studies will be to follow the coevolution of neural and social dynamics during the formation and restructuring of social groups of varying compositions. Similarly, future studies that perturb neural activity during social interactions will be able to dissect the circuits supporting the social behaviors described here.

In addition to bats, a wide range of species naturally interact in groups and exhibit a diversity of social structures and forms of communication. Extending the approach taken here to other species in order to uncover similarities and differences in neural repertoires for social communication may provide important insight into the neural mechanisms that facilitate successful group living.

Materials and methods summary

To study social-vocal interactions under ethologically relevant group conditions, we allowed groups of bats to interact freely in a dark enclosure and spontaneously vocalize while recording audio, video, and neural data. We used wireless tetrode-based electrophysiology to record single-neuron activity and local field potentials from the frontal cortices of multiple Egyptian fruit bats (*R. aegyptiacus*) simultaneously (these experiments were repeated in three separate groups of size n = 4, 5, and 8). The identity of the vocalizing bat was determined using a custom-made on-animal call detector device. For certain subsets of bats, we also performed the additional following experiments: (i) playback of vocalizations from a loudspeaker to either individual isolated bats or bats in the group setting; (ii) production of echolocation signals by an individual bat while freely flying; (iii) pairs of bats engaged in an operant vocalization production task; and (iv) tracking the position of individual bats in a larger enclosure to determine social-spatial preferences of members in the group. Single-neuron firing rates were calculated and used to determine which neurons modulated their firing rates while producing a vocalization (self-responsive) or listening to other bats' vocalizations (other-responsive). Single-neuron firing rates were also used to determine whether responsive neurons encoded the acoustic features of vocalizations. Identity selectivity of single neurons was assessed by training classifiers to use firing rates around vocalizations in order to predict the identity of the vocalizer. The power in the high-frequency band (70 to 150 Hz) of the local field potential was calculated and used to determine the level of interbrain correlation and Granger causality between the brains of individual bats around vocalizations. Responsivity of single neurons, identity selectivity of single neurons, and interbrain correlation measures were compared between spontaneous and trained vocalizations. Elicited identity selectivity accuracy and interbrain correlation measures were also compared between vocalizations produced by individual bats that exhibited one of two different modes of social-spatial preferences: (i) those preferring to position themselves close to other group members (in-cluster) or (ii) those preferring to remain farther away from other group members (out-of-cluster). Further details of the methods are provided in the supplementary materials.

REFERENCES AND NOTES

1. R. M. Seyfarth, D. L. Cheney, P. Marler, Vervet monkey alarm calls: Semantic communication in a free-ranging primate. *Anim. Behav.* **28**, 1070–1094 (1980). doi: 10.1016/S0003-3472(80)80097-2
2. J. E. Elie, F. E. Theunissen, The vocal repertoire of the domesticated zebra finch: A data-driven approach to decipher the information-bearing acoustic features of communication signals. *Anim. Cogn.* **19**, 285–315 (2016). doi: 10.1007/s10071-015-0933-6; pmid: 26581377
3. Y. Prat, M. Taub, Y. Yovel, Everyday bat vocalizations contain information about emitter, addressee, context, and behavior. *Sci. Rep.* **6**, 39419 (2016). doi: 10.1038/srep39419; pmid: 28005079

4. S. R. Hage, A. Nieder, Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. *Nat. Commun.* **4**, 2409 (2013). doi: 10.1038/ncomms3409; pmid: 24008252
5. S. D. Washington, J. S. Kanwal, DSCF neurons within the primary auditory cortex of the mustached bat process frequency modulations present within social calls. *J. Neurophysiol.* **100**, 3285–3304 (2008). doi: 10.1152/jn.90442.2008; pmid: 18768643
6. I. M. Carruthers, R. G. Natan, M. N. Geffen, Encoding of ultrasonic vocalizations in the auditory cortex. *J. Neurophysiol.* **109**, 1912–1927 (2013). doi: 10.1152/jn.00483.2012; pmid: 23324323
7. D. E. Okobi Jr.., A. Banerjee, A. M. M. Matheson, S. M. Phelps, M. A. Long, Motor cortical control of vocal interaction in neotropical singing mice. *Science* **363**, 983–988 (2019). doi: 10.1126/science.aau9480; pmid: 30819963
8. N. M. Adreani, P. B. D'Amelio, M. Gahr, A. Ter Maat, Life-Stage Dependent Plasticity in the Auditory System of a Songbird Is Signal and Emitter-Specific. *Front. Neurosci.* **14**, 588672 (2020). doi: 10.3389/fnins.2020.588672; pmid: 33343284
9. S. Hoffmann *et al*., Duets recorded in the wild reveal that interindividually coordinated motor control enables cooperative behavior. *Nat. Commun.* **10**, 2577 (2019). doi: 10.1038/s41467-019-10593-3; pmid: 31189912
10. C. Fichtel, M. Manser, in *Animal Behaviour: Evolution and Mechanisms* (Springer, 2010), pp. 29–54.
11. G. Kerth, Causes and Consequences of Sociality in Bats. *Bioscience* **58**, 737–746 (2008). doi: 10.1641/B580810
12. G. G. Kwiecinski, T. A. Griffiths, Rousettus egyptiacus. *Mamm. Species* **95**, 1–9 (1999). doi: 10.2307/3504411
13. L. Harten *et al*., Persistent producer-scrounger relationships in bats. *Sci. Adv.* **4**, e1603293 (2018). doi: 10.1126/sciadv.1603293; pmid: 29441356
14. L. Harten, Y. Prat, S. Ben Cohen, R. Dor, Y. Yovel, Food for Sex in Bats Revealed as Producer Males Reproduce with Scrounging Females. *Curr. Biol.* **29**, 1895–1900.e3 (2019). doi: 10.1016/j.cub.2019.04.066; pmid: 31130455
15. Y. Prat, M. Taub, Y. Yovel, Vocal learning in a social mammal: Demonstrated by isolation and playback experiments in bats. *Sci. Adv.* **1**, e1500019 (2015). doi: 10.1126/sciadv.1500019; pmid: 26601149
16. D. Genzel, J. Desai, E. Paras, M. M. Yartsev, Long-term and persistent vocal plasticity in adult bats. *Nat. Commun.* **10**, 3372 (2019). doi: 10.1038/s41467-019-11350-2; pmid: 31358755
17. S. U. Nummela, V. Jovanovic, L. de la Mothe, C. T. Miller, Social Context-Dependent Activity in Marmoset Frontal Cortex Populations during Natural Conversations. *J. Neurosci.* **37**, 7036–7047 (2017). doi: 10.1523/JNEUROSCI.0702-17.2017; pmid: 28630255
18. L. M. Romanski, B. B. Averbeck, M. Diltz, Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. *J. Neurophysiol.* **93**, 734–747 (2005). doi: 10.1152/jn.00675.2004; pmid: 15371495
19. A. K. Kuhlen, C. Bogler, S. E. Brennan, J. D. Haynes, Brains in dialogue: Decoding neural preparation of speaking to a conversational partner. *Soc. Cogn. Affect. Neurosci.* **12**, 871–880 (2017). doi: 10.1093/scan/nsx018; pmid: 28338791
20. L. Kingsbury *et al*., Correlated Neural Activity and Encoding of Behavior across Brains of Socially Interacting Animals. *Cell* **178**, 429–446.e16 (2019). doi: 10.1016/j.cell.2019.05.022; pmid: 31230711
21. S. W. C. Chang, J. F. Gariépy, M. L. Platt, Neuronal reference frames for social decisions in primate frontal cortex. *Nat. Neurosci.* **16**, 243–250 (2013). doi: 10.1038/nn.3287; pmid: 23263442
22. W. Zhang, M. M. Yartsev, Correlated Neural Activity across the Brains of Socially Interacting Bats. *Cell* **178**, 413–428.e22 (2019). doi: 10.1016/j.cell.2019.05.023; pmid: 31230710
23. S. R. Hage, A. Nieder, Audio-vocal interaction in single neurons of the monkey ventrolateral prefrontal cortex. *J. Neurosci.* **35**, 7030–7040 (2015). doi: 10.1523/JNEUROSCI.2371-14.2015; pmid: 25948255
24. L. J. Silbert, C. J. Honey, E. Simony, D. Poeppel, U. Hasson, Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E4687–E4696 (2014). doi: 10.1073/pnas.1323812111; pmid: 25267658
25. C. T. Miller, A. W. Thomas, S. U. Nummela, L. A. de la Mothe, Responses of primate frontal cortex neurons during natural vocal communication. *J. Neurophysiol.* **114**, 1158–1171 (2015). doi: 10.1152/jn.01003.2014; pmid: 26084912
26. K. Weineck, F. García-Rosales, J. C. Hechavarría, Neural oscillations in the fronto-striatal network predict vocal output in bats. *PLOS Biol.* **18**, e3000658 (2020). doi: 10.1371/journal.pbio.3000658; pmid: 32191695
27. F. P. Möhres, E. Kulzer, Über die Orientierung der Flughunde (Chiroptera - Pteropodidae). *Z. Vgl. Physiol.* **38**, 1–29 (1956). doi: 10.1007/BF00338621
28. W. J. Lee *et al*., Tongue-driven sonar beam steering by a lingual-echolocating fruit bat. *PLOS Biol.* **15**, e2003148 (2017). doi: 10.1371/journal.pbio.2003148; pmid: 29244805
29. T. Sugihara, M. D. Diltz, B. B. Averbeck, L. M. Romanski, Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. *J. Neurosci.* **26**, 11138–11147 (2006). doi: 10.1523/JNEUROSCI.3550-06.2006; pmid: 17065454
30. F. Babiloni, L. Astolfi, Social neuroscience and hyperscanning techniques: Past, present and future. *Neurosci. Biobehav. Rev.* **44**, 76–93 (2014). doi: 10.1016/j.neubiorev.2012.07.006; pmid: 22917915
31. A. Pérez, M. Carreiras, J. A. Duñabeitia, Brain-to-brain entrainment: EEG interbrain synchronization while speaking and listening. *Sci. Rep.* **7**, 4190 (2017). doi: 10.1038/s41598-017-04464-4; pmid: 28646190
32. G. J. Stephens, L. J. Silbert, U. Hasson, Speaker-listener neural coupling underlies successful communication. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 14425–14430 (2010). doi: 10.1073/pnas.1008662107; pmid: 20660768
33. A. Stolk *et al*., Cerebral coherence between communicators marks the emergence of meaning. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 18183–18188 (2014). doi: 10.1073/pnas.1414886111; pmid: 25489093
34. S. Dikker *et al*., Brain-to-Brain Synchrony Tracks Real-World Dynamic Group Interactions in the Classroom. *Curr. Biol.* **27**, 1375–1380 (2017). doi: 10.1016/j.cub.2017.04.002; pmid: 28457867
35. O. Mann *et al*., Finding Your Friends at Densely Populated Roosting Places: Male Egyptian Fruit Bats (*Rousettus aegyptiacus*) Distinguish between Familiar and Unfamiliar Conspecifics. *Acta Chiropt.* **13**, 411–417 (2011). doi: 10.3161/150811011X624893
36. R. Schmälzle, F. E. K. Häcker, C. J. Honey, U. Hasson, Engaged listeners: Shared neural processing of powerful political speeches. *Soc. Cogn. Affect. Neurosci.* **10**, 1137–1143 (2015). doi: 10.1093/scan/nsu168; pmid: 25653012
37. R. Hyon *et al*., Similarity in functional brain connectivity at rest predicts interpersonal closeness in the social network of an entire village. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 33149–33160 (2020). doi: 10.1073/pnas.2013606117; pmid: 33318188
38. J. Yang, H. Zhang, J. Ni, C. K. W. De Dreu, Y. Ma, Within-group synchronization in the prefrontal cortex associates with intergroup conflict. *Nat. Neurosci.* **23**, 754–760 (2020). doi: 10.1038/s41593-020-0630-x; pmid: 32341541
39. J. B. Kobler, S. F. Isbey, J. H. Casseday, Auditory pathways to the frontal cortex of the mustache bat, Pteronotus parnellii. *Science* **236**, 824–826 (1987). doi: 10.1126/science.2437655; pmid: 2437655
40. J. S. Kanwal, M. Gordon, J. P. Peng, K. Heinz-Esser, Auditory responses from the frontal cortex in the mustached bat, Pteronotus parnellii. *Neuroreport* **11**, 367–372 (2000). doi: 10.1097/00001756-200002070-00029; pmid: 10674488
41. A. Eiermann, K. H. Esser, Auditory responses from the frontal cortex in the short-tailed fruit bat Carollia perspicillata. *Neuroreport* **11**, 421–425 (2000). doi: 10.1097/00001756-200002070-00040; pmid: 10674499
42. L. López-Jury, A. Mannel, F. García-Rosales, J. C. Hechavarria, Modified synaptic dynamics predict neural activity patterns in an auditory field within the frontal cortex. *Eur. J. Neurosci.* **51**, 1011–1025 (2020). doi: 10.1111/ejn.14600; pmid: 31630441
43. J. D. Pettigrew, B. C. Maseko, P. R. Manger, Primate-like retinotectal decussation in an echolocating megabat, Rousettus aegyptiacus. *Neuroscience* **153**, 226–231 (2008). doi: 10.1016/j.neuroscience.2008.02.019; pmid: 18367343
44. A. Thiele, R. Rübsamen, K. P. Hoffmann, Anatomical and physiological investigation of auditory input to the superior colliculus of the echolocating megachiropteran bat Rousettus aegyptiacus. *Exp. Brain Res.* **112**, 223–236 (1996). doi: 10.1007/BF00227641; pmid: 8951391
45. K. Hu, Y. Li, X. Gu, H. Lei, S. Zhang, Brain structures of echolocating and nonecholocating bats, derived in vivo from magnetic resonance images. *Neuroreport* **17**, 1743–1746 (2006). doi: 10.1097/01.wnr.0000239959.91190.c8; pmid: 17047465
46. M. Nikaido *et al*., Comparative genomic analyses illuminate the distinct evolution of megabats within Chiroptera. *DNA Res.* **27**, dsaa021 (2020). doi: 10.1093/dnares/dsaa021; pmid: 32966557
47. E. Covey, Neurobiological specializations in echolocating bats. *Anat. Rec. A* **287**, 1103–1116 (2005). doi: 10.1002/ar.a.20254; pmid: 16217809
48. T. Ito *et al*., Three forebrain structures directly inform the auditory midbrain of echolocating bats. *Neurosci. Lett.* **712**, 134481 (2019). doi: 10.1016/j.neulet.2019.134481; pmid: 31494222
49. M. Kössl, J. Hechavarria, C. Voss, M. Schaefer, M. Vater, Bat auditory cortex – model for general mammalian auditory computation or special design solution for active time perception? *Eur. J. Neurosci.* **41**, 518–532 (2015). doi: 10.1111/ejn.12801; pmid: 25728173
50. M. Medalla, H. Barbas, Specialized prefrontal "auditory fields": Organization of primate prefrontal-temporal pathways. *Front. Neurosci.* **8**, 77 (2014). doi: 10.3389/fnins.2014.00077; pmid: 24795553
51. B. Plakke, L. M. Romanski, Auditory connections and functions of prefrontal cortex. *Front. Neurosci.* **8**, 199 (2014). doi: 10.3389/fnins.2014.00199; pmid: 25100931
52. J. B. Fritz, S. V. David, S. Radtke-Schuller, P. Yin, S. A. Shamma, Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. *Nat. Neurosci.* **13**, 1011–1019 (2010). doi: 10.1038/nn.2598; pmid: 20622871
53. Y. E. Cohen, F. Theunissen, B. E. Russ, P. Gill, Acoustic features of rhesus vocalizations and their representation in the ventrolateral prefrontal cortex. *J. Neurophysiol.* **97**, 1470–1484 (2007). doi: 10.1152/jn.00769.2006; pmid: 17135477
54. U. Hasson, A. A. Ghazanfar, B. Galantucci, S. Garrod, C. Keysers, Brain-to-brain coupling: A mechanism for creating and sharing a social world. *Trends Cogn. Sci.* **16**, 114–121 (2012). doi: 10.1016/j.tics.2011.12.007; pmid: 22221820
55. P. Barraza, A. Pérez, E. Rodríguez, Brain-to-Brain Coupling in the Gamma-Band as a Marker of Shared Intentionality. *Front. Hum. Neurosci.* **14**, 295 (2020). doi: 10.3389/fnhum.2020.00295; pmid: 32848670
56. J. W. Krakauer, A. A. Ghazanfar, A. Gomez-Marin, M. A. MacIver, D. Poeppel, Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron* **93**, 480–490 (2017). doi: 10.1016/j.neuron.2016.12.041; pmid: 28182904
57. M. M. Yartsev, The emperor's new wardrobe: Rebalancing diversity of animal models in neuroscience research. *Science* **358**, 466–469 (2017). doi: 10.1126/science.aan8865; pmid: 29074765
58. M. C. Rose, B. Styr, T. A. Schmid, J. E. Elie, M. M. Yartsev, Zenodo (2021). doi: 10.5281/zenodo.5152173
59. G. Neuweiler, Bau und Leistung des Flughundauges (Pteropus giganteus gig. Brünn.). *Z. Vgl. Physiol.* **46**, 13–56 (1962). doi: 10.1007/BF00340352
60. H. Q. Liu *et al*., *Sci. Rep.* **5**, 1–13 (2015).
61. E. Z. Lattenkamp, S. C. Vernes, L. Wiegrebe, Volitional control of social vocalisations and vocal usage learning in bats. *J. Exp. Biol.* **221**, jeb180729 (2018). doi: 10.1242/jeb.180729; pmid: 29880634
62. M. M. Yartsev, N. Ulanovsky, Representation of three-dimensional space in the hippocampus of flying bats. *Science* **340**, 367–372 (2013). doi: 10.1126/science.1235338; pmid: 23599496
63. A. Finkelstein *et al*., Three-dimensional head-direction coding in the bat brain. *Nature* **517**, 159–164 (2015). doi: 10.1038/nature14031; pmid: 25470055
64. M. Fukushima, D. Margoliash, The effects of delayed auditory feedback revealed by bone conduction microphone in adult zebra finches. *Sci. Rep.* **5**, 8800 (2015). doi: 10.1038/srep08800; pmid: 25739659
65. A. de Cheveigné, H. Kawahara, YIN, a fundamental frequency estimator for speech and music. *J. Acoust. Soc. Am.* **111**, 1917–1930 (2002). doi: 10.1121/1.1458024; pmid: 12002874
66. E. Glerean *et al*., Reorganization of functionally connected brain subnetworks in high-functioning autism. *Hum. Brain Mapp.* **37**, 1066–1079 (2016). doi: 10.1002/hbm.23084; pmid: 26686668
67. R. Q. Quiroga, Z. Nadasdy, Y. Ben-Shaul, Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Comput.* **16**, 1661–1687 (2004). doi: 10.1162/089976604774201631; pmid: 15228749
68. E. S. Page, Continuous inspection schemes. *Biometrika* **41**, 100–115 (1954). doi: 10.1093/biomet/41.1-2.100
69. L. Koepcke, G. Ashida, J. Kretzberg, Single and Multiple Change Point Detection in Spike Trains: Comparison of Different CUSUM Methods. *Front. Syst. Neurosci.* **10**, 51 (2016). doi: 10.3389/fnsys.2016.00051; pmid: 27445714

70. A. O. Nunez-Elizalde, A. G. Huth, J. L. Gallant, Voxelwise encoding models with non-spherical multivariate normal priors. *Neuroimage* **197**, 482–492 (2019). doi: 10.1016/j.neuroimage.2019.04.012; pmid: 31075394
71. S. Musall, M. T. Kaufman, A. L. Juavinett, S. Gluf, A. K. Churchland, Single-trial neural dynamics are dominated by richly varied movements. *Nat. Neurosci.* **22**, 1677–1686 (2019). doi: 10.1038/s41593-019-0502-4; pmid: 31551604
72. L. Barnett, A. K. Seth, The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference. *J. Neurosci. Methods* **223**, 50–68 (2014). doi: 10.1016/j.jneumeth.2013.10.018; pmid: 24200508
73. L. Barnett, A. B. Barrett, A. K. Seth, Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys. Rev. Lett.* **103**, 238701 (2009). doi: 10.1103/PhysRevLett.103.238701; pmid: 20366183
74. N. M. Timme, C. Lapish, A Tutorial for Information Theory in Neuroscience. *eNeuro* **5**, 1–40 (2018). doi: 10.1523/ENEURO.0052-18.2018; pmid: 30211307

ACKNOWLEDGMENTS

We thank W. Zhang, W. A. Liberti, E. Azim, R. T. Knight, Z. M. Williams, and R. Báez-Mendoza for comments on the manuscript; J. W. Pillow, F. E. Theunissen, and S. Popham for advice on acoustic encoding modelling; W. Zhang and R. T. Knight for advice on interbrain analyses; D. Genzel for advice on acoustic recording and analysis; L. Loomis and J. Chase for animal training; A. Rakuljic and B. Shvareva for assistance with histology; and G. Lawson, K. Jensen, J. Frohlich and the staff of the Office of Laboratory Animal Care for support with animal husbandry and care. **Funding:** Supported by NIH (DP2-DC016163), NIMH (1-R01MH25387-01), the New York Stem Cell Foundation (NYSCF-R-NI40), the Alfred P. Sloan Foundation (FG-2017-9646), the Brain Research Foundation (BRFSG-2017-09), a Packard Fellowship (2017-66825), the Klingenstein Simons Fellowship, the Human Frontiers Science Program, the Pew Charitable Trust (00029645), the McKnight Foundation, the Dana Foundation (M.M.Y.), and a Human Frontiers Science Program postdoctoral fellowship (B.S.). M.M.Y. is a New York Stem Cell Foundation–Robertson Investigator. **Author contributions:** M.C.R. and M.M.Y. designed the research, analyzed the data and wrote the original draft of the paper. M.C.R. performed experiments. M.C.R. and B.S. designed the research, conducted the experiments of the group social-spatial behavior, and together with M.M.Y. analyzed the data and revised the manuscript. T.A.S. and J.E.E. designed and developed the operant task. M.C.R. and J.E.E. implemented the multi-animal acoustic and neurophysiological recording methods. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The data presented in this study are available at (*58*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.aba9584
Materials and Methods
Figs. S1 to S27
Tables S1 to S4
References (*59*–*74*)

21 February 2020; resubmitted 15 February 2021
Accepted 7 September 2021
10.1126/science.aba9584

Your Legacy to Science

AN ESTATE GIFT TO THE
AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

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

By making AAAS a beneficiary of your will, trust, retirement plan or life insurance policy, you become a member of our 1848 Society, joining Thomas Edison, Alexander Graham Bell and the many distinguished individuals whose vision led to the creation of AAAS and our world-renowned journal, *Science*, so many years ago.

Unlike many of its peers, *Science* is not for-profit. Your estate gift would provide long-term financial stability and durable annual income that will support operations and competitive innovation for years to come. **This support is vital.**

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

—Prof. Elisabeth Ervin-Blankenheim, 1848 Society member

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

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

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

1848 SOCIETY AAAS

cut here

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

PLEASE CONTACT ME AT:

Name: _____

Address: _____

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

Email: _____ Phone: _____

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

RESEARCH ARTICLES

CORONAVIRUS

A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa

Eduan Wilkinson[1,2]†, Marta Giovanetti[3,4]†, Houriiyah Tegally[1]†, James E. San[1]†, Richard Lessells[1], Diego Cuadros[5], Darren P. Martin[6,7], David A. Rasmussen[8,9], Abdel-Rahman N. Zekri[10], Abdoul K. Sangare[11], Abdoul-Salam Ouedraogo[12], Abdul K. Sesay[13], Abechi Priscilla[14], Adedotun-Sulaiman Kemi[14], Adewunmi M. Olubusuyi[15], Adeyemi O. O. Oluwapelumi[16], Adnène Hammami[17], Adrienne A. Amuri[18,19], Ahmad Sayed[20], Ahmed E. O. Ouma[21], Aida Elargoubi[22,23], Nnennaya A. Ajayi[24], Ajogbasile F. Victoria[14], Akano Kazeem[14], Akpede George[25], Alexander J. Trotter[26], Ali A. Yahaya[27], Alpha K. Keita[28,29], Amadou Diallo[30], Amadou Kone[31], Amal Souissi[32], Amel Chtourou[17], Ana V. Gutierrez[26], Andrew J. Page[26], Anika Vinze[33], Arash Iranzadeh[6,7], Arnold Lambisia[34], Arshad Ismail[35], Audu Rosemary[36], Augustina Sylverken[37], Ayoade Femi[14], Azeddine Ibrahimi[38], Baba Marycelin[39], Bamidele S. Oderinde[39], Bankole Bolajoko[14], Beatrice Dhaala[40], Belinda L. Herring[27], Berthe-Marie Njanpop-Lafourcade[27], Bronwyn Kleinhans[41], Bronwyn McInnis[10], Bryan Tegomoh[42], Cara Brook[43,44], Catherine B. Pratt[45], Cathrine Scheepers[35,46], Chantal G. Akoua-Koffi[47], Charles N. Agoti[34,48], Christophe Peyrefitte[30], Claudia Daubenberger[49], Collins M. Morang'a[50], D. James Nokes[34,51], Daniel G. Amoako[35], Daniel L. Bugembe[40], Danny Park[33], David Baker[26], Deelan Doolabh[7], Deogratius Ssemwanga[40,52], Derek Tshiabuila[1], Diarra Bassirou[30], Dominic S. Y. Amuzu[50], Dominique Goedhals[53], Donwilliams O. Omuoyo[34], Dorcas Maruapula[54], Ebenezer Foster-Nyarko[26], Eddy K. Lusamaki[18,19], Edgar Simulundu[55], Edidah M. Ong'era[34], Edith N. Ngabana[18,19], Edwin Shumba[56], Elmostafa El Fahime[57], Emmanuel Lokilo[18], Enatha Mukantwari[58], Eromon Philomena[14], Essia Belarbi[59], Etienne Simon-Loriere[60], Etilé A. Anoh[47], Fabian Leendertz[59], Faida Ajili[61], Fakayode O. Enoch[62], Fares Wasfi[63], Fatma Abdelmoula[32,64], Fausta S. Mosha[27], Faustinos T. Takawira[65], Fawzi Derrar[66], Feriel Bouzid[32], Folarin Onikepe[14], Fowotade Adeola[67], Francisca M. Muyembe[18,19], Frank Tanser[68,69,70], Fred A. Dratibi[27], Gabriel K. Mbunsu[19], Gaetan Thilliez[26], Gemma L. Kay[26], George Githinji[34,71], Gert van Zyl[41,72], Gordon A. Awandare[50], Grit Schubert[59], Gugu P. Maphalala[73], Hafaliana C. Ranaivoson[44], Hajar Lemriss[74], Happi Anise[14], Haruka Abe[75], Hela H. Karray[17], Hellen Nansumba[76], Hesham A. Elgahzaly[77], Hlanai Gumbo[65], Ibtihel Smeti[32], Ikhlas B. Ayed[32], Ikponmwosa Odia[25], Ilhem Boutiba Ben Boubaker[78,79], Imed Gaaloul[22], Inbal Gazy[80], Innocent Mudau[7], Isaac Ssewanyana[76], Iyaloo Konstantinus[81], Jean B. Lekana-Douk[82], Jean-Claude C. Makangara[18,19], Jean-Jacques M. Tamfum[18,19], Jean-Michel Heraud[30,44], Jeffrey G. Shaffer[83], Jennifer Giandhari[1], Jingjing Li[84], Jiro Yasuda[75], Joana Q. Mends[85], Jocelyn Kiconco[52], John M. Morobe[34], John O. Gyapong[85], Johnson C. Okolie[14], John T. Kayiwa[40], Johnathan A. Edwards[68,86], Jones Gyamfi[85], Jouali Farah[87], Joweria Nakaseegu[52], Joyce M. Ngoi[50], Joyce Namulondo[52], Julia C. Andeko[82], Julius J. Lutwama[40], Justin O'Grady[26], Katherine Siddle[33], Kayode T. Adeyemi[14], Kefentse A. Tumedi[88], Khadija M. Said[34], Kim Hae-Young[89], Kwabena O. Duedu[85], Lahcen Belyamani[38], Lamia Fki-Berrajah[17], Lavanya Singh[1], Leonardo de O. Martins[26], Lynn Tyers[7], Magalutcheemee Ramuth[91], Maha Mastouri[22,23], Mahjoub Aouni[22], Mahmoud el Hefnawi[92], Maitshwarelo I. Matsheka[88], Malebogo Kebabonye[93], Mamadou Diop[30], Manel Turki[32], Marietou Paye[33], Martin M. Nyaga[94], Mathabo Mareka[95], Matoke-Muhia Damaris[96], Maureen W. Mburu[34], Maximillian Mpina[49,97,98], Mba Nwando[99], Michael Owusu[100], Michael R. Wiley[45], Mirabeau T. Youtchou[101], Mitoha O. Ayekaba[97], Mohamed Abouelhoda[102,103], Mohamed G. Seadawy[104], Mohamed K. Khalifa[20], Mooko Sekhele[95], Mouna Ouadghiri[38], Moussa M. Diagne[30], Mulenga Mwenda[105], Mushal Allam[35], My V. T. Phan[40], Nabil Abid[79,106], Nadia Touil[107], Nadine Rujeni[108,109], Najla Kharrat[32], Nalia Ismael[110], Ndongo Dia[30], Nedio Mabunda[110], Nei-yuan Hsiao[7,111], Nelson B. Silochi[97], Ngoy Nsenga[27], Nicksy Gumede[27], Nicola Mulder[112], Nnaemeka Ndodo[99], Norosoa H Razanajatovo[44], Nosamiefan Iguosadolo[14], Oguzie Judith[14], Ojide C. Kingsley[113], Okogbenin Sylvanus[25], Okokhere Peter[25], Oladiji Femi[114], Olawoye Idowu[14], Olumade Testimony[14], Omoruyi E. Chukwuma[67], Onwe E. Ogah[115], Chika K. Onwuamah[36,138], Oshomah Cyril[25], Ousmane Faye[30], Oyewale Tomori[14], Pascale Ondoa[56], Patrice Combe[116], Patrick Semanda[76], Paul E. Oluniyi[14], Paulo Arnaldo[110], Peter K. Quashie[50], Philippe Dussart[44], Phillip A. Bester[53], Placide K. Mbala[18,19], Reuben Ayivor-Djanie[85], Richard Njouom[117], Richard O. Phillips[118], Richmond Gorman[118], Robert A. Kingsley[26], Rosina A. A. Carr[85], Saâd El Kabbaj[119], Saba Gargouri[17], Saber Masmoudi[32], Safietou Sankhe[30], Salako B. Lawal[36], Samar Kassim[77], Sameh Trabelsi[120], Samar Metha[33], Sami Kammoun[121], Sanaâ Lemriss[122], Sara H. A. Agwa[77], Sébastien Calvignac-Spencer[59], Stephen F. Schaffner[33], Seydou Doumbia[31], Sheila M. Mandanda[18,19], Sherihane Aryeetey[123], Shymaa S. Ahmed[123], Siham Elhamoumi[33], Soafy Andriamandimby[44], Sobajo Tope[14], Sonia Lekana-Douki[82], Sophie Prosolek[26], Soumeya Ouangraoua[124,125], Steve A. Mundeke[18,19], Steven Rudder[26], Sumir Panji[112], Sureshnee Pillay[1], Susan Engelbrecht[41,72], Susan Nabadda[76], Sylvie Behillil[126], Sylvie L. Budiaki[95], Sylvie van der Werf[126], Tapfumanei Mashe[65], Tarik Aanniz[38], Thabo Mohale[35], Thanh Le-Viet[26], Tobias Schindler[49,97], Ugochukwu J. Anyaneji[1], Ugwu Chinedu[14], Upasana Ramphal[1,69,127], Uwanibe Jessica[14], Uwem George[14], Vagner Fonseca[1,4,128], Vincent Enouf[126], Vivianne Gorova[129,130], Wael H. Roshdy[123], William K. Ampofo[50], Wolfgang Preiser[41,72], Wonderful T. Choga[54,131], Yaw Bediako[50], Yeshnee Naidoo[1], Yvan Butera[108,132,133], Zaydah R. de Laurent[34], Amadou A. Sall[30], Ahmed Rebai[32], Anne von Gottberg[35,139], Bourema Kouriba[12], Carolyn Williamson[7,69,111], Daniel J. Bridges[105], Ihekweazu Chikwe[99], Jinal N. Bhiman[35,139], Madisa Mine[134], Matthew Cotten[40,135], Sikhulile Moyo[54,136], Simani Gaseitsiwe[54,136], Ngonda Saasa[55], Pardis C. Sabeti[33], Pontiano Kaleebu[40], Yenew K. Tebeje[21], Sofonias K. Tessema[21], Christian Happi[14], John Nkengasong[21], Tulio de Oliveira[1,2,69,137]*

The progression of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic in Africa has so far been heterogeneous, and the full impact is not yet well understood. In this study, we describe the genomic epidemiology using a dataset of 8746 genomes from 33 African countries and two overseas territories. We show that the epidemics in most countries were initiated by importations predominantly from Europe, which diminished after the early introduction of international travel restrictions. As the pandemic progressed, ongoing transmission in many countries and increasing mobility led to the emergence and spread within the continent of many variants of concern and interest, such as B.1.351, B.1.525, A.23.1, and C.1.1. Although distorted by low sampling numbers and blind spots, the findings highlight that Africa must not be left behind in the global pandemic response, otherwise it could become a source for new variants.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) emerged in late 2019 in Wuhan, China (*1*, *2*). Since then, the virus has spread to all corners of the world, causing almost 150 million cases of COVID-19 and more than 3 million deaths by the end of April 2021. Throughout the pandemic, it has been noted that Africa accounts for a relatively low proportion of reported cases and deaths—by the end of April 2021, there had been ~4.5 million cases and ~120,000 deaths on the continent, corresponding to less than 4% of the global burden. However, emerging data from seroprevalence surveys and autopsy studies in some African countries suggest that the true number of infections and deaths may be severalfold higher than reported (*3*, *4*). In addition, a recent analysis has shown that in

many African countries, the second wave of the pandemic was more severe than the first wave (*5*).

The first cases of COVID-19 on the African continent were reported in Nigeria, Egypt, and South Africa between mid-February and early March 2020, and most countries had reported cases by the end of March 2020 (*6*–*8*). These early cases were concentrated among airline travelers returning from regions of the world with high levels of community transmission. Many African countries introduced early public health and social measures, including international travel controls, quarantine for returning travelers, and internal lockdown measures, to limit the spread of the virus and give health services time to prepare (*5*, *9*). The initial phase of the epidemic was then heterogeneous, with relatively high case numbers reported in North Africa and southern Africa, and fewer cases reported in other regions.

From the onset of the pandemic, genomic surveillance has been at the forefront of the COVID-19 response in Africa (*10*). Rapid implementation of SARS-CoV-2 sequencing by various laboratories in Africa enabled genomic data to be generated and shared from the early imported cases. In Nigeria, the first genome sequence was released just 3 days after the announcement of the first case (*6*). Similarly, in Uganda, a sequencing program was set up rapidly to facilitate virus tracing, and the collection of samples for sequencing began immediately upon confirmation of the first case (*11*). In South Africa, the Network for Genomic Surveillance in South Africa (NGS-SA) was established in March 2020, and within weeks, genomic analysis was helping to characterize outbreaks and community transmission (*12*).

Genomic surveillance has also been critical for monitoring ongoing SARS-CoV-2 evolution and detection of new SARS-CoV-2 variants in Africa. Intensified sampling by NGS-SA in the Eastern Cape Province of South Africa in November 2020, in response to a rapid resurgence of cases, led to the detection of B.1.351 (501Y.V2) (*13*). This variant was subsequently designated a variant of concern (VOC) by the World Health Organization (WHO), owing to evidence of increased transmissibility (*14*) and resistance to neutralizing antibodies elicited by natural infection and vaccines (*15*–*17*).

In this study, we performed phylogenetic and phylogeographic analyses of SARS-CoV-2 genomic data from 33 African countries and two overseas territories to help characterize the dynamics of the pandemic in Africa. We show that the early introductions were predominantly from Europe, but that as the pandemic progressed, there was increasing spread between African countries. We also describe the emergence and spread of a number of key SARS-CoV-2 variants in Africa and highlight how the spread of B.1.351 (501Y.V2) and other variants contributed to the more severe second wave of the pandemic in many countries.

All author affiliations are listed at the end of this paper.
*Corresponding author. Email: tulio@sun.ac.za
†These authors contributed equally to this work.

SARS-CoV-2 genomic data

By 5 May 2021, 14,504 SARS-CoV-2 genomes had been submitted to the GISAID database (*18*) from 38 African countries and two overseas territories (Mayotte and Réunion) (Fig. 1A). Overall, this corresponds to approximately one sequence per ~300 reported cases. Almost half of the sequences were from South Africa (n = 5362), consistent with it being responsible for almost half of the reported cases in Africa. Overall, the number of sequences correlates closely with the number of reported cases per country (Fig. 1B). The countries and territories with the highest coverage of sequencing (defined as genomes per reported case) are Kenya (n = 856, one sequence per ~203 cases), Mayotte (n = 721, one sequence per ~21 cases), and Nigeria (n = 660, one sequence per ~250 cases). Although genomic surveillance started early in many countries, few have evidence of consistent sampling across the whole year. Half of all African genomes were deposited in the first 10 weeks of 2021, suggesting intensified surveillance in the second wave after the detection of B.1.351 (501Y.V2) and other variants (Fig. 1, C and D).

Genetic diversity and lineage dynamics in Africa

Of the 10,326 genomes retrieved from GISAID by the end of March 2021, 8746 genomes passed quality control and met the minimum metadata requirements. These genomes from Africa were compared in a phylogenetic framework with 11,891 representative genomes from around the world. Ancestral location state reconstruction of the dated phylogeny (hereafter referred to as discrete phylogeographic reconstruction) allowed us to infer the number of viral imports and exports between Africa and the rest of the world, and between individual African countries. African genomes in this study spanned the whole global genetic diversity of SARS-CoV-2, a pattern that largely reflects multiple introductions over time from the rest of the world (Fig. 2A).

In total, we detected at least 757 [95% confidence interval (CI): 728 to 786] viral introductions into African countries between the start of 2020 and February 2021, more than half of which occurred before the end of May 2020. Although the early phase of the pandemic was dominated by importations from outside Africa, predominantly from Europe, there was then a shift in the dynamics, with an increasing number of importations from other African countries as the pandemic progressed (Fig. 2, B and C). A rarefaction analysis in which we systematically subsampled genomes shows that vastly more introductions would have likely been identified with increased sampling in Africa or globally, suggesting that the introductions we identified are really just the "ears of the hippo," or a small part of a larger problem (fig. S1).

South Africa, Kenya, and Nigeria appear as major sources of importations into other African countries (Fig. 2D), although this is likely to be influenced by these three countries having the greatest number of deposited sequences. Particularly notable is the southern African region, where South Africa is the source for a large proportion (~80%) of the importations to other countries in the region. The North African region demonstrates a different pattern to the rest of the continent, with more viral introductions from Europe and Asia (particularly the Middle East) than from other African countries (fig. S2).

Africa has also contributed to the international spread of the virus, with at least 324 (95% CI: 295 to 353) exportation events from Africa to the rest of the world detected in this dataset. Consistent with the source of importations, most exports were to Europe (41%), Asia (26%), and North America (14%). As with the number of importations, exports were relatively evenly distributed over the 1-year period (fig. S3). However, an increase in the number of exportation events occurred between December 2020 and March 2021, which coincided with the second wave of infections in Africa and with some relaxations of travel restrictions around the world.

The early phase of the pandemic was characterized by the predominance of lineage B.1. This was introduced multiple times to African countries and has been detected in all but one of the countries included in this analysis. After its emergence in South Africa, B.1.351 became the most frequently detected SARS-CoV-2 lineage found in Africa (n = 1769, ~20%) (Fig. 1C). It was first sampled on 8 October 2020 in South Africa (*13*) and has since spread to 20 other African countries.

As air travel came to an almost complete halt in March and April 2020, the number(s) of detectable viral imports into Africa decreased and the pandemic entered a phase that was characterized in sub-Saharan Africa by sustained low levels of within-country movements and occasional international viral movements between neighboring countries, presumably via road and rail links between these. Though some border posts between countries were closed during the initial lockdown period (table S1), others remained open to allow trade to continue. Regional trade in southern Africa was only slightly affected by lockdown restrictions and quickly rebounded to prepandemic levels (fig. S4) after the relaxation of restrictions between June 2020 and December 2020.

Although lineage A viruses were imported into several African countries, they only account for 1.3% of genomes sampled in Africa. Despite lineage A viruses initially causing many localized clustered outbreaks, each the result of independent introductions to several countries (e.g., Burkina Faso, Côte d'Ivoire, and Nigeria), they were later largely replaced by lineage B viruses as the pandemic evolved. This is possibly due to the increased transmissibility of lineage B viruses by virtue of the D614G (Asp^{614}→Gly) mutation in the spike protein (*19*, *20*). However, there is evidence of an increasing prevalence of lineage A viruses in some African countries (*11*). In particular, A.23.1 emerged in East Africa and appears to be rapidly increasing in prevalence in Uganda and Rwanda (*11*). Furthermore, a highly divergent variant from lineage A was recently identified in Angola from individuals arriving from Tanzania (*21*).

Emergence and spread of new SARS-CoV-2 variants

To determine how some of the key SARS-CoV-2 variants are spreading within Africa, we performed phylogeographic analyses on the VOC B.1.351, the variant of interest (VOI) B.1.525, and two additional variants that emerged and that we designated as VOIs for this analysis (A.23.1 and C.1.1). These African VOCs and VOIs have multiple mutations on the spike glycoprotein, and a molecular clock analysis of these four datasets provided strong evidence that these four lineages are evolving in a clock-like manner (Fig. 3, A and B).

B.1.351 was first sampled in South Africa in October 2020, but phylogeographic analysis suggests that it emerged earlier, around August 2020. It is defined by 10 mutations in the spike protein, including K417N (Lys^{417}→Asn), E484K (Glu^{484}→Lys), and N501Y (Asn^{501}→Tyr) in the receptor binding domain (Fig. 3B). After its emergence in the Eastern Cape, it spread extensively within South Africa (Fig. 4A). By November 2020, the variant had spread into neighboring Botswana and Mozambique, and by December 2020, it had reached Zambia and Mayotte. Within the first 3 months of 2021,

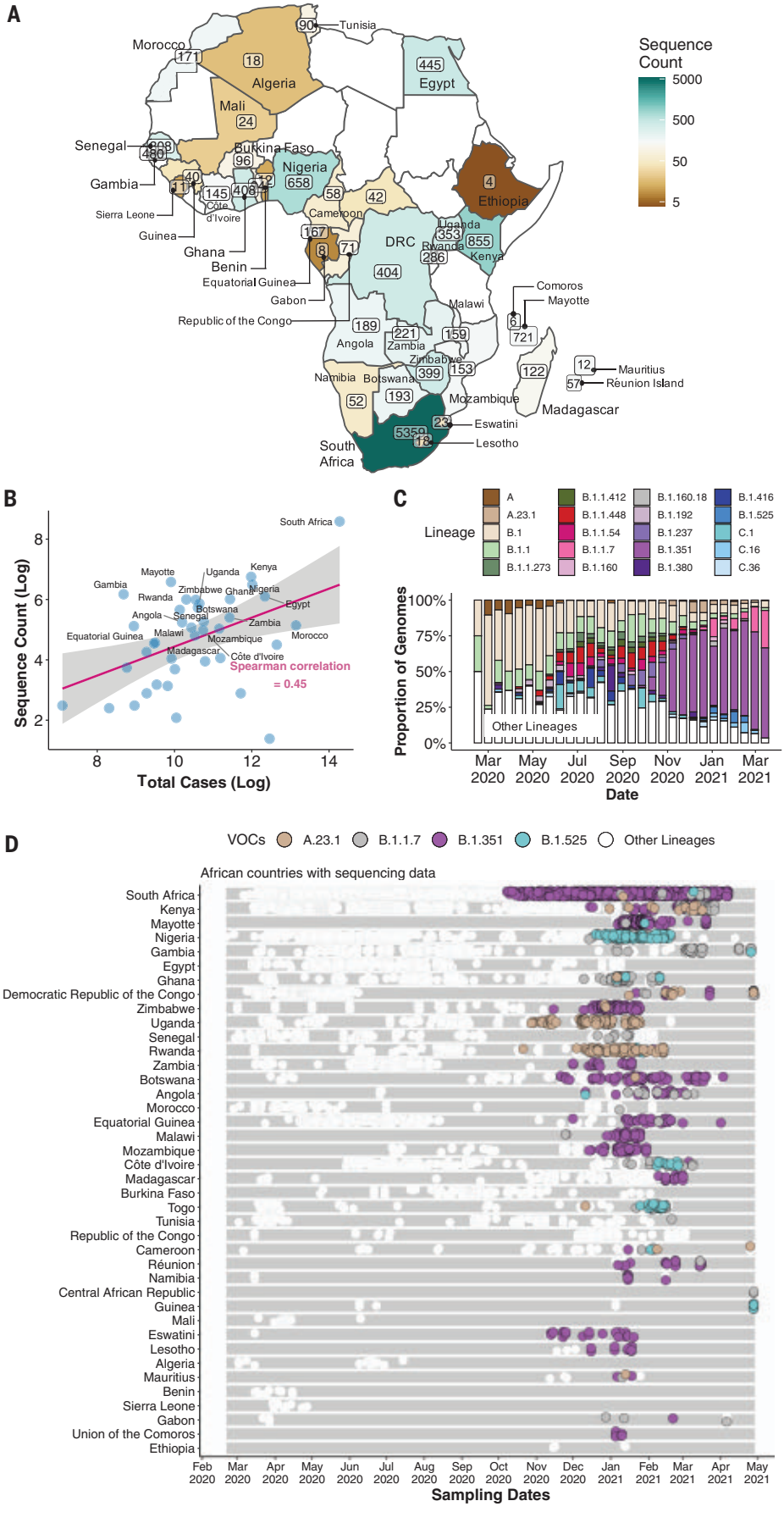


Fig. 1. SARS-CoV-2 sequences in Africa. (**A**) Map of the African continent with the number of SARS-CoV-2 sequences reflected in GISAID as of 5 May 2021. (**B**) Regression plot of the number of viral sequences versus the number of reported COVID-19 cases in various African countries as of 5 May 2021. Countries with >500 sequences are labeled. The shaded region indicates the 95% confidence interval. (**C**) Progressive distribution of the top 20 PANGO lineages on the African continent. (**D**) Temporal sampling of SARS-CoV-2 sequences in African countries (ordered by total number of sequences) through time, with VOCs of note highlighted and annotated according to their PANGO lineage assignment.

further exports from South Africa into Botswana, Zimbabwe, Mozambique, and Zambia occurred. By March 2021, B.1.351 had become the dominant lineage within most southern African countries as well as the overseas territories of Mayotte and Réunion (fig. S5). Our phylogeographic reconstruction also demonstrates movement of B.1.351 into East and Central Africa directly from southern Africa. Our discrete phylogeographic analysis of a wider sample of B.1.351 isolates demonstrates the spread of the lineage into West Africa. This patient from West Africa had a known travel history to Europe, so it is possible that the patient acquired the infection while in Europe or in transit and not from other African sources (fig. S6).

B.1.525 is a VOI defined by six substitutions in the spike protein [Q52R (Gln^{52}→Arg), A67V (Ala^{67}→Val), E484K, D614G, Q677H (Gln^{677}→His), and F888L (Phe^{888}→Leu)] and two deletions in the N-terminal domain [HV69-70Δ (deletion of His and Val at positions 69 and 70) and Y144Δ (deletion of Tyr at position 144)]. This was first sampled in the United Kingdom in mid-December 2020, but our phylogeographic reconstruction suggests that the variant originated in Nigeria in November 2020 [95% highest posterior density (HPD) 2020-11-01 to 2020-12-03] (Fig. 4B). Since then, it has spread throughout much of Nigeria and neighboring Ghana. Given sparse sampling from other neighboring countries within West and Central Africa (Fig. 1, A and C), the extent of the spread of this VOI in the region is not clear. Beyond Africa, this VOI has spread to Europe and the United States (fig. S6).

We designated A.23.1 and C.1.1 as VOIs for the purposes of this analysis because they present good examples of the continued evolution of the virus within Africa (*11*, *13*). Lineage A.23, characterized by three spike mutations [F157L (Phe^{157}→Leu), V367F (Val^{367}→Phe), and Q613H (Gln^{613}→His)], was first detected in a Ugandan prison in Amuru in July 2020 (95% HPD: 2020-07-15 to 2020-08-02). From there, the lineage was transmitted to Kitgum prison, possibly facilitated by the transfer of prisoners. Subsequently, the A.23 lineage spilled into the

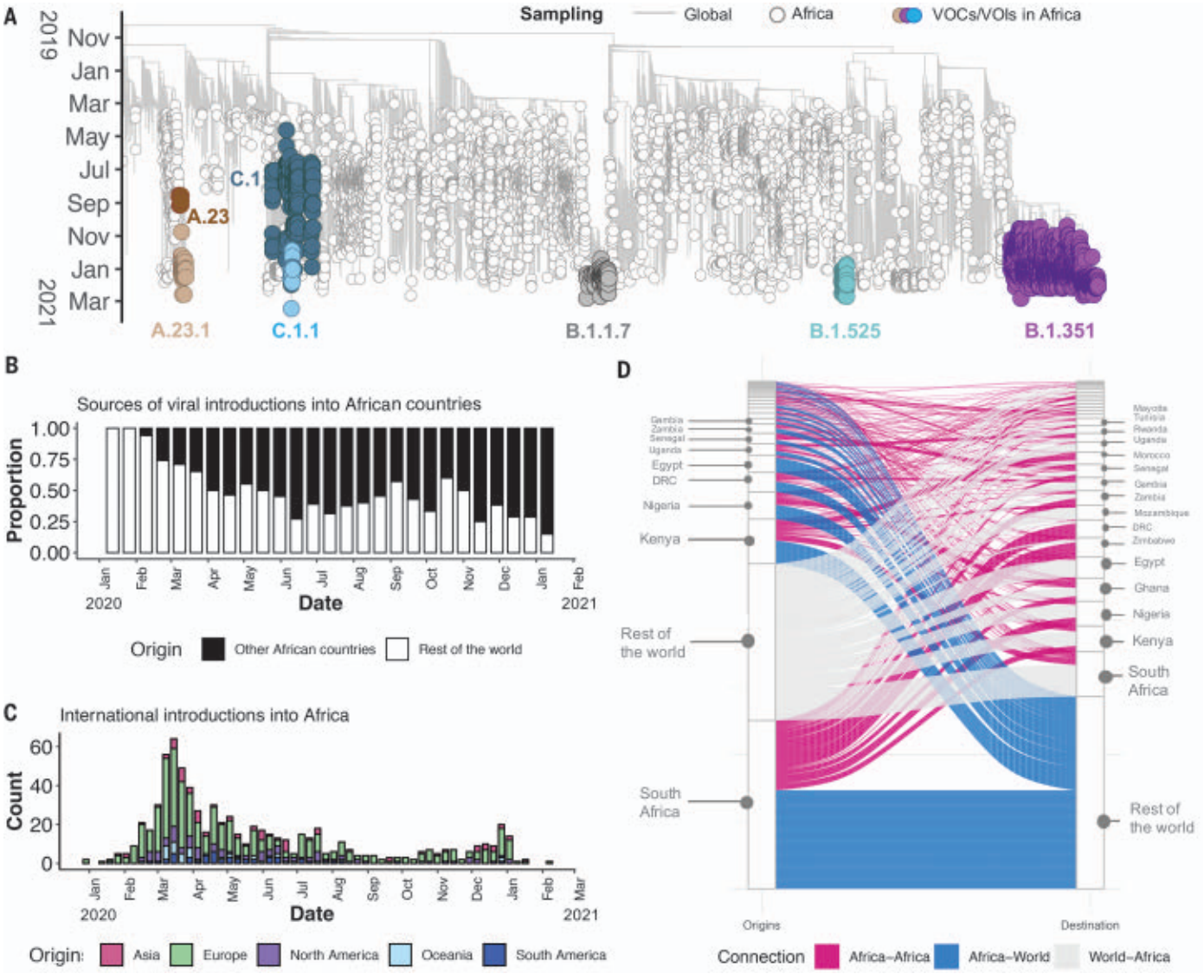


Fig. 2. Phylogenetic reconstruction of the SARS-CoV-2 pandemic on the continent of Africa. (**A**) Time-resolved maximum likelihood tree containing 8746 high-quality African SARS-CoV-2 near-full-genome sequences analyzed against a backdrop of global reference sequences. VOIs and VOCs are highlighted on the phylogeny. (**B**) Sources of viral introductions into African countries characterized as external introductions from the rest of the world versus internal introductions from other African countries. (**C**) Total external viral introductions over time into Africa. (**D**) The number of viral imports and exports into and out of various African countries depicted as internal (between African countries, in pink) or external (between African and non-African countries, in blue and gray).

general population and spread to Kampala, adding other spike mutations [R102I (Arg^{102}→Ile), L141F (Leu^{141}→Phe), E484K, and P681R (Pro^{681}→Arg)] along with additional mutations in nsp3, nsp6, ORF8, and ORF9, prompting a new lineage classification, A.23.1 (Fig. 3, A and B). Since the emergence of A.23.1 in September 2020 (95% HPD: 2020-09-02 to 2020-09-28), it has spread regionally into neighboring Rwanda and Kenya and has now also reached South Africa and Botswana in the south and Ghana in the west (Fig. 4C). However, our phylogeographic reconstruction of A.23.1 suggests that the introduction into Ghana may have occurred via Europe (fig. S6), whereas the introductions

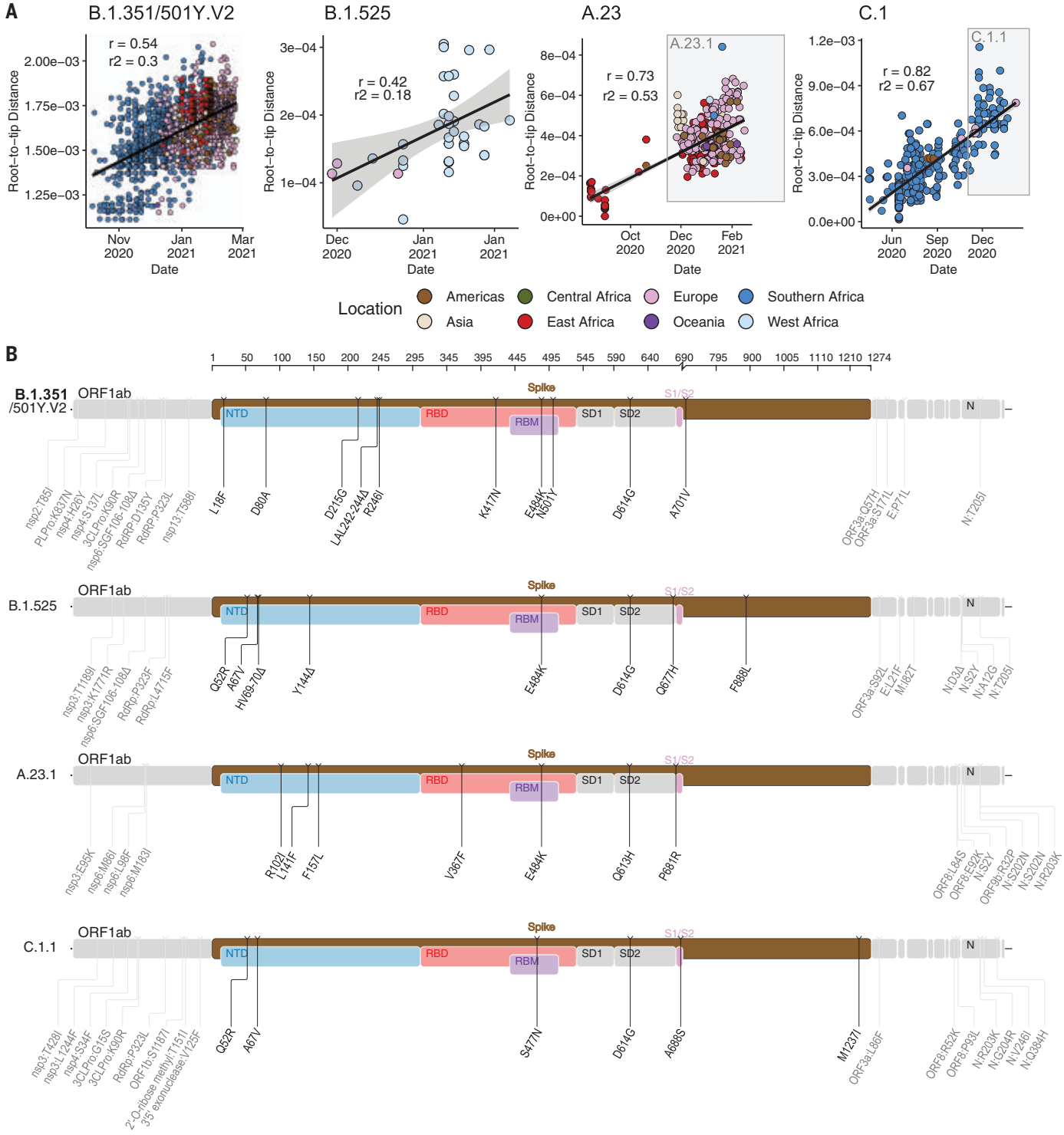


Fig. 3. Genetic profile of VOCs and VOIs under investigation. (**A**) Root-to-tip regression plots for four lineages of interest. C.1 and A.23 show continued evolution into VOIs C.1.1 and A.23.1, respectively. *r*, coefficient of correlation; r^2, coefficient of determination. (**B**) Genome maps of four VOCs and VOIs, where the spike region is shown in detail and in color and the rest of the genome is shown in gray. ORF, open reading frame; NTD, N-terminal domain; RBD, receptor binding domain; RBM, receptor binding motif; SD1, subdomain 1; SD2, subdomain 2.

into southern Africa likely occurred directly from East Africa. This is consistent with epidemiological data suggesting that the case detected in South Africa was a contact of an individual who had recently traveled to Kenya.

Lineage C.1 emerged in South Africa in March 2020 (95% HPD: 2020-03-13 to 2020-04-17) during a cluster outbreak before the first wave of the epidemic (*13*). C.1.1 is defined by the spike mutations S477N (Ser^{477}→Asn), A688S (Ala^{688}→Ser), and M1237I (Met^{1237}→Ile) and also contains the Q52R and A67V mutations similar to B.1.525 (Fig. 3B). A continuous trait phylogeographic reconstruction of the movement dynamics of these lineages suggests that C.1 emerged in the city of Johannesburg and spread within South Africa during the first wave (Fig. 4D). Independent exports of C.1 from South Africa led to regional spread to Zambia (June to July 2020) and Mozambique (July to August 2020), and the evolution to C.1.1 seems to have occurred in Mozambique around mid-September 2020 (95% HPD: 2020-09-07 to 2020-10-05). An in-depth analysis of SARS-CoV-2 genotypes from Mozambique suggests that the C.1.1 lineage was the most prevalent in the country until the introduction of B.1.351, which has dominated the epidemic since (fig. S5).

The VOC B.1.1.7, which was first sampled in Kent, England, in September 2020 (*22*), has also increased in prevalence in several African countries (fig. S5). To date, this VOC has been detected in 11 African countries, as well as the Indian Ocean islands of Mauritius and Mayotte (fig. S7). The time-resolved phylogeny suggests that this lineage was introduced into Africa on at least 16 occasions between November 2020 and February 2021, with evidence of local transmission in Nigeria and Ghana.

Conclusions

Our phylogeographic reconstruction of past viral dissemination patterns suggests a strong epidemiological linkage between Europe and Africa, with 64% of detectable viral imports into Africa originating in Europe and 41% of detectable viral exports from Africa landing in Europe (Fig. 1C). This phylogeographic analysis also

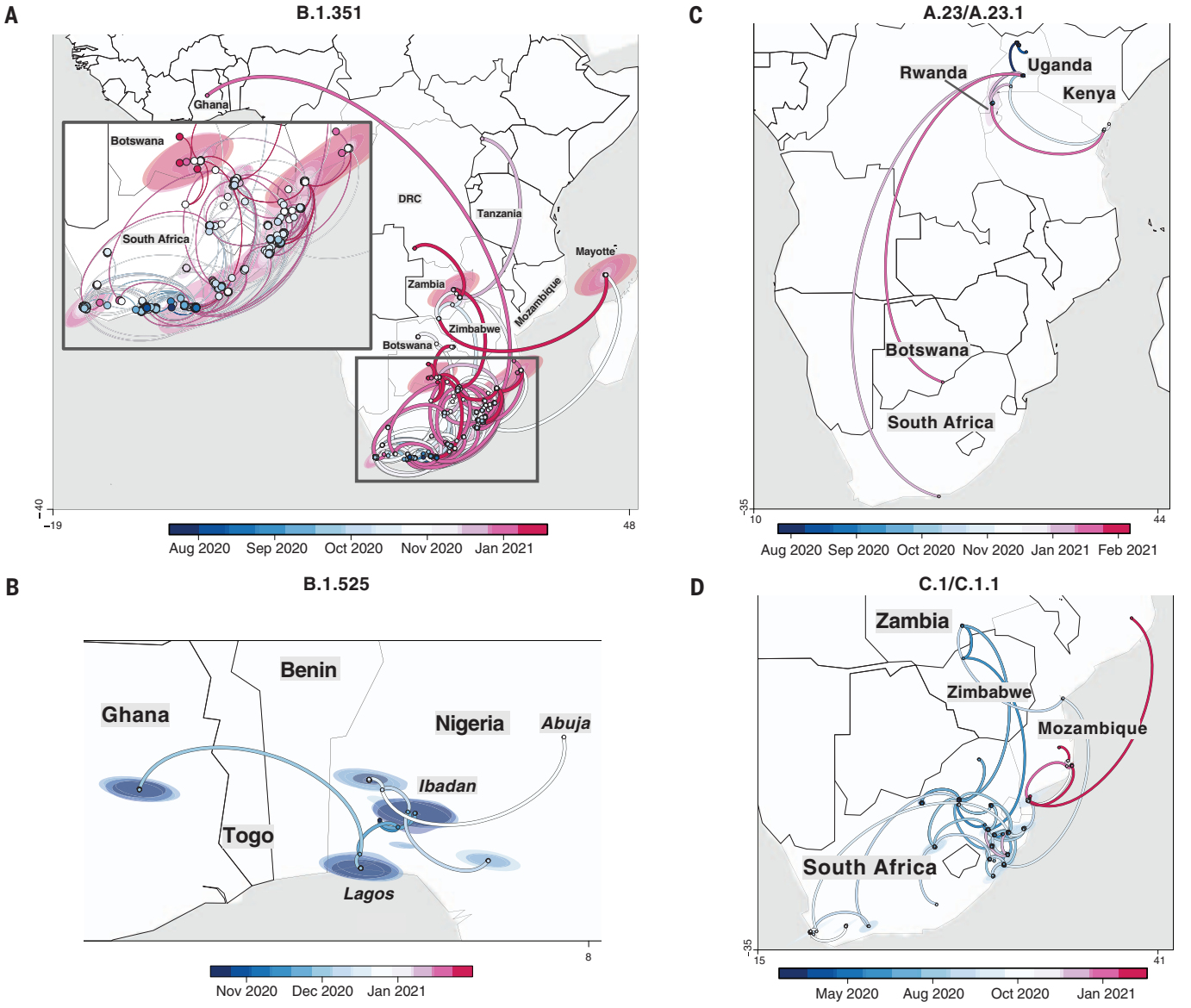


Fig. 4. Phylogeographic reconstruction of the spread of four VOCs and VOIs across the African continent. (**A** to **D**) Phylogeographic reconstruction of the spread of four VOCs and VOIs across the African continent using sequences showing strict continuous transmission across geographical regions: B.1.351 (A), B.1.525 (B), A.23/A.23.1 (C), and C.1/C.1.1 (D). Curved lines denote the direction of transmission in the counterclockwise direction. Solid lines show transmission paths as inferred by phylogeographic reconstruction and colored by date, whereas dashed lines show the known travel history of the particular case considered.

suggests a changing pattern of viral diffusion into and within Africa over the course of 2020. In almost all instances, the earliest introductions of SARS-CoV-2 into individual African countries were from countries outside Africa.

High rates of COVID-19 testing and consistent genomic surveillance in the south of the continent have led to the early identification of VOCs such as B.1.351 and VOIs such as C.1.1 (*13*). Since the discovery of these southern African variants, several other SARS-CoV-2 VOIs have emerged in different parts of the world, including elsewhere on the African continent, such as B.1.525 in West Africa and A.23.1 in East Africa. There is strong evidence that both of these VOIs are rising in frequency in the regions where they have been detected, which suggests that they may possess higher fitness than other variants in these regions. Although more-focused research on the biological properties of these VOIs is needed to confirm whether they should be considered VOCs, it would be prudent to assume the worst and focus on limiting their spread. It will be important to investigate how these different variants compete against one another if they occupy the same region.

Our focused phylogenetic analysis of the B.1.351 lineage revealed that in the final months of 2020, this variant spread from South Africa into neighboring countries, reaching as far north as the Democratic Republic of the Congo (DRC) by February 2021. This spread may have been facilitated through rail and road networks that form major transport arteries linking South Africa's ocean ports to commercial and industrial centres in Botswana, Zimbabwe, Zambia, and the southern parts of the DRC. The rapid, apparently unimpeded spread of B.1.351 into these countries suggests that current land-border controls that are intended to curb the international spread of the virus are ineffective. Perhaps targeted testing of cross-border travelers, genotyping of positive cases, and the focused tracking of frequent cross-border travelers, such as long distance truckers, would more effectively contain the spread of future VOCs and VOIs that emerge within this region.

The dominance of VOIs and VOCs in Africa has important implications for vaccine rollouts on the continent. For one, slow rollout of vaccines in most African countries creates an environment in which the virus can replicate and evolve: This will almost certainly produce additional VOCs, any of which could derail the global fight against COVID-19. Conversely, with the already widespread presence of known variants, difficult decisions about balancing reduced efficacy and availability of vaccines have to be made. This also highlights how crucial it is that trials are done. From a public health perspective, genomic surveillance is only one item in the toolkit of pandemic preparedness. It is important that such work is closely followed by genotype-to-phenotype research to determine the actual relevance of continued evolution of SARS-CoV-2 and other emerging pathogens.

The rollout of vaccines across Africa has been painfully slow (figs. S8 and S9). There have, however, been notable successes that suggest that the situation is not hopeless. The small island nation of the Seychelles had vaccinated 70% of its population by May 2021. Morocco has kept pace with many developed nations and, by mid-March, had vaccinated ~16% of its population. Rwanda, one of Africa's most resource-constrained countries, had, within 3 weeks of obtaining its first vaccine doses in early March, managed to provide first doses to ~2.5% of its population. For all other African countries, at the time of writing, vaccine coverage (first dose) was <1.0% of the general population.

The effectiveness of molecular surveillance as a tool for monitoring pandemics is largely dependent on continuous and consistent sampling through time, rapid virus genome sequencing, and rapid reporting. When this is achieved, molecular surveillance can ensure the early detection of changing pandemic characteristics. Further, when such changes are discovered, molecular surveillance data can also guide public health responses. In this regard, the molecular surveillance data that are being gathered by most African countries are less useful than they could be. For example, the time lag between when virus samples are taken and when sequences for these samples are deposited in sequence repositories is so great in some cases that the primary utility of genomic surveillance data is lost (fig. S10). This lag is driven by several factors, depending on the laboratory or country in question: (i) lack of reagents owing to disruptions in global supply chains, (ii) lack of equipment and infrastructure within the originating country, (iii) scarcity of technical skills in laboratory methods or bioinformatic support, and (iv) hesitancy by some health officials to release data. More-recent sampling and prompt reporting is crucial to reveal the genetic characteristics of currently circulating viruses in these countries.

The patchiness of African genomic surveillance data is therefore the main weakness of our study. However, there is evidence that the situation is improving, with ~50% of African SARS-CoV-2 genome sequences having been submitted to the GISAID database within the first 10 weeks of 2021. Although the precise factors underlying this surge in sequencing efforts are unclear, an important driver is almost certainly increased global interest in genomic surveillance after the discovery of multiple VOCs and VOIs since December 2020. We cannot reject that the observed increase in exports from Africa may be due to intensified sequencing activity after the detection of variants around the world. It is important to note here that phylogeographic reconstruction of viral spread is highly dependent on sampling where there is the caveat that the exact routes of viral movements between countries cannot be inferred if there is no sampling in connecting countries. Furthermore, our efforts to reconstruct the movement dynamics of SARS-CoV-2 across the continent are almost certainly biased by uneven sampling between different African countries. It is not a coincidence that we identified South Africa, Kenya, and Nigeria, which have sampled and sequenced the most SARS-CoV-2 genomes, as major sources of viral transmissions between sub-Saharan African countries. However, these countries also had the highest number of infections, which may decrease the sampling biases (Fig. 1A).

The reliability of genomic surveillance as a tool to prevent the emergence and spread of dangerous variants is dependent on the intensity with which it is embraced by national public health programs. As with most other parts of the world, the success of genomic surveillance in Africa requires that more samples are tested for COVID-19, higher proportions of positive samples are sequenced within days of sampling, and persistent analyses of these sequences are performed for concerning signals such as (i) the presence of novel nonsynonymous mutations at genomic sites associated with pathogenicity and immunogenicity, (ii) evidence of positive selection at codon sites where nonsynonymous mutations are observed, and (iii) evidence of lineage expansions. Despite limited sampling, Africa has identified many of the VOCs and VOIs that are being transmitted across the world. Detailed characterization of the variants and their impact on vaccine-induced immunity is of extreme importance. If the pandemic is not controlled in Africa, we may see the production of vaccine escape variants that may profoundly affect the population in Africa and across the world.

REFERENCES AND NOTES

1. C. Wang, P. W. Horby, F. G. Hayden, G. F. Gao, *Lancet* **395**, 470–473 (2020).
2. Q. Li *et al*., *N. Engl. J. Med.* **382**, 1199–1207 (2020).
3. S. Uyoga *et al*., *Science* **371**, 79–82 (2021).
4. L. Mwananyanda *et al*., *BMJ* **372**, n334 (2021).
5. S. J. Salyer *et al*., *Lancet* **397**, 1265–1275 (2021).
6. P. Oluniyi, First African SARS-CoV-2 genome sequence from Nigerian COVID-19 case. Virological (2020); https://virological.org/t/first-african-sars-cov-2-genome-sequence-from-nigerian-covid-19-case/421.
7. M. A. Medhat, M. El Kassas, *J. Glob. Health* **10**, 010368 (2020).
8. M. Allam *et al*., *Microbiol. Resour. Announc.* **9**, e00572-e20 (2020).
9. N. Haider *et al*., *BMJ Glob. Health* **5**, e003319 (2020).
10. S. C. Inzaule, S. K. Tessema, Y. Kebede, A. E. Ogwell Ouma, J. N. Nkengasong, *Lancet Infect. Dis.* **21**, e281–e289 (2021).
11. D. L. Bugembe *et al*., *medRxiv* 2021.02.08.21251393 (2021); https://doi.org/10.1101/2021.02.08.21251393.
12. J. Giandhari *et al*., *Int. J. Infect. Dis.* **103**, 234–241 (2021).
13. H. Tegally *et al*., *Nat. Med.* **27**, 440–446 (2021).
14. C. A. Pearson, T. W. Russell, N. Davies, A. J. Kucharski, Estimates of severity and transmissibility of novel SARS-CoV-2 variant 501Y.V2 in South Africa. CMMID Repository (2021); https://cmmid.github.io/topics/covid19/sa-novel-variant.html.

15. S. Cele *et al*., *medRxiv* 2021.01.26.21250224 (2021); https://doi.org/10.1101/2021.01.26.21250224.
16. S. A. Madhi *et al*., *N. Engl. J. Med.* **384**, 1885–1898 (2021).
17. C. K. Wibmer *et al*., *Nat. Med.* **27**, 622–625 (2021).
18. Y. Shu, J. McCauley, *Euro Surveill.* **22**, 30494 (2017).
19. E. Volz *et al*., *Cell* **184**, 64–75.e11 (2021).
20. B. Korber *et al*., *Cell* **182**, 812–827.e19 (2020).
21. T. de Oliveira *et al*., *medRxiv* 2021.03.30.21254323 (2021); https://doi.org/10.1101/2021.03.30.21254323.
22. B. Meng *et al*., *Cell Rep.* **35**, 109292 (2021).
23. S. E. James, T. Houriiyah, krisp-kwazulu-natal/africa-covid19-genomics: A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa - Code and scripts. Zenodo (2021); https://doi.org/10.5281/zenodo.5386379.

ACKNOWLEDGMENTS

We acknowledge the authors from the originating laboratories and the submitting laboratories, who generated and shared, via GISAID, the genetic sequence data on which this research is based (table S4). We also acknowledge the contribution of K. Maria from the NGS-SA platform for their contribution toward the sequencing effort in Cape Town, South Africa. Similarly, we thank A. M. Elsaame, S. M. Elsayed, and R. M. Darwish from the Faculty of Medicine Ain Shams Research Institute (MASRI) for their efforts toward sequencing in Egypt. We thank S. Bane, M. Sanogo, D. Diallo, A. Combo Georges Togo, and A. Coulibaly from the University Clinical Research Centre (UCRC) at the University of Sciences, Techniques, and Technologies of Bamako for the contribution they have made toward sequencing efforts in Mali. We acknowledge the contribution of M. Moeti and A. Salam Gueye from the WHO for their contribution toward combating SARS-CoV-2 on the African continent. We further wish to extend acknowledgment to S. Lutucuta and J. Morais from the Angolan Ministry of Health for their continued hard work with regards to SARS-CoV-2 sampling, sequencing, and pandemic response in Angola. From Malawi we wish to acknowledge the work of B. Chilima, B. Mvula, and M. Chitenje from the Malawian Ministry of Health for their work on the COVID-19 response within the country. **Funding:** The University of Ghana (WACCBIP) team was funded by a Wellcome/African Academy of Sciences Developing Excellence in Leadership Training and Science (DELTAS) grant (DEL-15-007 and 107755/Z/15/Z: Awandare); National Institute of Health Research (NIHR) (17.63.91) grants using UK aid from the UK government for a global health research group for genomic surveillance of malaria in West Africa (Wellcome Sanger Institute, UK) and the global research unit for Tackling Infections to Benefit Africa (TIBA partnership, University of Edinburgh); and a World Bank African Centres of Excellent grant (WACCBIP-NCDs: Awandare). Project ADAGE PRFCOV19-GP2 (2020-2022) includes 40 researchers from the Center of Biotechnology of Sfax, the University of Sfax, the University of Monastir, the University Hospital Hédi Chaker of Sfax, the Military Hospital of Tunis, and Dacima Consulting. Ministry of Higher Education and Scientific Research and Ministry of Health of the Republic of Tunisia. The Uganda contributions were funded by the UK Medical Research Council (MRC/UKRI) and the UK Department for International Development (DFID) under the MRC/DFID concordat agreement (grant agreement number NC_PC_19060) and by the Wellcome, DFID–Wellcome Epidemic Preparedness–Coronavirus (grant agreement number 220977/Z/20/Z) awarded to M.C. Work from Quadram Institute Bioscience was funded by The Biotechnology and Biological Sciences Research Council Institute Strategic Programme Microbes in the Food Chain BB/R012504/1 and its constituent projects BBS/E/F/000PR10348, BBS/E/F/000PR10349, BBS/E/F/000PR10351, and BBS/E/F/000PR10352 and by the Quadram Institute Bioscience BBSRC–funded Core Capability Grant (project number BB/CCG1860/1). The Africa Pathogen Genomics Initiative (Africa PGI) at the Africa CDC is supported by the Bill & Melinda Gates Foundation (INV018978 and INV018278), Illumina Inc, the US Centers for Disease Control and Prevention (CDC), and Oxford Nanopore Technologies. Sequences generated in Zambia through PATH were funded by the Bill & Melinda Gates Foundation. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation. Funding for sequencing in Côte d'Ivoire, Burkina Faso, and part of the sequencing in the DRC was granted by the German Federal Ministry of Education and Research (BMBF). Sequencing efforts from Morocco have been supported by Academie Hassan II of Science and Technology, Morocco. Funding for surveillance, sampling, and testing in Madagasar was provided by the WHO, the CDC (grant U5/IP000812-05), the US Agency for International Development (USAID; cooperation agreement 72068719CA00001), and the Office of the Assistant Secretary for Preparedness and Response in the US Department of Health and Human Services (DHHS; grant number IDSEP190051-01-0200). Funding for sequencing was provided by the Bill & Melinda Gates Foundation (GCE/ID OPP1211841), Chan Zuckerberg Biohub, and the Innovative Genomics Institute at UC Berkeley. The Botswana Harvard AIDS Institute was supported by the following funding: H3ABioNet through funding from the National Institutes of Health Common Fund (U41HG006941)—H3ABioNet is an initiative of the Human Health and Heredity in Africa Consortium (H3Africa) program of the African Academy of Science (AAS); DHHS–NIH–National Institute of Allergy and Infectious Diseases (NIAID) (5K24AI131928-04 and 5K24AI131924-04); Sub-Saharan African Network for TB/HIV Research Excellence (SANTHE); a DELTAS Africa Initiative (grant DEL-15-0060—the DELTAS Africa Initiative is an independent funding scheme of the AAS's Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [grant 107752/Z/15/Z] and the UK government; and the South African Medical Research Council (SAMRC) and the Department of Technology and Innovation as part of the Network for Genomic Surveillance in South Africa (NGS-SA) and the Stellenbosch University Faculty of Medicine & Health Sciences, Strategic Equipment Fund. D.P.M. is funded by the Wellcome Trust (Wellcome Trust grant 222574/Z/21/Z). Sequencing activities at the NICD were supported by a conditional grant from the South African National Department of Health as part of the emergency COVID-19 response; a cooperative agreement between the National Institute for Communicable Diseases of the National Health Laboratory Service and the U.S. Centers for Disease Control and Prevention (grant number 5 U01IP001048-05-00); the African Society of Laboratory Medicine (ASLM) and Africa Centers for Disease Control and Prevention through a sub-award from the Bill and Melinda Gates Foundation grant number INV-018978; the UK Foreign, Commonwealth and Development Office and Wellcome (Grant no 221003/Z/20/Z); the South African Medical Research Council (Reference number SHIPNCD 76756); the UK Department of Health and Social Care, managed by the Fleming Fund and performed under the auspices of the SEQAFRICA project. Furthermore, pandemic surveillance in South Africa and Senegal was supported in part through NIH grant U01 AI151698 for the United World Antiviral Research Network (UWARN). Support for pandemic surveillance from the Tulio de Oliveira group to other African countries is funded by the Rockefeller Foundation. Sequencing efforts in the DRC were funded by the Bill & Melinda Gates Foundation under grant INV-018030 awarded to C.B.P. and further supported by funding from the Africa CDC through the ASLM (African Society of Laboratory Medicine) for Accelerating SARS-CoV-2 Genomic Surveillance in Africa. Sequencing efforts in Rwanda were commissioned by the NIHR Global Health Research program (16/136/33) using UK aid from the UK government (funding to E.M. and N. R. through TIBA partnership) and additional funds from the government of Rwanda through RBC/National Reference Laboratory in collaboration with the Belgian Development Agency (ENABEL) for additional genomic sequencing at GIGA Research Institute–Liege/Belgium. The sequencing effort in Equatorial Guinea was supported by a public-private partnership, the Bioko Island Malaria Elimination Project, composed of the government of Equatorial Guinea Ministries of Mines and Hydrocarbons, and Health and Social Welfare, Marathon EG Production Limited, Noble Energy, Atlantic Methanol Production Company, and EG LNG. Sample collection and typing in Mali were supported by Fondation Merieux–France, and sequence efforts have been supported by the Enable and Enhance Initiative of the German Federal Government's Security Cooperation against Biological Threats in the G5 Sahel Region. The Nigeria work was made possible by support from Flu Lab and a cohort of generous donors through TED's Audacious Project, including the ELMA Foundation, MacKenzie Scott, the Skoll Foundation, and Open Philanthropy. Further Nigeria funding came from grants from the NIAID (www.niaid.nih.gov), NIH-H3Africa (https://h3africa.org) (U01HG007480 and U54HG007480), and the World Bank grant (worldbank.org) (ACE IMPACT project) to C.H. Analysis for the Gabon strains was supported by the Science and Technology Research Partnership for Sustainable Development (SATREPS), Japan International Cooperation Agency (JICA), and Japan Agency for Medical Research and Development (AMED) (grant number JP20jm0110013) and a grant from AMED (grant number JP20wm0225003). Sequencing at KEMRI-Wellcome Trust Research Programme site in Kenya was supported by the National Institute for Health Research (NIHR) (project references 17/63/82 and 16/136/33), using UK aid from the UK Government to support global health research, and the UK Foreign, Commonwealth and Development Office and Wellcome Trust (grant# 102975; 220985). **Author contributions:** Conceptualization: E.W., H.T., J.G.S., J.O., J.O.G., K.O.D., R.A.D., R.A.K., R.L., S.K.T., S.Ma., T.d.O. Methodology: A.-R.N.Z., A.Re., C.N.A., D.P.M., D.A.R., E.W., H.T., J.A.E., J.Gy., J.G.S., K.H.Y., K.O.D., L.d.O.M., M.A.B., M.C., M.G., M.M.N., M.V.T.P., P.A., T.d.O., V.F. Investigation: A.E., A.Ir., A.-R.N.Z., A.Re., A.So., A.v.G., C.A.K., C.W., D.C., D.J.N., D.P.M., D.A.R., D.S., E.M., E.N.N., E.W., F.L., G.G., H.T., J.A.E., J.E.S., J.Gy., J.G.S., J.-J.M.T., J.L., J.Nk., J.Q.M., K.H.Y., M.-M.D., M.A.B., M.C., M.G., M.Mar., M.Mas., M.S., M.V.T.P., N.A., N.H.R., N.K., P.K., R.A.A.C., R.A.D., R.G., S.A.M., S.F.A., S.Ma., S.O., T.L.V., V.F., W.P. Sampling: A.-S.K., A.N.A., A.A.A., A.A.S., A.D., A.E., A.E.O.O., A.F., A.G., A.H., A.Ib., A.Ka., A.K.Sa., A.K.Se., A.L., A.M.O., A.O.O.O., A.P., A.Ro., A.Sy., A.F.V., A.V.G., A.v.G., B.B., B.L.H., B.Ko., B.Kl., B.Ma., B.N., B.S.O., B.T., C.N.A., C.D., C.P., C.S., C.W., D.Bas., D.C., D.D., D.G., D.G.A., D.J.N., D.S., E.K.L., E.M., E.M.O., E.N.N., E.P., E.Sh., F.Ab., F.Ad., F.O.E., F.M.M., F.S.M., F.O., F.T., F.T.T., G.A.A., G.G., G.K.M., G.P.M., G.T., G.v.Z., H.C., H.A.E., H.N., I.C., I.Gaa., I.Gaz., I.K., I.M., I.O., I.Ss., J.C.A., J.-C.M., J.B.L.-D., J.E.S., J.Gi., J.Gy., J.J.L., J.K., J.L., J.M.H., J.M.M., J.Nam., J.Nak., J.T.K., K.T.A., K.M.S., L.B., M.-M.D., M.Al., M.C., M.D., M.e.H., M.G.S., M.K.K., M.Mp., M.Mar., M.Mas., M.M.D., M.N., M.Ou., M.O.A., M.R., M.S., M.W.M., M.T.Y., N.A., N.G., N.H., N.H.R., N.Ig., N.K., N.Ma., N.Nd., N.B.S., N.S., O.Ch., O.E.C., O.Fa., O.Fe., O.I., O.J., O.C.K., O.E.O., O.P., O.S., O.Te., P.E.O., P.A.B., P.C., P.C.S., P.D., P.K., P.K.Q., P.O., P.S., R.A.D., R.G., R.N., S.A., S.S.A., S.B., S.B.L., S.D., S.EI., S.E.K., S.F.A., S.Gas., S.Kam., S.L., S.L.D., S.Me., S.Mo., S.M.M., S.N., S.Pi., S.S., S.To., T.L.V., T.Ma., T.S., U.C., U.G., U.J., U.R., V.G., W.K.A., W.T.C., W.P., W.H.R., Y.Bu., Y.K.T., Y.N., Z.R.D. Sequencing: A.-S.Kam., A.N.A., A.A.A., A.A.S., A.C., A.D., A.E.O.O., A.F., A.Ib., A.Is., A.Ko., A.Ka., A.K.K., A.K.Sa., A.K.Se., A.L., A.-R.N.Z., A.P., A.Sa., A.So., A.Sy., A.S.O., A.J.T., A.F.V., A.V.G., A.v.G., A.A.Y., B.B., B.D., B.L.H., B.Ko., B.Kl., B.Mc., B.N., B.T., C.N.A., C.B., C.B.P., C.D., C.M.M., C.P., C.S., D.Bas., D.D., D.G., D.G.A., D.J.B., D.L.B., D.M., D.O.O., D.P., D.S.Y.A., D.T., E.E.F., E.F.N., E.K.L., E.L., E.M.O., E.P., E.Sh., E.Si., E.S.L., F.Ab., F.Aj., F.A.D., F.D., F.M.M., F.S.M., F.O., F.T.T., F.W., G.A.A., G.G., G.P.M., G.T., G.v.Z., H.Ab., H.An., H.C., H.C.R., H.A.E., H.G., H.H.K., H.N., I.B.-B.B., I.C., I.Gaa., I.Gaz., I.K., I.M., I.Ss., J.C.A., J.B., J.-C.M., J.B.L.-D., J.F., J.Gi., J.Gy., J.J.L., J.K., J.M.H., J.M.M., J.M.N., J.Nam., J.Nak., J.Q.M., J.T.K., J.Y., K.T.A., K.M.S., K.O.D., K.S., K.A.T., L.B., L.F., L.S., L.T., M.-M.D., M.Al., M.A.B., M.C., M.D., M.e.H., M.G.S., M.I.M., M.K.K., M.Mi., M.Mp., M.Mw., M.M.D., M.M.N., M.Ow., M.Ou., M.O.A., M.V.T.P., M.W.M., M.T.Y., N.D., N.G., N.H., N.Ig., N.Is., N.Ma., N.Nd., N.Ns., N.B.S., N.S., N.T., O.Cy., O.Ch., O.E.C., O.Fa., O.I., O.J., O.Te., P.A., P.A.B., P.C.S., P.D., P.E.O., P.K., P.K.Q., P.K.M., P.O., P.S., R.A.A.C., R.G., R.N., R.O.P., S.A., S.S.A., S.B., S.B.L., S.C.S., S.D., S.EI., S.En., S.E.K., S.Gar., S.Gas., S.H.Ab., S.Kas., S.L., S.L.D., S.Me., S.Mo., S.Ma., S.M.M., S.N., S.Pi., S.Pr., S.R., S.S., S.To., S.Tr., S.v.d.W., T.A., T.Ma., T.Mo., T.S., U.C., U.G., U.J., U.J.A., U.R., V.G., W.K.A., W.T.C., W.H.R., Y.Bu., Y.K.T., Y.N., Z.R.D. Visualization: A.C., A.Is., A.Ko., A.K.K., A.Sa., A.So., A.A.Y., B.T., C.B., C.M.M., D.Bak., D.O.O., D.P., D.A.R., D.S.Y.A., E.A.A., E.B., E.S.L., E.W., F.A.D., F.B., F.D., F.W., G.S., H.Ab., H.An., H.G., H.L., H.T., I.B.A., I.Ss., J.A.E., J.B., J.F., J.Gy., J.M.N., J.Y., K.H.Y., K.S., L.F., L.S., L.T., M.Ao., M.Al., M.G., M.T., M.V.T.P., M.R.W., N.D., N.Is., N.K., N.Ns., N.T., O.Cy., O.To., P.A., P.C.S., P.E.O., R.A.A.C., S.B., S.F.S., S.H.A., S.Kas., S.Ma., T.A., T.Mo., V.E., Y.Be. Funding acquisition: A.J.P., A.Re., A.v.G., B.Ko., C.N.A., C.A.K., C.B.P., C.W., D.C., D.J.B., D.J.N., F.L., G.A.A., G.G., G.P.M., H.C., J.E.S., J.-J.M.T., J.L., J.M.H., J.Nk., J.O., K.O.D., M-M.D., M.C., M.I.M., M.Mas., M.V.T.P., N.A., P.C.S., P.K., P.K.M., R.A.K., S.A.M., S.EI., S.Mo., S.v.d.W., T.d.O., W.P. Project administration: A.J.P., A.Re., A.F.V., A.v.G., B.Ko., C.W., D.J.B., D.J.N., E.W., F.Aj., F.T., G.A.A., G.P.M., G.S., G.T., H.C., J.C.O., J.-J.M.T., J.M.H., J.O., J.O.G., J.Y., K.O.D., M.C., M.K., M.Mar., M.P., M.V.T.P., M.R.W., N.R., O.To., P.C.S., P.K., P.K.M., R.A.K., S.A.M., S.EI., S.F.S., S.Gas., S.Mo., T.d.O. Supervision: A.J.P., A.Re., B.Ko., C.W., D.J.N., E.N.N., E.W., F.T., G.A.A., G.L.K., H.C., J.B., J.M.H., J.Nk., J.O., J.O.G., K.O.D., M.Al., M.C., M.I.M., M.Mar., M.M.N., M.S., N.Mu., N.R., P.C.S., P.K., P.K.M., R.A.K., S.EI., S.E.K., S.Gas., S.Me., S.Mo., S.Pa., T.d.O. Writing – original draft: A.K.Sa., A.-R.N.Z., B.Ko., D.P.M., E.W., F.T., G.L.K., H.T., J.B., J.-C.M., M.Al., M.A.B., M.C., M.G., M.Mi., N.Mu., R.L. Writing – review and editing: A.-R.N.Z., B.Ko., C.M.M., D.J.N., D.P.M., D.A.R., D.S.Y.A., D.T., E.K.L., E.L., E.S.L., E.W., H.T., J.E.S., J.G.S., L.d.O.M., M.A.B., M.C., M.e.H., P.K.Q., P.K.M., R.L., S.K.T., T.d.O., U.J.A. Author contributions are listed alphabetically. A full list of author abbreviations is included on the GitHub repository (https://github.com/krisp-kwazulu-natal/africa-covid19-genomics) (*23*). **Competing interests:** P.C.S. is a founder and shareholder of Sherlock Biosciences and is on the board and serves as shareholder of the Danaher Corporation. The authors declare no other conflicts of interest. **Data and materials availability:** All sequences that were used in the present study are listed in table S4 (accessible on the GitHub repository) along with their GISAID sequence IDs, dates of sampling, the originating and submitting laboratories, and main authors. All input files (e.g., alignments or XML files), all resulting output files, and scripts used in the study are shared publicly on GitHub (https://github.com/krisp-kwazulu-natal/africa-covid19-genomics) (*23*). 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.abj4336
Materials and Methods
Figs. S1 to S10
Tables S1 to S4
References (*24*–*38*)
MDAR Reproducibility Checklist

[1]KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), Nelson R Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa. [2]Centre for Epidemic Response and Innovation (CERI), School of Data Science and Computational Thinking, Stellenbosch University, Stellenbosch, South Africa. [3]Laboratorio de Flavivirus, Fundacao Oswaldo Cruz, Rio de Janeiro, Brazil. [4]Laboratório de Genética Celular e Molecular, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil. [5]Department of Geography and GIS, University of Cincinnati, Cincinnati, OH, USA. [6]Institute of Infectious Diseases and Molecular Medicine, Department of Integrative Biomedical Sciences, Computational Biology Division, University of Cape Town, Cape Town, South Africa. [7]Division of Medical Virology, Wellcome Centre for Infectious Diseases in Africa, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa. [8]Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC, USA. [9]Bioinformatics Research Center, North Carolina State University, Raleigh, NC, USA. [10]Cancer Biology Department, Virology and Immunology Unit, National Cancer Institute, Cairo University, Cairo 11796, Egypt. [11]Centre d'Infectiologie Charles Mérieux-Mali (CICM-Mali), Bamako, Mali. [12]Bacteriology and Virology Department Souro Sanou University Hospital, Bobo-Dioulasso, Burkina Faso. [13]MRCG at LSHTM Genomics Lab, Fajara, Gambia. [14]African Centre of Excellence for Genomics of Infectious Diseases (ACEGID), Redeemer's University, Ede, Osun State, Nigeria. [15]Department of Virology, College of Medicine, University of Ibadan, Ibadan, Nigeria. [16]Department of Medical Microbiology and Parasitology, Faculty of Basic Clinical Sciences, College of Health Sciences, University of Ilorin, Ilorin, Kwara State, Nigeria. [17]CHU Habib Bourguiba, Laboratory of Microbiology, Faculty of Medicine of sFax, University of sFax, sFax, Tunisia. [18]Pathogen Sequencing Lab, Institut National de Recherche Biomédicale (INRB), Kinshasa, Democratic Republic of the Congo. [19]Université de Kinshasa (UNIKIN), Kinshasa, Democratic Republic of the Congo. [20]Genomics Research Program, Children's Cancer Hospital, Cairo, Egypt. [21]Institute of Pathogen Genomics, Africa Centres for Disease Control and Prevention (Africa CDC), Addis Ababa, Ethiopia. [22]Laboratory of Transmissible Diseases and Biological Active Substances (LR99ES27), Faculty of Pharmacy of Monastir, Monastir, Tunisia. [23]Laboratory of Microbiology, University Hospital of Monastir, Monastir, Tunisia. [24]Internal Medicine Department, Alex Ekwueme Federal University Teaching Hospital, Abakaliki, Nigeria. [25]Irrua Specialist Teaching Hospital, Irrua, Nigeria. [26]Quadram Institute Bioscience, Norwich, UK. [27]World Health Organization, Africa Region, Brazzaville Congo. [28]Centre de Recherche et de Formation en Infectiologie de Guinée (CERFIG), Université de Conakry, Conakry, Guinea. [29]TransVIHMI, Montpellier University/IRD/INSERM, Montpellier, France. [30]Virology Department, Institut Pasteur de Dakar, Dakar, Senegal. [31]Mali-University Clinical Research Center (UCRC), Bamako, Mali. [32]Laboratory of Molecular and Cellular Screening Processes, Centre of Biotechnology of Sfax, University of Sfax, Sfax, Tunisia. [33]Broad Insitute of Harvard and MIT, Cambridge, MA, USA. [34]KEMRI-Wellcome Trust Research Programme/KEMRI-CGMR-C, Kilifi, Kenya. [35]National Institute for Communicable Diseases (NICD) of the National Health Laboratory Service (NHLS), Johannesburg, South Africa. [36]The Nigerian Institute of Medical Research, Yaba, Lagos, Nigeria [37]Institute of Virology, Charité – Universitätsmedizin, Berlin, Germany. [38]Medical Biotechnology Laboratory, Rabat Medical and Pharmacy School, Mohammed V University, Rabat, Morocco. [39]Department of Immunology, University of Maiduguri Teaching Hospital, P.M.B. 1414, Maiduguri, Nigeria. [40]MRC/UVRI and LSHTM Uganda Research Unit, Entebbe, Uganda. [41]Division of Medical Virology, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, Cape Town, South Africa. [42]The Biotechnology Center of the University of Yaoundé I, Cameroon and CDC Foundation, Yaounde, Cameroon. [43]Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA. [44]Virology Unit, Institut Pasteur de Madagascar, Antananarivo, Madagascar. [45]University of Nebraska Medical Center (UNMC), Omaha, NE, USA. [46]Antibody Immunity Research Unit, School of Pathology, University of the Witwatersrand, Johannesburg, South Africa. [47]CHU de Bouaké, Laboratoire/Unité de Diagnostic des Virus des Fièvres Hémorragiques et Virus Émergents, Bouaké, Côte d'Ivoire. [48]School of Public Health, Pwani University, Kilifi, Kenya. [49]Swiss Tropical and Public Health Institute, Basel, Switzerland. [50]West African Centre for Cell Biology of Infectious Pathogens (WACCBIP), Department of Biochemistry, Cell and Molecular Biology, University of Ghana, Accra, Ghana. [51]School of Life Sciences and Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research (SBIDER), University of Warwick, Coventry, UK. [52]Uganda Virus Research Institute, Entebbe, Uganda. [53]Division of Virology, National Health Laboratory Service and University of the Free State, Bloemfontein, South Africa. [54]Botswana Harvard AIDS Institute Partnership and Botswana Harvard HIV Reference Laboratory, Gaborone, Botswana. [55]University of Zambia, School of Veterinary Medicine, Department of Disease Control, Lusaka, Zambia. [56]African Society for Laboratory Medicine, Addis Ababa, Ethiopia. [57]Functional Genomic Platform/National Centre for Scientific and Technical Research (CNRST), Rabat, Morocco. [58]Rwanda National Reference Laboratory, Kigali, Rwanda. [59]Robert Koch-Institute, Berlin, Germany. [60]G5 Evolutionary Genomics of RNA Viruses, Institut Pasteur, Paris, France. [61]Research Unit of Autoimmune Diseases UR17DN02, Military Hospital of Tunis, University of Tunis El Manar, Tunis, Tunisia. [62]Department of Public Health, Ministry of Health, Ilorin, Kwara State, Nigeria. [63]Laboratory of Clinical Virology, Institut Pasteur de Tunis, Tunis, Tunisia. [64]Faculty of Pharmacy of Monastir, Monastir, Tunisia. [65]National Microbiology Reference Laboratory, Harare, Zimbabwe. [66]National Influenza Centre, Viral Respiratory Laboratory, Algiers, Algeria. [67]Medical Microbiology and Parasitology Department, College of Medicine, University of Ibadan, Ibadan, Nigeria. [68]Lincoln International Institute for Rural Health, University of Lincoln, Lincoln, UK. [69]Centre for the AIDS Programme of Research in South Africa (CAPRISA), Durban, South Africa. [70]Africa Health Research Institute, KwaZulu-Natal, Durban, South Africa. [71]Department of Biochemistry and Biotechnology, Pwani University, Kilifi, Kenya. [72]National Health Laboratory Service (NHLS), Tygerberg, Cape Town, South Africa. [73]Institution and Department, Ministry Of Health, COVID-19 Testing Laboratory, Mbabane, Kingdom of Eswatini. [74]Laboratory of Health Sciences and Technologies, High Institute of Health Sciences, Hassan 1st University, Settat, Morocco. [75]Department of Emerging Infectious Diseases, Institute of Tropical Medicine, Nagasaki University, Nagasaki, Japan. [76]Central Public Health Laboratories (CPHL), Kampala, Uganda. [77]Faculty of Medicine Ain Shams Research institute (MASRI), Ain Shams University, Cairo, Egypt. [78]Charles Nicolle Hospital, Laboratory of Microbiology, National Influenza Center, 1006 Tunis, Tunisia. [79]Laboratory of Transmissible Diseases and Biological Active Substances (LR99ES27), Faculty of Pharmacy of Monastir, University of Monastir, Monastir, Tunisia. [80]Department of Biochemistry and Molecular Biology, The Institute for Medical Research Israel-Canada, Hadassah Medical School, The Hebrew University of Jerusalem, Jerusalem, Israel. [81]Namibia Institute of Pathology, Windhoek, Namibia. [82]Centre Interdisciplinaires de Recherches Medicales de Franceville (CIRMF), Franceville, Gabon. [83]Department of Biostatistics and Data Science, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, USA. [84]Urban Health Collaborative, Dornsife School of Public Health, Drexel University, Philadelphia, PA, USA. [85]UHAS COVID-19 Testing and Research Centre, University of Health and Allied Sciences, Ho, Ghana. [86]Rollins School of Public Health, Emory University, Atlanta, GA, USA. [87]Anoual Laboratory, Casablanca, Morocco. [88]Botswana Institute for Technology Research and Innovation, Gaborone, Botswana. [89]New York University Grossman School of Medicine, New York City, NY, USA. [90]Centre de Recherches Medicales de Lambarene (CERMEL), Lambarene, Gabon. [91]Virology/Molecular Biology Department, Central Health Laboratory, Ministry of Health and Wellness, Mauritius. [92]Center of Scientific Excellence for Influenza Viruses, National Research Centre (NRC), Cairo Egypt. [93]Ministry of Health and Wellness, Gaborone, Botswana. [94]Next Generation Sequencing Unit and Division of Virology, Faculty of Health Sciences, University of the Free State, Bloemfontein 9300, South Africa. [95]National Reference Laboratory Lesotho, Maseru, Lesotho. [96]Centre for Biotechnology Research and Development, Kenya Medical Research Institute, Nairobi, Kenya. [97]Laboratorio de Investigaciones de Baney, Baney, Equatorial Guinea. [98]Ifakara Health Institute, Dar-es-Salaam, Tanzania. [99]Nigeria Centre for Disease Control, Abuja, Nigeria. [100]Department of Medical Diagnostics, Kumasi Centre for Collaborative Research in Tropical Medicine, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. [101]Department of Medical Laboratory Science, Niger Delta University, Bayelsa State, Nigeria. [102]Systems and Biomedical Engineering Department, Faculty of Engineering, Cairo University, Cairo 12613, Egypt. [103]King Faisal Specialist Hospital and Research Center, Riyadh, Kingdom of Saudi Arabia. [104]Biological Prevention Department, Main Chemical Laboratories, Egypt Army, Cairo, Egypt. [105]PATH, Lusaka, Zambia. [106]Department of Biotechnology, High Institute of Biotechnology of Sidi Thabet, University of Manouba, BP-66, 2020 Ariana-Tunis, Tunisia. [107]Genomic Center for Human Pathologies (GENOPATH), Faculty of Medicine and Pharmacy, Mohammed V University, Rabat, Morocco. [108]Rwanda National Joint Task Force COVID-19, Rwanda Biomedical Centre, Ministry of Health, Kigali, Rwanda. [109]School of Health Sciences, College of Medicine and Health Sciences, University of Rwanda, Kigali, Rwanda. [110]Instituto Nacional de Saude (INS), Maputo, Mozambique. [111]National Health Laboratory Service (NHLS), Cape Town, South Africa. [112]Computational Biology Division, Department of Integrative Biomedical Sciences, IDM, CIDRI Africa Wellcome Trust Centre, University of Cape Town, Cape Town, South Africa. [113]Virology Laboratory, Alex Ekwueme Federal University Teaching Hospital, Abakaliki, Nigeria. [114]Department of Epidemiology and Community Health, Faculty of Clinical Sciences, College of Health Sciences, University of Ilorin, Ilorin, Kwara State, Nigeria. [115]Alex Ekwueme Federal University Teaching Hospital, Abakaliki, Nigeria. [116]Mayotte Hospital Center, Mayotte, France. [117]Virology Service, Centre Pasteur of Cameroun, Yaounde, Cameroon. [118]Kumasi Centre for Collaborative Research in Tropical Medicine, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. [119]Laboratoire de Recherche et d'Analyses Médicales de la Gendarmerie Royale, Rabat, Morocco. [120]Clinical and Experimental Pharmacology Lab, LR16SP02, National Center of Pharmacovigilance, University of Tunis El Manar, Tunis, Tunisia. [121]CHU Hedi Chaker Sfax, Service de Pneumologie, Tunis, Tunisia. [122]Laboratoire de Recherche et d'Analyses Médicales de la Gendarmerie Royale, Rabat, Morocco. [123]Central Public Health Laboratories (CPHL), Cairo, Egypt. [124]Centre MURAZ, Ouagadougou, Burkina Faso. [125]National Institute of Public Health of Burkina Faso (INSP/BF), Ouagadougou, Burkina Faso. [126]National Reference Center for Respiratory Viruses, Molecular Genetics of RNA Viruses, UMR 3569 CNRS, University of Paris, Institut Pasteur, Paris, France. [127]Sub-Saharan African Network For TB/HIV Research Excellence (SANTHE), Durban, South Africa. [128]Coordenação Geral de Laboratórios de Saúde Pública/Secretaria de Vigilância em Saúde, Ministério da Saúde, Brasília, Distrito Federal, Brazil. [129]World Health Organization, WHO Lesotho, Maseru, Lesotho. [130]Med24 Medical Centre, Ruwa, Zimbabwe. [131]Division of Human Genetics, Department of Pathology, University of Cape Town, Cape Town, South Africa. [132]Center for Human Genetics, College of Medicine and Health Sciences, University of Rwanda, Kigali, Rwanda. [133]Laboratory of Human Genetics, GIGA Research Institute, Liège, Belgium. [134]National Health Laboratory, Gaborone, Botswana. [135]MRC-University of Glasgow Centre for Virus Research, Glasgow, UK.[136]Harvard T.H. Chan School of Public Health, Boston, MA, USA. [137]Department of Global Health, University of Washington, Seattle, WA, USA. [138]Centre for Human Virology and Genomics, Nigerian Institute of Medical Research, Yaba, Lagos, Nigeria. [139]School of Pathology, Faculty of Health Science, University of the Witwatersrand, Johannesburg, South Africa.

12 May 2021; accepted 3 September 2021
Published online 9 September 2021
10.1126/science.abj4336

ORGANIC CHEMISTRY

General method for iron-catalyzed multicomponent radical cascades–cross-couplings

Lei Liu[1]†, Maria Camila Aguilera[2]†, Wes Lee[1], Cassandra R. Youshaw[1], Michael L. Neidig[2]*, Osvaldo Gutierrez[1,3]*

Transition metal–catalyzed cross-coupling reactions are some of the most widely used methods in chemical synthesis. However, despite notable advantages of iron (Fe) as a potentially cheaper, more abundant, and less toxic transition metal catalyst, its practical application in multicomponent cross-couplings remains largely unsuccessful. We demonstrate 1,2-bis(dicyclohexylphosphino)ethane Fe–catalyzed coupling of α-boryl radicals (generated from selective radical addition to vinyl boronates) with Grignard reagents. Then, we extended the scope of these radical cascades by developing a general and broadly applicable Fe-catalyzed multicomponent annulation–cross-coupling protocol that engages a wide range of π-systems and permits the practical synthesis of cyclic fluorous compounds. Mechanistic studies are consistent with a bisarylated Fe(II) species being responsible for alkyl radical generation to initiate catalysis, while carbon-carbon bond formation proceeds between a monoarylated Fe(II) center and a transient alkyl radical.

Organoboron compounds are valuable and highly versatile reagents widely used in modern organic synthesis (*1*). In particular, the use of organoboron reagents in palladium-catalyzed Suzuki-Miyaura couplings is one of the top five most used reactions in drug discovery (*2*). Photoredox catalysis has further expanded the utility of alkyl organoboron compounds as versatile radical precursors for numerous transformations (*3*). More recently, vinyl organoboron reagents have been used as effective lynchpins in three-component nickel- and metallaphotoredox-catalyzed cross-coupling reactions, leading to alkyl boryl scaffolds primed for further functionalization (*4*–*9*). Despite these efforts, the equivalent iron (Fe)–catalyzed transformation remains highly desirable in pharmaceutical research because of Fe's low cost, abundance, and potential for distinct and complementary modes of reactivity.

Fe-catalyzed cross-couplings have enabled the union of diverse carbon (C)–centered radicals and organometallic partners (Fig. 1A). However, although organoboron reagents have found utility in Fe-catalyzed two-component cross-couplings (Suzuki-Miyaura), application in multicomponent cross-couplings remains an elusive transformation (Fig. 1B) (*10*–*15*). Here, we report the successful realization of Fe-catalyzed cross-coupling of α-boryl radicals (generated from selective radical addition to vinyl boronates) with Grignard reagents to form dicarbofunctionalized compounds (Fig. 1C). Furthermore, to address a long-standing challenge in multicomponent cross-couplings, we report a general Fe-catalyzed multicomponent annulation–cross-coupling (MAC) protocol that facilitates the practical synthesis of tetrafluoroethylene-containing carbocycles and derivatives, which were previously difficult to make (*16*). Last, spectroscopy experiments [in situ Mössbauer, electron paramagnetic resonance (EPR), and x-ray crystallography], density functional theory (DFT), and radical probes shed light on the mechanism of this transformation.

[1]Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, USA. [2]Department of Chemistry, University of Rochester, Rochester, NY 14627, USA. [3]Department of Chemistry, Texas A&M University, College Station, TX 77843, USA.
*Corresponding author. Email: og.labs@tamu.edu (O.G.); michael.neidig@rochester.edu (M.L.N.)
†These authors contributed equally to this work.

Development of a three-component coupling reaction

On the basis of our recently reported studies on Fe-catalyzed radical cross-couplings (*17*–*21*), we hypothesized that the electron-deficient nature of vinyl boronates and the rapid kinetics observed in the Fe-catalyzed Kumada cross-couplings could be coupled to engage transient α-boryl radicals in selective three-component radical cross-couplings (*22*–*24*). In this vein, we first tested the proposed three-component radical cross-coupling using a sterically hindered alkyl halide under the slow addition of aryl Grignard nucleophile to avoid competing two-component cross-coupling and biaryl formation. After extensive experimentation, we identified $FeCl_3$ (10 mol %) in combination with 1,2-bis(dicyclohexylphosphino)ethane **L1** (20 mol %) as an effective catalytic system that engages α-boryl radicals, presumably through regioselective Giese addition of *tert*-butyl radical to vinyl boronate **2a**, to form the desired product **4a** in 90% yield (tables S1 and S2). In contrast to existing nickel systems, the reaction proceeded in <1 hour at low temperatures. Precatalysts with weakly coordinating triflate or acetate groups diminished the efficiency of the system, whereas halides or acetylacetonate counterions had minor effect on the yields (table S2, entries 1 to 6). Furthermore, 1,2-bis(dicyclohexylphosphino)ethane **L1** was a distinctly effective ligand (table S2, entries 7 to 13). Control experiments demonstrated that both the ligand and Fe salt are crucial for the reaction to proceed (table S2, entries 17 to 19). Both alkyl iodides and bromides provided the desired product under these conditions (~90% yield), although diminished yields were observed with alkyl chlorides (table S2, entries 20 and 21).

Encouraged by these findings, we turned our attention to studying the generality of the three-component radical cascade transformation. As shown in Fig. 2, we observed a wide range of organomagnesium compounds to be suitable cross-coupling partners with alkyl α-boryl radicals. In particular, difunctionalization of vinyl boronates proceeded with good yields and excellent regioselectivity with mono- and disubstituted aryl Grignard nucleophiles that varied in electron density at the para- and meta positions (**4a** to **4q**) as well as (hetero)aryl nucleophiles (**4r**). Furthermore, in contrast to nickel and metallaphotoredox catalytic systems, this protocol is compatible with alkenyl and alkynyl nucleophiles, albeit with lower yields observed for the latter (**4s** to **4w**). However, ortho-substituted aryl Grignard reagents were less efficient, presumably because of increased steric demand (**4y**). Having established the reactivity with sp- and sp^2-hybridized Grignard nucleophiles, we next probed the alkyl halide scope (Fig. 2, bottom). A range of tertiary acyclic and cyclic aliphatic electrophiles afforded the desired products with good yields (**4a′** to **4f′**). Secondary alkyl halides (**4g′**) and tertiary alkyl halides bearing aryl or heteroatom substituents also formed the desired products (**4f′** and **4h′**). Last, we identified tertiary α-bromo esters as competent substrates, leading to **4i′** bearing both an ester and alkyl boron as versatile synthetic handles for further diversification. Despite substantial advances in the synthesis of organofluorinated compounds (*25*), selective and catalytic $C(sp^3)$–CF_2R bond formation remains challenging (*26*) and is exceedingly rare in Fe-catalyzed cross-couplings (*27*–*31*). Seeking to expand the alkyl radical scope, we investigated whether this protocol could provide direct access to versatile fluorinated alkyl boron compounds. As shown in Fig. 2, bottom, a wide range of fluoroalkyl radical precursors—including those containing alkyl, ester, silyl, heteroaryl, phenoxy, perfluoroalkyl, and protected aldehydes—proved compatible partners, which led to the desired 1,2-alkylfluorinated-aryl organoboron products (**4j′** to **4p′**) in good to excellent yield. The distinctive properties of C–F bonds (*32*–*35*), versatility of the alkyl boron bond, and practicality of this method are anticipated to provide rapid access to valuable fluoroalkyl boron building blocks for synthetic applications.

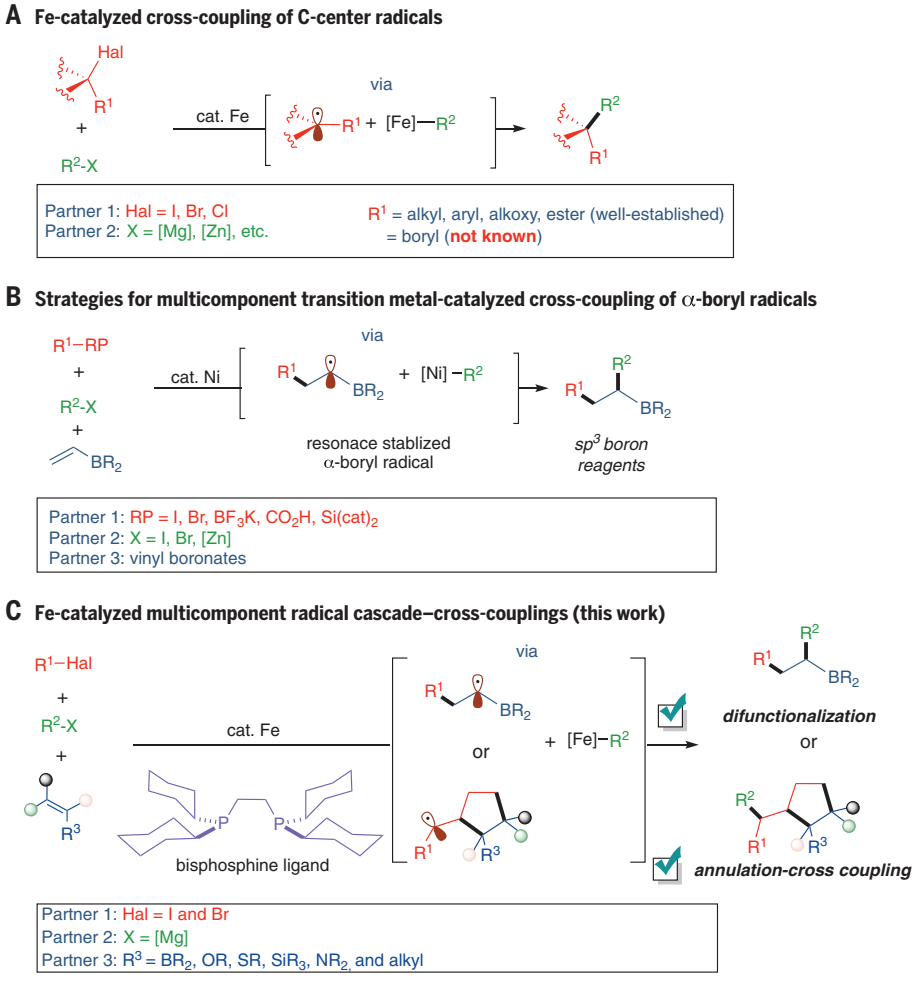


Fig. 1. Metal-catalyzed cross-coupling of α-boryl radicals. (**A**) Established methods for Fe-catalyzed C–C cross-coupling with alkyl radicals. (**B**) Current strategies for three-component trapping of α-boryl radicals. (**C**) Our report on the use of bisphosphine-iron complexes to promote radical cascade–cross-coupling reactions.

Extension to MACs

With the aim of expanding the applications of multicomponent radical cross-couplings, we turned our attention to alkyl halides with pendant alkenes as bifunctional coupling partners (*36*). Despite the utility of radical-based cyclization cascades and transition metal–catalyzed intramolecular cyclization-arylation in organic synthesis (*37*), the analogous intermolecular three-component radical cycloaddition-arylations have proven unexpectedly elusive. Furthermore, although incorporating CF_3, C_2F_5, and perfluoroalkyl alkyl groups in pharmaceutical research is common, practical and general synthetic methods for incorporation of the tetrafluoroethylene (-CF_2–CF_2-) moiety into cyclic compounds remains a challenge (*16*, *38*). We hypothesized that 4-bromo,3,3,4,4-tetrafluoro-1-butene **1s′,** which is commercially available and safe to handle, could serve as a general and practical lynchpin for the construction of tetrafluoroethylene-containing carbocycles that were previously hard to make. Similar conditions to those used for the three-component radical cross-coupling led to the MAC product **5a**, albeit in low yields because of competitive formation of dicarbofunctionalization product **4s′** (Fig. 3A). After screening of conditions, we found that by lowering the iron concentration and changing the catalyst-to-ligand ratio, we could shut down the dicarbofunctionalization pathway and increase yield of the annulation product **5a**. Presumably, lower catalyst concentration allows for more efficient 5-*exo* cyclization, leading to **int-2**, which in turn can undergo radical cross-coupling to form the annulation product (Fig. 3B).

With optimized conditions in hand, we next explored the generality with respect to alkene (Fig. 3C). Overall, a broad range of olefinic partners (**2a** to **zj**) (table S5) were found to be competent partners that lead to tetrafluoroethylene-containing drug-like scaffolds in one synthetic step. In particular, as shown in Fig. 3C, in addition to boron-substituted alkenes (**2a** and **2b**), vinyl silanes (**2c**), ketene acetals (**2d**), ethers (**2e** and **2v**), thioenols (**2f**), and enamines (**2g** and **2w**) were compatible. Furthermore, four-, five-, and six-membered (hetero)carbocycles bearing exo-cyclic alkenes (**2j** to **2r**) formed the corresponding spirocyclic compounds in good yields. We did not observe erosion of stereochemistry when using the enantiopure exo-methylene–containing pyrrolidine **2m** that led to the corresponding annulation-arylation product **5m**, as characterized by means of x-ray crystallography. In addition, this method allowed the gram-scale synthesis of spirocyclic compound **5p**, a derivative of sequosempervirin A (*39*), and fused bicyclic (hetero)cyclic structures, starting from the corresponding di- and trisubstituted cyclic (hetero)alkenes with good yields and modest to high regio- and diastereoselectivity (**2s** to **2x**). Acyclic olefins bearing alkyl chains with pendant functional groups—including aryl (**2z**), primary chloride (**2za**), alkene (**2zb**), alkyl boryl (**2zc**), unprotected alcohol (**2zd**) and amine (**2ze**), and alkyl ester (**2zf**)—were also competent partners. Tetrasubstituted alkenes and terminal alkynes also yielded the desired annulation-arylation products (**5zg**, **5zh**, **5zi**, and **5zj**), albeit in lower yields. To demonstrate potential for late-stage modification of bioactive compounds, we applied this protocol to natural products that bear alkene groups (**5zk** to **5zo**). We also explored the alkyl radical scope and found that other radical precursors could participate in the annulation (**5zp** to **5zu**).

We next explored the reaction scope of the nucleophile (Fig. 3D). In general, we found that para-substituted electron-rich and electron-poor aryl Grignard reagents formed the desired products (**5zv** to **5zza**). In addition, we found good yields across the board with aryl magnesium nucleophiles that bear electron-withdrawing groups (**5zza**, **5zzb**, **5zzf**, and **5zzg**); sterically hindered systems (**5zzc**, **5zzd, 5zze**, and **5zzi**), including those bearing C(sp^2)–Cl bonds for further functionalization; (hetero)aryls (**5zzk** and **5zzl**); and vinyl Grignard reagents (**5zzm** and **5zzn**).

Mechanistic studies

To elucidate the mechanism of this multicomponent iron-catalyzed cross-coupling, we used a combined spectroscopic, structural, computational, and organic synthetic approach. First, to provide direct insight into the iron intermediates involved in catalysis and enable the identification of the key iron species that initiates radical generation, we applied freeze-trapped 80 K ^{57}Fe Mössbauer and 10 K EPR spectroscopies, combined with single-crystal x-ray crystallography. For these spectroscopic studies, $^{57}FeBr_2$ was used as the starting Fe salt because it was a more accessible ^{57}Fe source, and it performed similarly to $FeCl_3$ under catalytic conditions. Two equivalents of **L1** (dcype) were

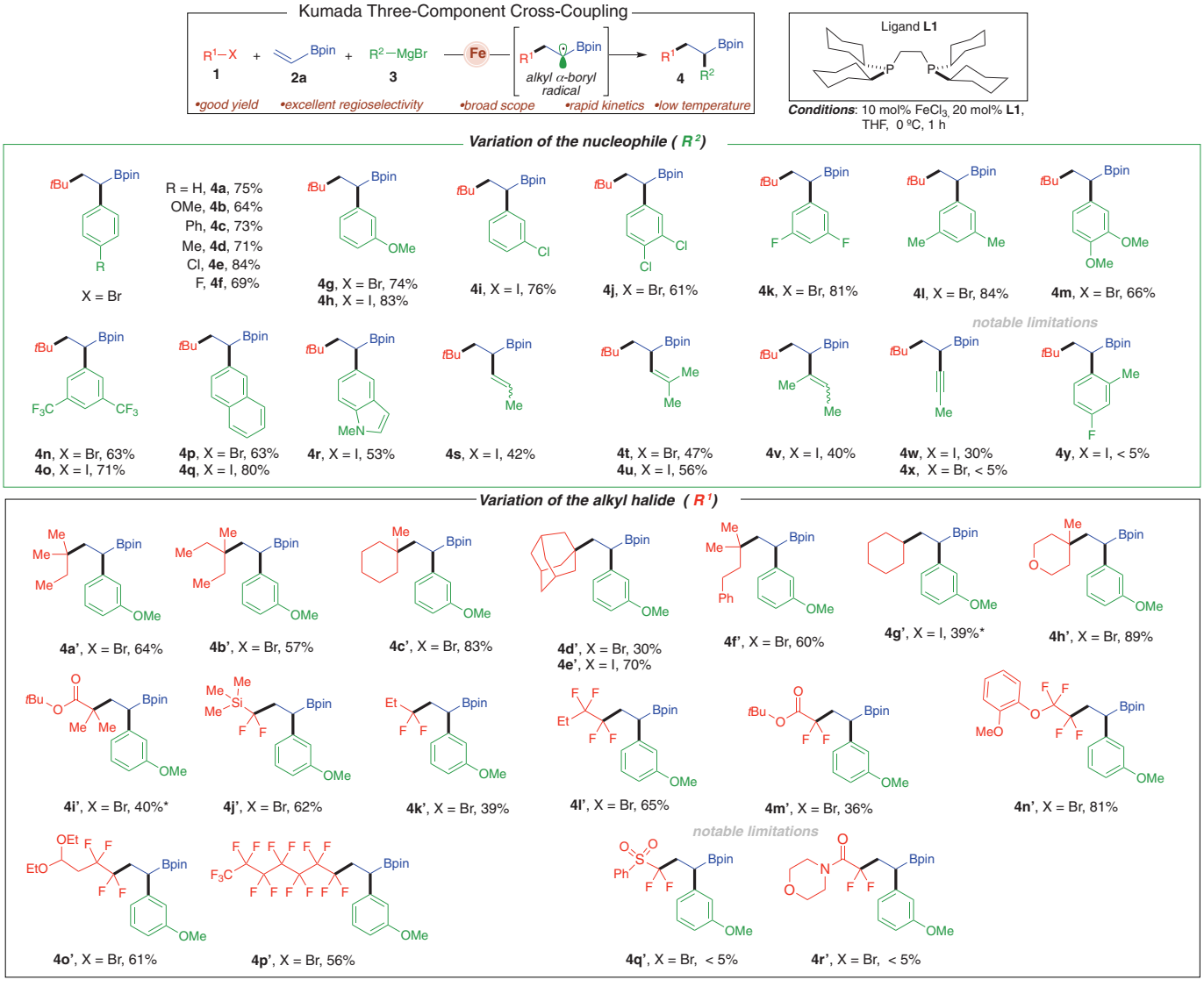


Fig. 2. Reaction scope of three-component dicarbofunctionalization of vinyl boronates by using bisphosphine-iron complexes as catalysts. Reactions were carried out on a 0.20-mmol scale at 0°C for 1 hour, performed with **1** (2.0 equiv), **2a** (1.0 equiv), Grignard reagent **3** (2.0 equiv) with THF (0.2 ml). Grignard reagent **3** was added dropwise by means of a syringe pump over 1 hour, isolated yield. *__1__ (5.0 equiv), Grignard **3** (5.5 equiv), and **3h**. Bpin, boronic acid pinacol ester; Me, methyl; Ph, phenyl; *t*Bu, *tert*-butyl.

combined with $^{57}FeBr_2$ in tetrahydrofuran (THF) for 10 min at 0°C, after which the solution was freeze-trapped in liquid N_2; an 80 K Mössbauer spectrum then revealed the formation of a single Fe species (δ = 0.72 mm/s and $|\Delta E_Q| = 3.21$ mm/s, where δ is the isomer shift and $|\Delta E_Q|$ is the quadrupole splitting) (table S6), which was later isolated and characterized by x-ray crystallography and Evans method nuclear magnetic resonance (NMR) as the high-spin Fe(II) dihalide complex Fe(dcype)Br_2 (**1**) [effective magnetic moment (μ_{eff}) = 5.2(1), δ = 0.73 mm/s, and $|\Delta E_Q| = 3.13$ mm/s for isolated material; numbers in parentheses indicate standard deviation] (Fig. 4A). The reaction of **1** with 1 equiv of the aryl Grignard reagent 3-methoxyphenyl magnesium bromide at 0°C resulted in the formation of a single Fe species with Mössbauer parameters of δ = 0.51 mm/s and $|\Delta E_Q| = 2.49$ mm/s, which is consistent with previously reported monoarylated Fe(II) bisphosphine complexes (Fig. 4A) (*40*, *41*). X-ray crystallography combined with Evans method NMR confirmed this species as the distorted tetrahedral, monoaryl high-spin Fe(II) complex Fe(dcype)BrAr (Ar, 3-$MeOC_6H_4$) [μ_{eff} = 5.0(2), δ = 0.52 mm/s, and $|\Delta E_Q| = 2.45$ mm/s for isolated material] (**2**).

Addition of a second equivalent of aryl Grignard reagent led to a color change of the solution from pale to brilliant yellow over the course of 1 min, which further evolved to dark green over 5 min. At that point, freeze-trapped Mössbauer spectroscopy indicated the formation of two new iron species, **3** and **4**, corresponding to 50 and 10% of the total Fe concentration, respectively (Fig. 4A). At a shorter reaction time (1 min), only 25% of species **2** had been converted to **3** without any formation of **4**. Further experiments revealed that species **4** could be accessed in higher amounts if excess aryl Grignard reagent is added (fig. S4A). This observation suggested the possible identity of **4** as a reduced Fe species, which was subsequently confirmed with x-ray crystallography to be (dcype)Fe{η^6-[3,3′-$(OMe)_2$-1,1′-$(C_6H_4)_2$]} (Fig. 4A and fig. S4B). The identification of this Fe(0) complex suggested that **3** was likely an Fe(II) bisaryl species formed before reductive elimination (δ = 0.23 mm/s and $|\Delta E_Q| =$ 4.35 mm/s). This assignment was subsequently

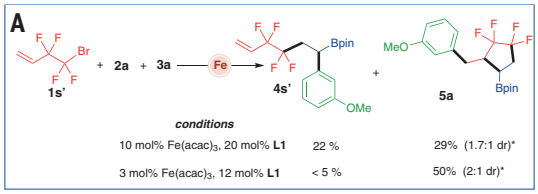
confirmed with crystallographic analysis as the bisarylated species Fe(dcype)Ar_2 (***3***), in which the large increase in the quadrupole splitting of ***3*** is consistent with the distorted square planar geometry of this Fe(II) complex. Although crystals of ***3*** repeatedly decayed over the course of the data collection, which limited the quality of the structure, the atomic assignments, overall geometry, connectivity, and identification of this complex are unambiguous. The product ***3*** was too thermally unstable for further characterization. However, literature precedent for $S = 1$ distorted square planar Fe phosphine compounds bearing two mesityl ligands, including their comparable Mössbauer parameters to ***3***, supports assignment of ***3*** as an intermediate-spin ($S = 1$) Fe(II) complex (*41*, *42*). This molecular geometry is atypical for bisaryl-Fe(II)-bisphosphines beyond mesityl complexes, which suggests an increased donor strength for **L1** that results from the cyclohexyl substituents. In addition, the identification of **4** indicates that ***3*** undergoes a two-electron reduction pathway similar to that previously observed in Fe(II)-SciOPP species (SciOPP, 1,2-{bis[3,5-di(*tert*-butyl)phenyl]phosphino}benzene) (*40*). Although all stoichiometric reactions were performed with an excess of **L1**, as prescribed for the catalytic reaction, all the iron species identified in situ contained only one bisphosphine per iron center.

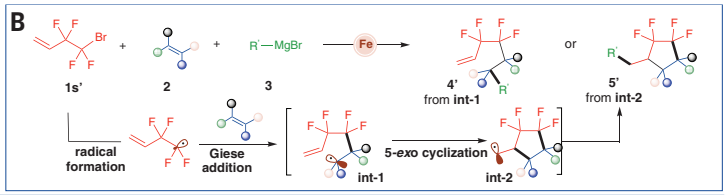
We proceeded to evaluate the reactivity of the identified, transmetalated Fe(II)-aryl-bisphosphine species toward electrophile (2-iodo-2-methylpropane) to determine their potential for radical initiation in catalysis. Pseudo–single-turnover studies for the reaction of monoarylated species **2** (generated in situ) with an excess of 2-iodo-2-methylpropane (20 equiv) in the presence of vinyl boronic acid pinacol ester indicated that **2** is reactive toward electrophile at an observed rate (k_{obs}) of ~0.04 min^{-1} (figs. S5 and S6), which resulted in three-component product formation. However, the observed rate of reaction is far too slow to be catalytically relevant, including the initial radical generation, considering the average turnover frequency during catalysis (~0.17 min^{-1}). Conversely, Mössbauer spectroscopic studies indicated that ***3*** is highly reactive toward electrophile. The reaction of a mixture of **2** and ***3*** (generated in situ after 1 min) with excess 2-iodo-2-methylpropane (20 equiv) leads to the complete consumption of species ***3*** within 25 s with concomitant generation of complex **1**, whereas the complex **2** in solution does not react with electrophile, which is consistent with its aforementioned slow reactivity (Fig. 4B). However, when a similar reaction is performed in the presence of vinyl boronic acid pinacol ester, species **2** is also consumed (fig. S7), which suggests a likely recombination of the secondary radical (formed after addition of the tertiary radical to the alkene) with **2** to generate product (further insights are available in the supplementary materials, materials and methods). Although complex **4** was also found to be highly reactive toward excess electrophile (20 equiv) (fig. S8), only undesired side products, including the two-component coupling of the electrophile and alkene, were observed to form, which is consistent with the lack of the aryl component required to form the three-component product. Furthermore, the reaction of the bisarylated species ***3*** with electrophile is faster (<25 s) than its transformation to complex **4** (>1 min); thus, **4** is unlikely to be generated in any substantial amount under catalytic conditions, which is consistent with no observation of **4** during catalysis. However, formation of **4** is more facile when only 1 equiv of **L1** is used (fig. S9), providing one possible role of excess ligand in achieving optimal yields in catalysis, although other roles of the excess phosphine, such as coordination to magnesium cations, cannot be excluded (*43*). Overall, these reactivity studies identify the distorted square planar bisarylated Fe(II) complex ***3*** as the key iron species responsible for the initial radical formation with the alkyl electrophile, which is required as the first step to initiate catalysis. Details about the nature of C–C bond formation were studied by means of DFT calculations.

We bolstered these stoichiometric studies with in situ iron speciation studies during catalysis: 80 K Mössbauer and 10 K EPR spectroscopy experiments were carried out on freeze-trapped reaction samples at various time points throughout the catalytic reaction (10, 30, and 50 min) (fig. S10). The distribution of species during catalysis consisted of ~48% **1** and ~52% **2** by Mössbauer spectroscopy; no EPR active species were observed. The presence of **2** in such large quantities during catalysis is consistent with the previous observation that **2** reacts slowly with electrophile. Complex ***3*** being undetectable during catalysis is also consistent with the prior observation of its rapid reactivity toward electrophile.

We next considered the success of **L1** as the supporting ligand because other related bisphosphines resulted in substantially decreased product formation. To understand this effect, we compared the distribution of species formed under catalytically relevant conditions with tetraethyl ligand **L3**. The Mössbauer spectrum of the freeze-trapped solution after reaction of $^{57}FeBr_2$ with 2 equiv of **L3** [1,2-bis(diethylphosphino)ethane (depe)] at 0°C shows the formation of a single Fe species ($\delta = 0.47$ mm/s and $|\Delta E_Q| = 1.51$ mm/s), which corresponds to the previously reported, distorted octahedral Fe(II) complex Fe(depe)$_2$$Br_2$ (**5**) (fig. S11) (*44*). Complex **5** also preferentially forms over the 1:1 **L3**:Fe complex, even when only 1 equiv of **L3** is used (fig. S12). Reaction of **5** with aryl Grignard (1 or 2 equiv) at 0°C led to a color change from brilliant yellow-green to orange within 5 min. Freeze-trapped Mössbauer spectroscopy indicated the formation of a single new Fe species **6** ($\delta = 0.30$ mm/s and $|\Delta E_Q| = 0.27$ mm/s) (fig. S13), even at extended reaction times, with the reduced isomer shift (relative to complex **5**) and small quadrupole splitting consistent with a distorted octahedral, arylated low-spin Fe(II) complex, as expected for the transmetalation of **5** with aryl Grignard reagent (*40*). Combined with additional reaction data at room temperature (fig. S13), this species is assigned to the monoarylated complex Fe(depe)$_2$BrAr. Reaction of **6** with 20 equiv of 2-iodo-2-methylpropane resulted in no consumption of this coordinatively saturated iron species, even at extended time points (20 min), which indicates limited or no reactivity toward electrophile (fig. S14). The observed slow transmetalation of **5** and lack of reactivity of **6** described above are consistent with their presence as the major Fe species in solution during catalysis (fig. S15). In the presence of excess nucleophile, a $S = 1/2$ Fe species could also be observed (fig. S16), which likely corresponds to a five-coordinate Fe(depe)$_2$X complex (X = Br or Ar), consistent with previously reported Fe(I) complexes formed as a result of the reaction of Fe-bisphosphines with aryl Grignard reagents (*44*, *45*). A coordinatively saturated, bisarylated Fe(depe)$_2$$Ar_2$ complex may also form in situ before reduction to Fe(I), although this species could not be unambiguously observed. The reduced Fe(I) species was found to be reactive toward electrophile in the presence of alkene (fig. S17), but the formation of only side products, including the two-component coupling of the electrophile and alkene, was observed, which is consistent with this Fe(I) species being unproductive for catalysis. Overall, these observations are consistent with the poor catalytic performance when using **L3** and highlight the importance of steric effects on the bisphosphine ligands in promoting the formation of coordinatively unsaturated Fe(II) species capable of initiating the radical formation as well as undergoing recombination to generate the desired product selectively.

Next, using vinyl cyclopropane as radical probe **2zp** (Fig. 5A), we observed the 1,5-dicarbofunctionalization product **4′** and no annulation product **5′** to be consistent with faster alkyl radical ring opening ($k \sim 10^7$ s^{-1}) than radical 5-*exo* cyclization ($k \sim 10^5$/s) and arylation ($k \sim 10^4$ s^{-1}) (*19*, *46*). In addition, consistent with the intermediacy of the alkyl radical, we also observed stereoconvergence in





C

5a, 41%, 2.1:1 dr

5b, 45%, 1.5:1 dr

5c, 52%, 1.1:1 dr

5d, 91%

5e, 81%, 1.2:1 dr

5f, 39%, 4.5:1 dr

5g, 33%, > 20:1 dr

5h, 68%, 10:1 dr (x-ray)

cis-5h (major)

5i, 54%, 15:1 dr

5j, 72%

5k, 83%

5l, 63%

5m, 61%

5n, 81%, 2.4:1 dr

5o, 93%

5p, 77%
gram scale [1.02 g]

5q, 47%, 7.1:1 dr

5r, 75%

5s, 48%, 7.7:1 dr

x-ray (major)

5t, 45%, 13.6:1.9:1 dr

5u, 59%, 4:1 dr

5v, 49%, 4.5:1 dr

5w, 74%, 1.4:1 dr

5x, 40%†, 10:1 dr

5y, 86%

5z, 59%, 8:1 dr

5za, 47%, 5.4: 1 dr

5zb, 44%, 7.3:1 dr

5zc, 28%, 6.7:1 dr

5zd, 57%‡, 3.7:1 dr

5ze, 49%‡, 3.9:1 dr

5zf, 67%, 1.4:1 dr

5zg, 66%†

5zh, 38%†

5zi, 30%†

5zj, 32%†

5zk, 40%‡, 8.4:7.9:1.3:1 dr

5zl, 51%, 38:1:1 dr

5zm, 57%‡, 2.1:1.7:1 dr

5zn, 88%, 4.9:1 dr

5zo, 84%, 4.9:2.5:1:1 dr

5zp, 75%, 6:1 dr

5zq, 31%, > 20:1 dr

5zr, 24%, > 20:1 dr

5zs, 30%, > 20:1 dr

5zt, 44%†,#, 10:1 dr

5zu, 56%†,#, 1.4:1 dr

D

5zv, 78%

5zw, 86%

5zx, 83%

5zy, 93%

x-ray

5zz, 80%

5zza, 87%

5zzb, 81%

5zzc, 79%

5zzd, 79%

5zze, 50%

5zzf, 59%

5zzg, 86%

5zzh, 82%

5zzi, 56%

5zzj, 87%

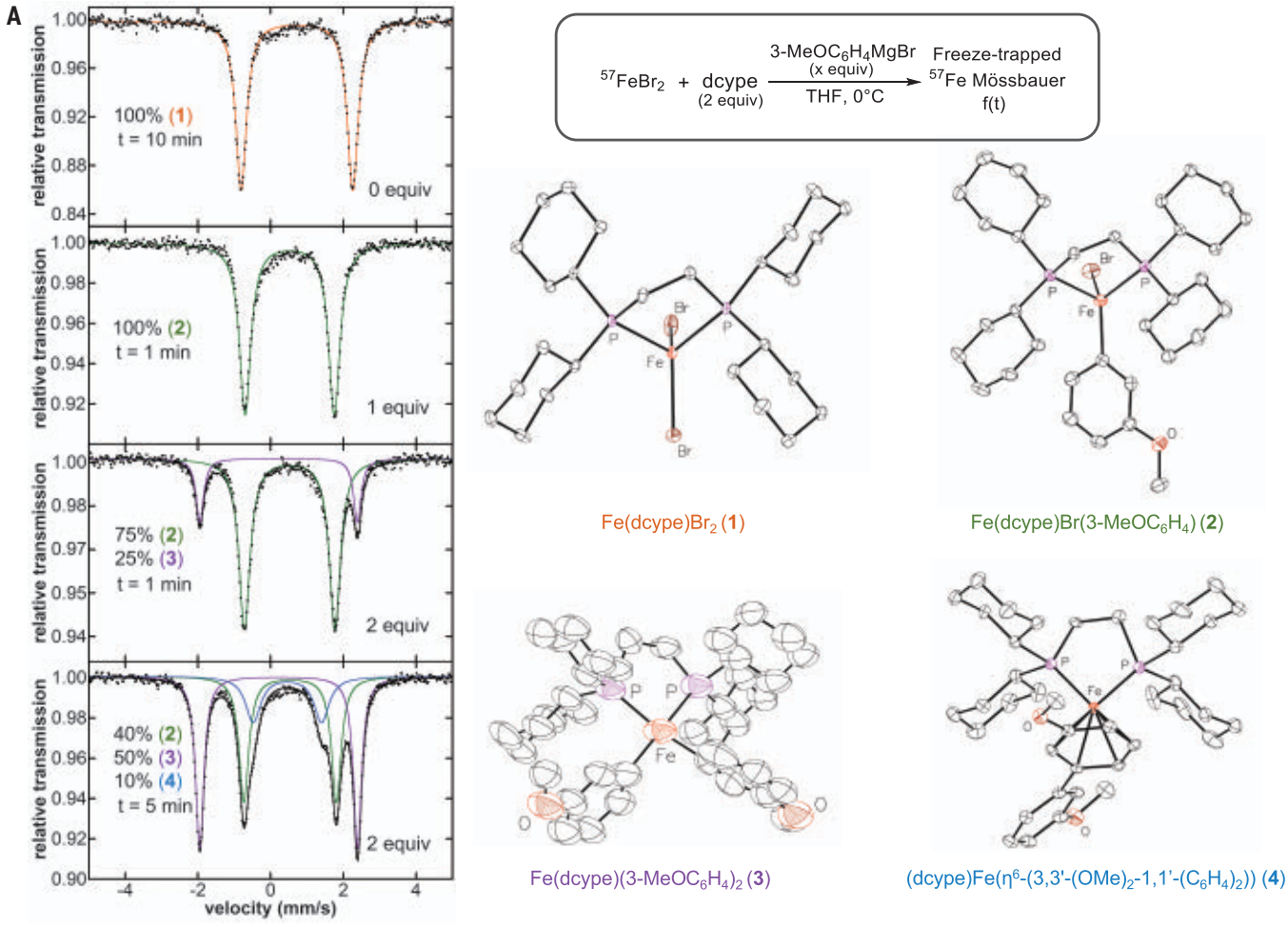
5zzk, 39%

5zzl, 42%

5zzm, 51%

5zzn, 39%

Fig. 3. Reaction scope of three-component annulation-arylation reaction sequence catalyzed by bisphosphine-iron catalysts. Reactions were carried out on a 0.20-mmol scale at 0°C for 1 hour, performed with **1** (1.0 equiv), **2** (1.5 equiv) Grignard reagent **3** (2.0 equiv), $Fe(acac)_3$ (3 mol %), and **L1** (12 mol %) with THF (0.2 ml). Grignard reagent **3** was added dropwise via a syringe pump over 1 hour. Reported yields and dr are from isolated yields. acac, acetylacetonate; Boc, *tert*-Butyloxycarbonyl; dr, diastereomeric ratio. *^{1}H NMR yield with CH_2Br_2 as internal standard. †Alkene or alkyne (14 equiv). ‡Grignard **3** (3.0 equiv). #No additional solvent.



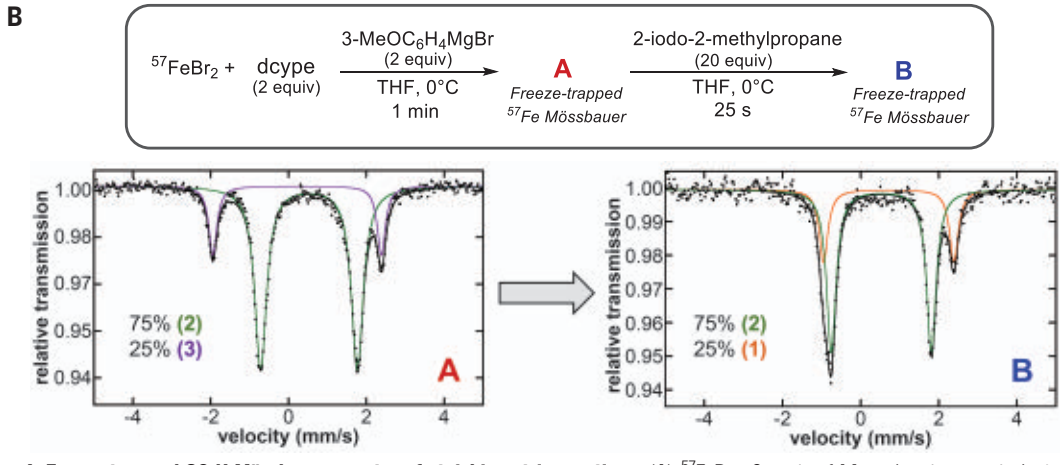


Fig. 4. Freeze-trapped 80 K Mössbauer spectra of stoichiometric reactions. (**A**) $^{57}FeBr_2$, 2 equiv of **L1**, and various equivalents of 3-$MeOC_6H_4MgBr$ (left). Combining SC-XRD and Mössbauer studies of crystalline material, the individual components were assigned as Fe(dcype)Br_2 (**1**) (orange), Fe(dcype)Br(3-$MeOC_4H_6$) (**2**) (green), Fe(dcype)(3-$MeOC_4H_6)_2$ (**3**) (purple), and (dcype)Fe{η^6-[3,3'-$(OMe)_2$-1,1'-$(C_4H_6)_2$]} (**4**) (blue). Thermal ellipsoids are shown at 50% probability. (**B**) The freeze-trapped 80 K Mössbauer spectrum of the in situ formed iron species upon reaction of $^{57}FeBr_2$ and 2 equiv of **L1**, with 2 equiv of 3-$MeOC_6H_4MgBr$ for 1 min (left) and following subsequent reaction with 2-iodo-2-methylpropane for 25 s (right).

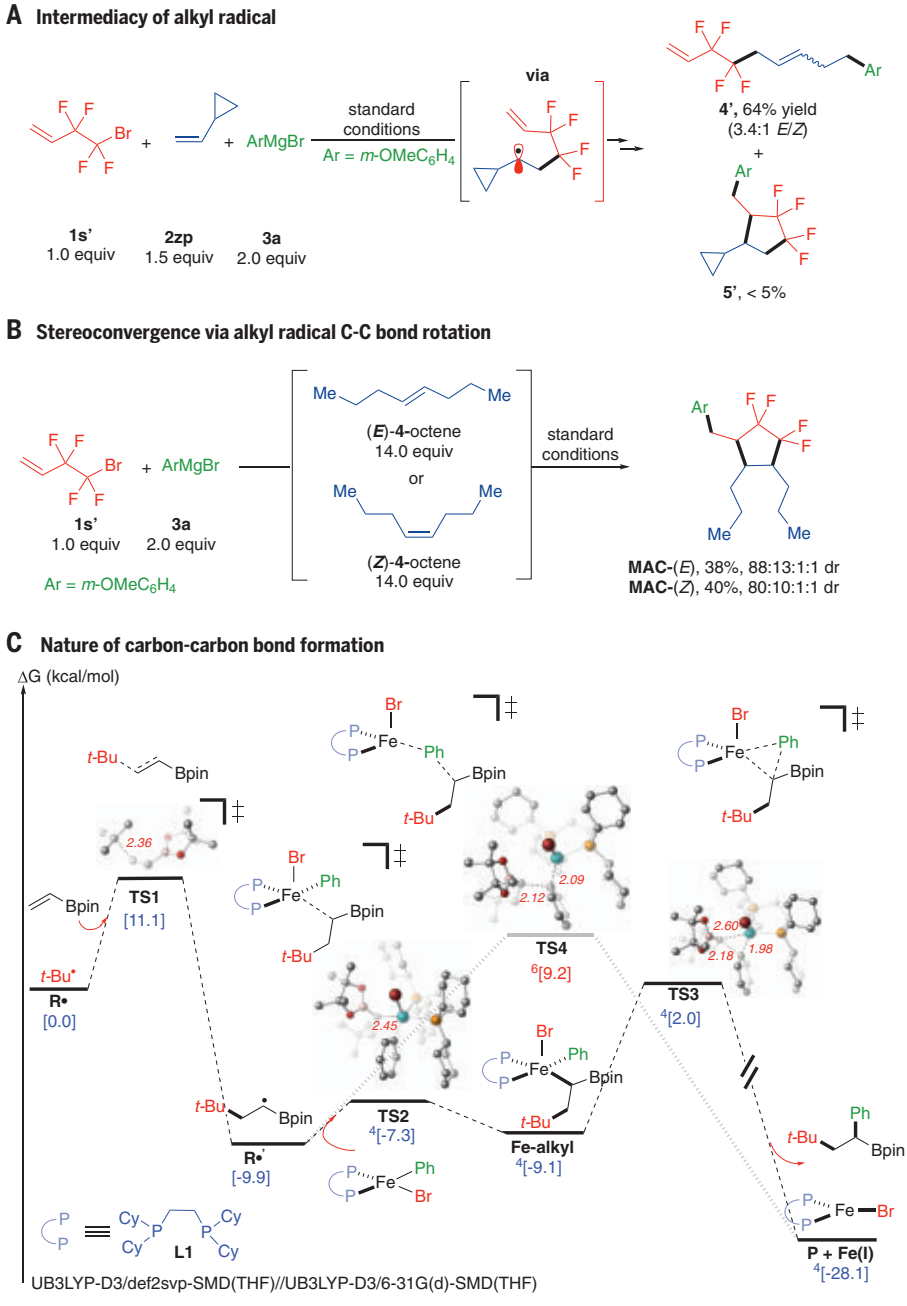


Fig. 5. Experimental and DFT calculations insights into the mechanism. (**A** and **B**) Standard conditions were carried out on a 0.20-mmol scale at 0°C; Grignard reagent **3a** was added dropwise by means of a syringe pump over 1 hour. Reported yields and dr are from isolated yields. (**C**) Computed lowest-energy pathway for the Fe-catalyzed multicomponent radical cascade cross-coupling reaction leading to dicarbofunctionalization of vinyl boronates. TS, transition state structure.

the three-component annulation, leading to the same product outcome by using either *E* or *Z* 4-octene, which suggests rapid equilibration (through sigma-bond rotation) before 5-*exo* radical cyclization-arylation (Fig. 5B). Last, to gain insight into the nature of the C–C bond formation, we turned to DFT calculations, which supported initial radical formation from an intermediate-spin distorted square planar bisarylated Fe(II) species through halogen abstraction (barrier, ~20 kcal/mol), which led to Fe(III) and tertiary radical (fig. S19). In turn, as shown in Fig. 5C, radical addition to vinyl boronate is predicted to be fast (barrier, ~11 kcal/mol) and irreversible, leading to α-boryl radical **R•'**. Then, this radical can rapidly and reversibly add (barrier, ~2.6 kcal/mol) to the corresponding high-spin distorted tetrahedral, monoarylated Fe(II) species, which leads to a distorted square pyramidal Fe(III)-alkyl intermediate. Subsequent irreversible reductive elimination by means of quartet spin state leads to the desired product and Fe(I) species, which can then restart the catalytic cycle (*17*, *47*). We also located the competing outer-sphere C–C forming transition state, but this pathway is less likely because of a much higher barrier (~19 kcal/mol) in comparison with the inner-sphere stepwise C–C bond formation (Fig. 5C, **TS4**). (Further mechanistic discussion and alternative pathways are provided in fig. S22.)

Overall, we anticipate that the method disclosed here will provide a practical and general route to functionalization of electron-rich and electron-deficient alkenes with various alkyl and hetero-substituents and application to late-stage functionalization of bioactive molecules.

REFERENCES AND NOTES

1. E. Fernández, A. Whiting, *Synthesis and Application of Organoboron Compounds* (Springer, 2015).
2. J. P. Wolfe, R. A. Singer, B. H. Yang, S. L. Buchwald, *J. Am. Chem. Soc.* **121**, 9550–9561 (1999).
3. J. C. Tellis, D. N. Primer, G. A. Molander, *Science* **345**, 433–436 (2014).
4. M. W. Campbell, J. S. Compton, C. B. Kelly, G. A. Molander, *J. Am. Chem. Soc.* **141**, 20069–20078 (2019).
5. A. García-Domínguez, R. Mondal, C. Nevado, *Angew. Chem. Int. Ed.* **58**, 12286–12290 (2019).
6. R. S. Mega, V. K. Duong, A. Noble, V. K. Aggarwal, *Angew. Chem. Int. Ed.* **59**, 4375–4379 (2020).
7. S.-Z. Sun, Y. Duan, R. S. Mega, R. J. Somerville, R. Martin, *Angew. Chem. Int. Ed.* **59**, 4370–4374 (2020).
8. M. Chierchia, P. Xu, G. J. Lovinger, J. P. Morken, *Angew. Chem. Int. Ed.* **58**, 14245–14249 (2019).
9. X.-X. Wang, X. Lu, S.-J. He, Y. Fu, *Chem. Sci.* **11**, 7950–7956 (2020).
10. T. Hatakeyama *et al.*, *J. Am. Chem. Soc.* **132**, 10674–10676 (2010).
11. H. M. O'Brien *et al.*, *Nat. Catal.* **1**, 429–437 (2018).
12. M. P. Crockett, A. S. Wong, B. Li, J. A. Byers, *Angew. Chem. Int. Ed.* **59**, 5392–5397 (2020).
13. R. B. Bedford *et al.*, *Organometallics* **33**, 5940–5943 (2014).
14. N. Kumar, R. R. Reddy, N. Eghbarieh, A. Masarwa, *Chem. Commun.* **56**, 13–25 (2020).
15. J. C. Lo *et al.*, *J. Am. Chem. Soc.* **139**, 2484–2503 (2017).
16. J. Václavík, I. Klimánková, A. Budinská, P. Beier, *Eur. J. Org. Chem.* **2018**, 3554–3593 (2018).
17. W. Lee, J. Zhou, O. Gutierrez, *J. Am. Chem. Soc.* **139**, 16126–16133 (2017).
18. L. Liu, W. Lee, M. Yuan, O. Gutierrez, *Comments Inorg. Chem.* **38**, 210–237 (2018).
19. L. Liu, W. Lee, J. Zhou, S. Bandyopadhyay, O. Gutierrez, *Tetrahedron* **75**, 129–136 (2019).
20. L. Liu *et al.*, *Chem. Sci.* **11**, 3146–3151 (2020).
21. L. Liu *et al.*, *Chem. Sci.* **11**, 8301–8305 (2020).
22. M. L. Neidig *et al.*, *Acc. Chem. Res.* **52**, 140–150 (2019).
23. J. D. Sears, P. G. N. Neate, M. L. Neidig, *J. Am. Chem. Soc.* **140**, 11872–11883 (2018).
24. D. G. Hall, Structure, properties, and preparation of boronic acid derivatives: in *Boronic acids* (Wiley, 2005), chap. 1, pp. 1–133.
25. M.-C. Belhomme, T. Besset, T. Poisson, X. Pannecoucke, *Chemistry* **21**, 12836–12865 (2015).
26. L. An, C. Xu, X. Zhang, *Nat. Commun.* **8**, 1460 (2017).
27. L. An, Y.-L. Xiao, S. Zhang, X. Zhang, *Angew. Chem. Int. Ed.* **57**, 6921–6925 (2018).
28. X. X. Rong, H.-Q. Pan, W. R. Dolbier Jr., B. E. Smart, *J. Am. Chem. Soc.* **116**, 4521–4522 (1994).
29. J. Derosa, O. Apolinar, T. Kang, V. T. Tran, K. M. Engle, *Chem. Sci.* **11**, 4287–4296 (2020).
30. D. V. Avila *et al.*, *J. Am. Chem. Soc.* **116**, 99–104 (1994).
31. X. Lin, F. Zheng, F.-L. Qing, *Organometallics* **31**, 1578–1582 (2012).
32. D. O'Hagan, *Chem. Soc. Rev.* **37**, 308–319 (2008).
33. K. Müller, C. Faeh, F. Diederich, *Science* **317**, 1881–1886 (2007).
34. S. Purser, P. R. Moore, S. Swallow, V. Gouverneur, *Chem. Soc. Rev.* **37**, 320–330 (2008).

35. M. A. Miller, E. M. Sletten, *ChemBioChem* **21**, 3451–3462 (2020).
36. J. Boström, D. G. Brown, R. J. Young, G. M. Keserü, *Nat. Rev. Drug Discov.* **17**, 709–727 (2018).
37. A. Studer, D. P. Curran, *Angew. Chem. Int. Ed.* **55**, 58–102 (2016).
38. L. Li *et al.*, *Angew. Chem. Int. Ed.* **56**, 9971–9975 (2017).
39. S. Maity, S. Ghosh, *Tetrahedron Lett.* **48**, 3355–3358 (2007).
40. S. L. Daifuku, J. L. Kneebone, B. E. R. Snyder, M. L. Neidig, *J. Am. Chem. Soc.* **137**, 11432–11444 (2015).
41. S. L. Daifuku, M. H. Al-Afyouni, B. E. R. Snyder, J. L. Kneebone, M. L. Neidig, *J. Am. Chem. Soc.* **136**, 9132–9143 (2014).
42. E. J. Hawrelak *et al.*, *Inorg. Chem.* **44**, 3103–3111 (2005).
43. A. M. Messinis *et al.*, *Nat. Catal.* **2**, 123–133 (2019).
44. D. J. Evans, R. A. Henderson, A. Hills, D. L. Hughes, K. E. Oglieve, *J. Chem. Soc., Dalton Trans.* (7): 1259–1265 (1992).
45. R. B. Bedford *et al.*, *Organometallics* **33**, 5767–5780 (2014).
46. M. Newcomb, in *Encyclopedia of Radicals in Chemistry, Biology and Materials*, C. Chatgilialoglu and A. Studer, Eds. (Wiley, 2012).
47. A. K. Sharma *et al.*, *J. Am. Chem. Soc.* **139**, 16117–16125 (2017).

ACKNOWLEDGMENTS

We thank P. Y. Zavalij (University of Maryland) and W. W. Brennessel (University of Rochester) for help with x-ray diffraction and Y. Li (University of Maryland) for help with the high-resolution mass spectrometry. **Funding:** Funding was provided by the National Institutes of Health (grant R35GM137797 to O.G. and R01GM111480 to M.L.N.) and the National Science Foundation (CAREER 1751568 to O.G.). **Author contributions:** L.L. developed the method and performed DFT calculations, and M.C.A. carried out the mechanistic studies. W.L. carried out DFT calculations and chromatography. C.R.Y. performed materials preparation and chromatography. O.G. and M.L.N. directed the investigations and wrote the manuscript with revisions provided by the other authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Crystallographic data are available free of charge from the Cambridge Crystallographic Data Center under reference nos. CCDC-2085294, CCDC-2085295, CCDC-2085297, CCDC-2085565, CCDC-2085566, CCDC-2085567, and CCDC-2085568. All other data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj6005
Materials and Methods
Supplementary Text
Figs. S1 to S22
Tables S1 to S13
References (*48*–*69*)

24 May 2021; accepted 9 August 2021
10.1126/science.abj6005

GUT INFLAMMATION

Identification of a choroid plexus vascular barrier closing during intestinal inflammation

Sara Carloni[1,2], Alice Bertocchi[2]†, Sara Mancinelli[1,2]†, Martina Bellini[1,2], Marco Erreni[1,2], Antonella Borreca[2,3], Daniele Braga[2], Silvia Giugliano[2], Alessandro M. Mozzarelli[1,2], Daria Manganaro[5], Daniel Fernandez Perez[5], Federico Colombo[2], Antonio Di Sabatino[4], Diego Pasini[5,6], Giuseppe Penna[2], Michela Matteoli[2,3], Simona Lodato[1,2]*, Maria Rescigno[1,2]*

Up to 40% of patients with inflammatory bowel disease present with psychosocial disturbances. We previously identified a gut vascular barrier that controls the dissemination of bacteria from the intestine to the liver. Here, we describe a vascular barrier in the brain choroid plexus (PVB) that is modulated in response to intestinal inflammation through bacteria-derived lipopolysaccharide. The inflammatory response induces PVB closure after gut vascular barrier opening by the up-regulation of the wingless-type, catenin-beta 1 (Wnt/β-catenin) signaling pathway, rendering it inaccessible to large molecules. In a model of genetically driven closure of choroid plexus endothelial cells, we observed a deficit in short-term memory and anxiety-like behavior, suggesting that PVB closure may correlate with mental deficits. Inflammatory bowel disease–related mental symptoms may thus be the consequence of a deregulated gut–brain vascular axis.

Around 30% of patients with inflammatory bowel disease (IBD) present with extraintestinal manifestations of morbidity (*1*). An axis connecting the gut, liver, and brain has long been postulated, but the players involved have not been clearly identified, particularly regarding the contribution of the vascular component. We recently demonstrated the existence of a gut vascular barrier (GVB) that resembles the blood–brain barrier (BBB) and directly links the gut with the liver (*2*–*5*). We showed an increase of GVB leakiness and dysfunction during *Salmonella* infection (*2*, *3*), diet-induced nonalcoholic steatohepatitis (*4*), or metastatic colorectal cancer (*5*), leading to an impaired gut–liver axis connection. The augmented GVB permeability and bacterial translocation into the liver and spleen were associated with increased detection of the endothelial cell–specific plasmalemma vesicle–associated protein (PV1) (*2*–*8*), a crucial molecule for the development of diaphragms in caveolae, trans-endothelial channels, and fenestrae (*9*–*11*). Detection of PV1 by MECA-32 antibody is associated with increased vascular permeability, presumably because of a conformational change or more accessibility of the recognized PV1 epitope (*12*). Even though the intestinal endothelium has long been recognized as a key player in gut inflammation, its role in IBD and its extraintestinal manifestations has yet to be defined (*13*).

In addition to causing severe physical symptoms, active IBD is associated with manifestation of anxiety and depression in up to 40% of patients (*14*), along with deterioration in cognitive functions (*15*, *16*). Similarly, the preclinical model of dextran sodium sulfate (DSS) colitis is associated with increased anxiety- and depression-like behaviors (*17*, *18*) and alterations in the limbic system (*19*). There is increasing evidence that circulating pro-inflammatory mediators in IBD patients may contribute to the progression of several central nervous system (CNS) disorders (*20*–*22*). However, it is so far not known whether these neurological diseases are a sequel, a prelude, or a combination of both to having a chronic disease.

The CNS has a complex vasculature system, which includes the BBB and the blood–cerebrospinal fluid (CSF) barrier (BCSFB). The BBB is composed of continuous, nonfenestrated vessels and acts as a tight regulator of CNS homeostasis in both physiological and pathological conditions (*23*–*25*). The BCSFB consists of choroid plexus (CP) fenestrated capillaries surrounded by stromal space and an outer cuboidal epithelial layer. The CP epithelial lining, which is continuous with the ependyma, has been considered a barrier because, unlike the endothelium, epithelial cells are sealed to one another by tight junctions (TJs) (*26*, *27*). In contrast to the BBB, under homeostatic conditions, the endothelium of the CP is permeable (*28*).

We hypothesized that mental disorders in IBD may be the consequence of vascular dysregulation. The opening of the GVB would allow the spreading of inflammation to distant sites, including the brain, which may result in mental deficits.

RESULTS

Intestinal inflammation associates with GVB impairment and recruitment of inflammatory cells to distal organs

We investigated the status of the GVB in patients with ulcerative colitis (UC) (Fig. 1A). Immunofluorescence analysis revealed a marked increase in PV1 staining in both ulcerated sites and neighboring tissues of UC specimens compared with healthy controls (Fig. 1B, quantified in Fig. 1C; $P = 0.0009, 0.0020$), suggesting that the GVB of patients with UC is compromised. After a computational analysis using

[1]Humanitas University, Department of Biomedical Sciences, 20072 Pieve Emanuele (Milan), Italy. [2]IRCCS Humanitas Research Hospital, 20089 Rozzano, Milan, Italy. [3]Institute of Neuroscience, National Research Council, c/o Humanitas Research Hospital, 20089 Rozzano, Milan, Italy. [4]Department of Internal Medicine, San Matteo Hospital Foundation, University of Pavia, 27100 Pavia, Italy. [5]Department of Experimental Oncology, European Institute of Oncology (IEO), IRCCS, 20139 Milan, Italy. [6]Department of Health Sciences, University of Milan, 20142 Milan, Italy.
***Corresponding author. Email: maria.rescigno@hunimed.eu (M.R.); simona.lodato@hunimed.eu (S.L.)**
†These authors contributed equally to this work.

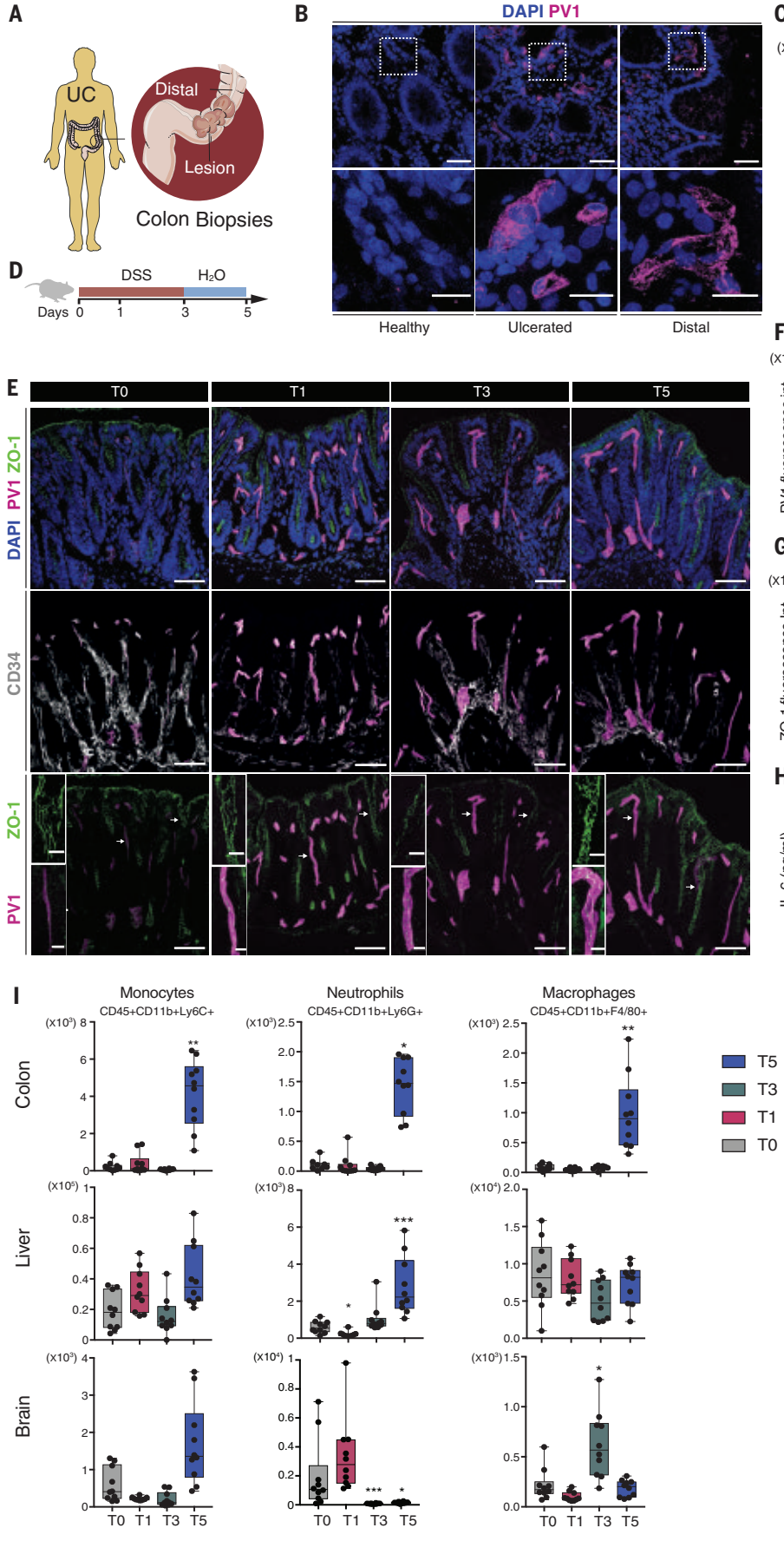


Fig. 1. UC leads to GVB impairment and fosters the recruitment of inflammatory cells to distal organs. (**A**) Colon biopsies of patients with UC were analyzed for GVB integrity. (**B** and **C**) Representative confocal images (B) and relative quantification expressed as fluorescence intensity (C) of PV1 (magenta) in human colon biopsy sections of healthy (H, $n = 6$), ulcerated (U, $n = 8$), and distal part with respect to the ulceration portion (D, $n = 4$) of the colon. (**D**) Timeline of DSS administration to C57BL/6 mice. (**E**) Representative confocal images showing PV1 (magenta), ZO-1 (green) expression, and CD34$^+$ (gray) blood vessels, cell nuclei [4′,6-diamidino-2-phenylindole (DAPI), blue] of colon cryosections. Scale bars, 50 μm; for zoom, 20 μm. (**F** and **G**) Fluorescence intensity quantification of PV1 (F) among CD34$^+$ vessels and ZO-1 (G). For PV1, $n = 5$ (T0 to T3) and $n = 4$ (T5). For ZO-1, $n = 5$ (T0, T3) and $n = 4$ (T1, T5). For (A) to (G), data from one experiment representative of three independent biological repetitions are presented. (**H**) Quantification of interleukin-6 cytokine (pg/ml) in plasma. $n = 12$ (T0), $n = 13$ (T1, T3), and $n = 8$ (T5). For (C), (F), (G), and (H), comparisons were performed with ordinary one-way ANOVA and Sidak's multiple-comparisons test. (**I**) Absolute numbers of monocytes (CD11b$^+$Ly6C$^+$), neutrophils (CD11b$^+$Ly6G$^+$), and macrophages (CD11b$^+$F4/80$^+$) gated on CD45$^+$ live cells. $n = 10$ (T0-T5). Comparisons were performed with nonparametric Kruskal-Wallis test and Dunn's multiple-comparisons test. For (H) and (I), data from three pooled experiments are represented using boxplots. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.

the online R2 database (*29*), we found that patients with UC displayed elevated lipopolysaccharide (LPS)–binding protein concentrations in sera (*30*) (fig. S1; $P < 0.0001$), suggesting that GVB disruption is accompanied by translocation of bacterial products.

Next, we evaluated whether a similar GVB derangement could be observed in a murine model of DSS experimental colitis (Fig. 1D). GVB morphology, colon length, and body weight were analyzed at two early time points [1 (T1) or 3 (T3) days after DSS treatment] and at the

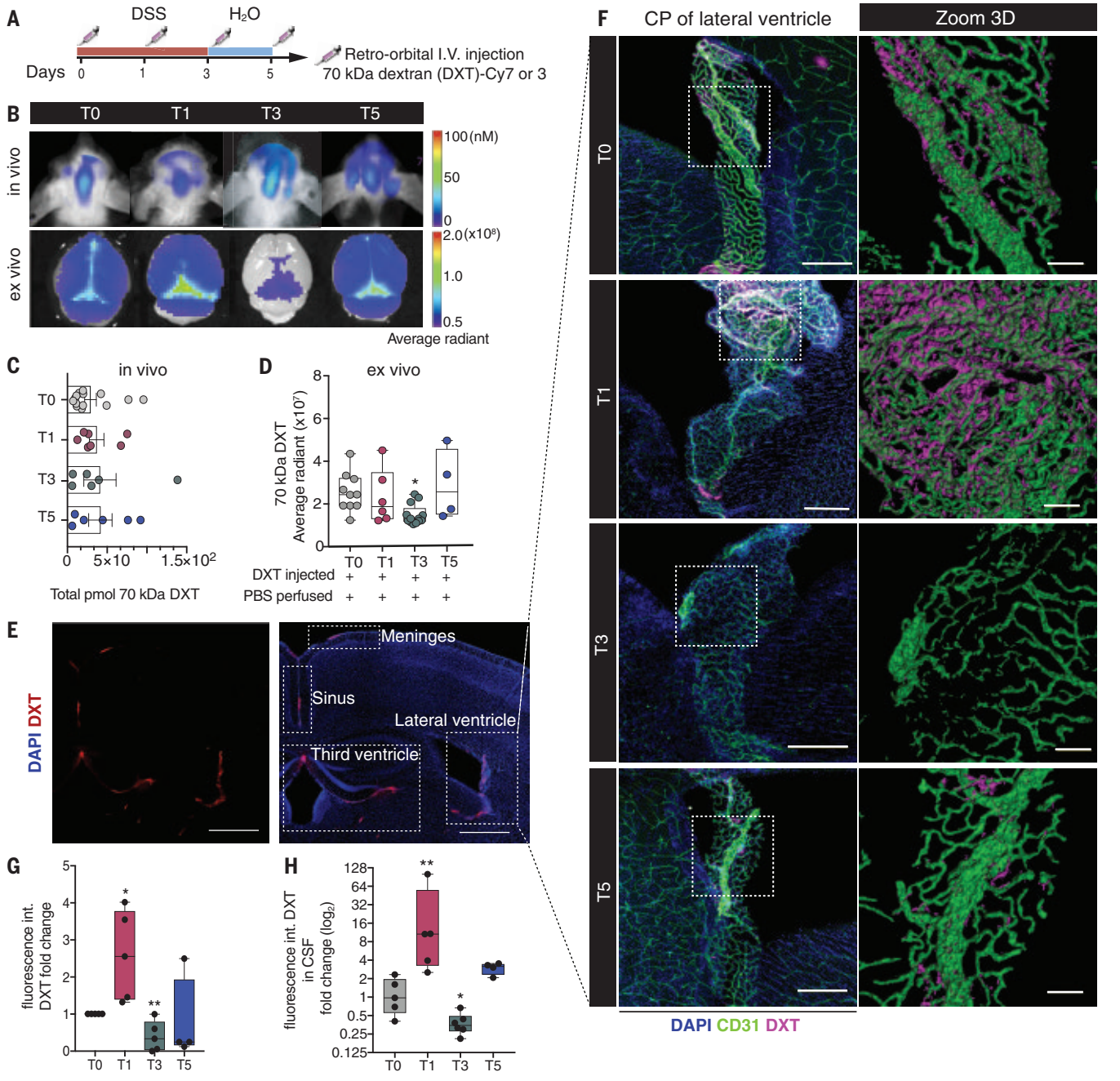


Fig. 2. Intestinal inflammation leads to alteration of CP permeability. (**A**) DSS administration to C57BL/6 mice and intravenous injection with 300 μg of 70-kDa Cy7/Cy3–conjugated dextran. (**B**) In vivo (top) and ex vivo (bottom) representative images of 70-kDa Cy7–conjugated dextran signal detected in the brain. (**C** and **D**) In vivo (C) and ex vivo (D) quantification of 70-kDa Cy7–conjugated dextran accumulation in the brain. For in vivo (C), $n = 9$ (T0, T3), $n = 6$ (T1), and $n = 4$ (T5). For ex vivo (D), $n = 10$ (T0), $n = 6$ (T1), $n = 13$ (T3), and $n = 4$ (T5). Comparisons between groups were performed with ordinary one-way ANOVA and Sidak's multiple-comparisons test. Data from two pooled independent experiments are represented using bar graphs with means ± SEM (C) and boxplots (D). (**E**) Representative confocal images of 70-kDa Cy3-dextran (red) distribution in the brain. Scale bars, 500 μm. (**F**) Representative confocal images showing 70-kDa Cy3-dextran (magenta) accumulation and CD31$^+$ (green) blood vessels in CP of brain lateral ventricles and 3D isosurface magnified reconstructions. Scale bars, 50 μm; for zoom, 20 μm. For (E) and (F), cell nuclei are stained with DAPI (blue). (**G**) Quantifications of extravasated 70-kDa Cy3-dextran. $n = 8$ (T0), $n = 5$ (T1, T3), and $n = 4$ (T5). (**H**) Quantifications of 70-kDa Cy3-dextran accumulation within CSF. $n = 5$ (T0, T1), $n = 6$ (T3), and $n = 4$ (T5). For (G) and (H), data from one experiment representative of two independent biological repetitions are presented as fold change. Data are represented using boxplots. Comparisons were performed with nonparametric Kruskal-Wallis test and Dunn's multiple-comparisons test. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$.

initiation of the recovery phase (2 days after stopping DSS treatment: T5). Already at T1, we observed a decrease of colon length and body weight, which significantly progressed through T3 ($P = 0.0007$, $P = 0.0003$) by T5 ($P = 0.0070$, $P < 0.0001$) (fig. S2, A and B). Concomitantly, we found a significant and persistent increase over time of PV1 on CD34$^+$ endothelial cells ($P = 0.0363$, $P = 0.0271$, $P = 0.0009$) (Fig. 1E, quantified in Fig. 1F), indicating an increment of vascular leakiness (*2*–*4*). Nevertheless, we did not observe any change in TJ-associated protein zonula occludens-1 (ZO-1) (fig. S2C) in intestinal endothelial cells (iECs), suggesting a specific remodeling of trans-cellular permeability of the intestinal vascular barrier.

To assess the integrity of the intestinal epithelial barrier, we evaluated the expression of ZO-1 in intestinal epithelial cells (iEpCs), whose decreased expression correlates with increased epithelial permeability (*4*). During DSS treatment, we observed a significant decrease in ZO-1 expression until T3 ($P = 0.0422$). Conversely, after 2 days of DSS deprivation (T5), ZO-1 expression increased to levels even higher than those of untreated mice ($P = 0.0021$), suggesting that iEpCs are the first to recover during

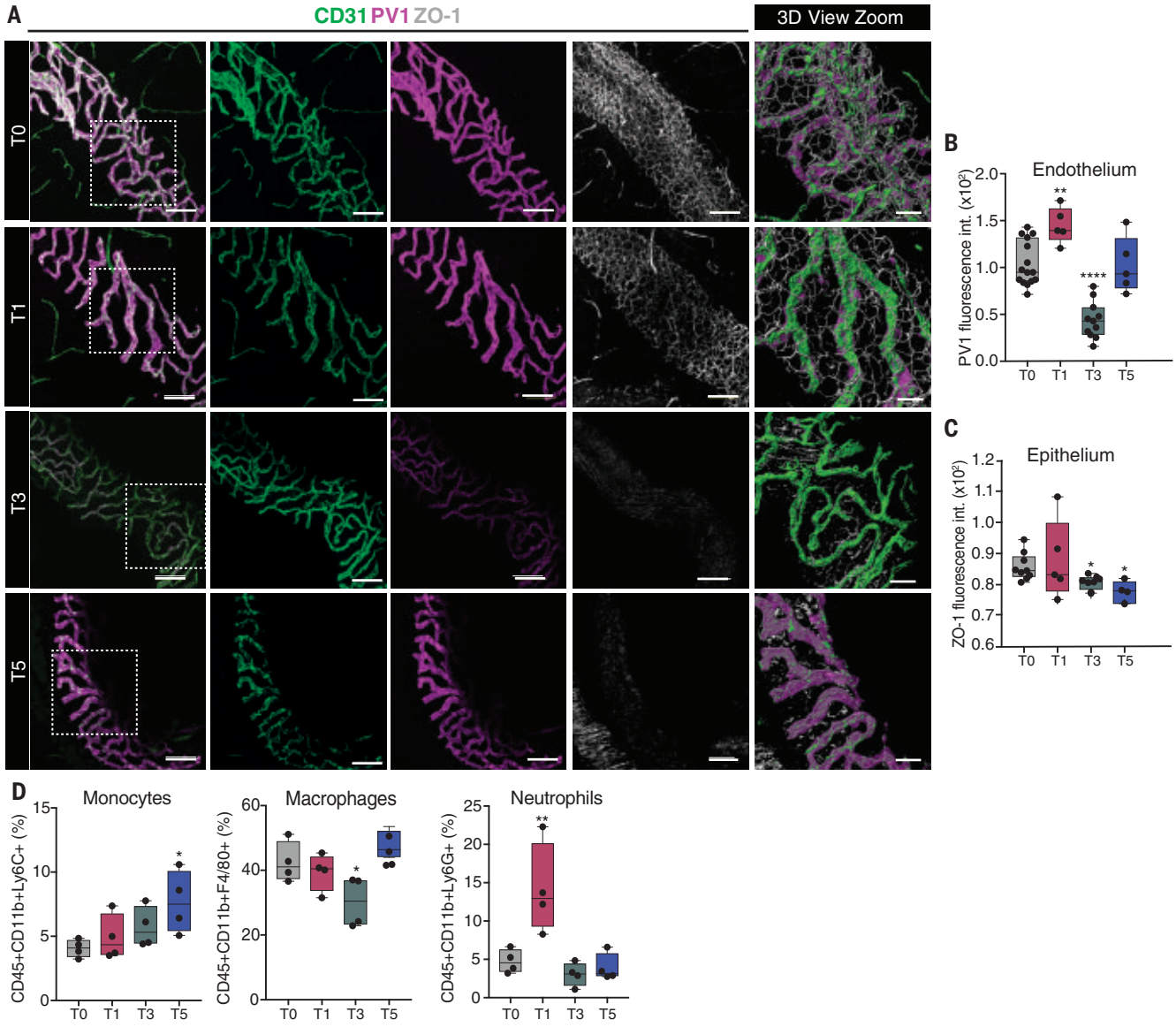


Fig. 3. Intestinal inflammation leads to PVB impairment and fosters the recruitment of inflammatory cells. (**A**) Representative confocal images showing PV1 (magenta) detection in CD31$^+$ (green) blood vessels and ZO-1 (gray) in brain lateral CP of untreated (T0) and DSS-treated mice (T1, T3, and T5). Scale bars, 50 μm. Last column: 3D isosurface reconstructions showing a qualitative representation of magnified acquired structure. Scale bars, 20 μm. Brain sections were at the same rostrocaudal level according to bregma coordinates (between B = 0.02 and B = −1.94). (**B** and **C**) Fluorescence intensity quantification of PV1 in CD31$^+$ blood endothelial cells (B) and ZO-1 outside CD31$^+$ blood vessels (C) expressed as arbitrary units (A.U.). For (B), $n = 15$ (T0), $n = 5$ (T1), $n = 11$ (T3), and $n = 5$ (T5). For (C), $n = 9$ (T0), $n = 5$ (T1), $n = 8$ (T3), and $n = 4$ (T5). Data from three independent biological repetitions are represented. (**D**) Flow cytometric analysis of inflammatory monocytes (CD11b$^+$Ly6C$^+$), neutrophils (CD11b$^+$Ly6G$^+$), and macrophages (CD11b$^+$F4/80$^+$) isolated from CP of untreated (T0) and DSS-treated mice (T1, T3, and T5). Graphs show the frequency (%) of each cell type calculated among CD45$^+$ live cells. Data from three pooled CPs for each point are represented. $n = 4$ (T0 to T5). All comparisons were performed with nonparametric Kruskal-Wallis test and Dunn's multiple-comparisons test, comparing each group (T1, T3, and T5) with T0. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$. For (B) to (D), data are represented using boxplots.

intestinal inflammation upon removal of the insult(s) (Fig. 1E, quantified in Fig. 1G).

However, despite the recovery of intestinal epithelial barrier integrity at T5, we found a progressive increase of the proinflammatory cytokine interleukin-6 in the plasma of treated mice compared with controls (Fig. 1H), concurrent with increased endothelial PV1 detection. This correlated with increased absolute number and frequency of circulating monocytes as early as T1 until T3, but there was no significant difference in circulating neutrophils and macrophages (fig. S2, D and E). Thus, we hypothesized that derangement of the GVB would lead to the recruitment of inflammatory leukocytes to the gut. To investigate this, we analyzed the inflammatory infiltrate associated with the colonic lamina propria over time. Mice were intracardially perfused with phosphate-buffered saline (PBS) before tissue isolation to exclude the contribution of circulating immune cells. Cytofluorimetric analysis of isolated immune cells from the lamina propria revealed an increase in both absolute number and frequency of inflammatory monocytes, neutrophils, and macrophages at T5 (Fig. 1I; fig. S3A; fig. S4A, colon; and fig. S4B).

Consistent with the induction of systemic inflammation, an increased absolute number and percentage of innate immune cells were also observed in the liver and the brain but with organ-specific fluctuations and temporal dynamics (Fig. 1I; fig. S3, B and C; fig. S4A, liver and brain; and fig. S4, C and D). The pattern of inflammatory monocytes was similar across the three organs, with an increase in absolute numbers at T5. Neutrophils decreased in the liver at T1 and, at the same time, were recruited into the brain. At T3, neutrophils were markedly reduced in the brain until T5, whereas in the liver and colon, they accumulated at T5. In the brain at T3, we observed an increase in the absolute number and frequency of macrophages (Fig. 1I and fig. S4, A and D) and resident microglia (fig. S5, A and B), the main surveillance immune cell population in the CNS (*31*, *32*).

Intestinal inflammation leads to an altered microglial phenotype

During neuroinflammation, scavenging resting microglial cells acquire an activated state that displays inflammatory and phagocytic features (*31*). To assess the microglial state during DSS-mediated intestinal inflammation, we performed morphological analysis in the hippocampal CA1 region. From T1 to T5, microglial cells showed a significant shortening ($P < 0.0001$) of branch length (fig. S5C, quantified in fig. S5D). At T3, ramified microglia displayed increased numbers of branches and junctions, suggesting an operating scavenging activity followed by an amoeboid-like phenotype, which was characterized by a deramification (decreased number of branches and junctions, and shortening of branch length) at T5 (fig. S5C, quantified in fig. S5D). Because microglia have an amoeboid-like morphology that is strictly associated with their reactive state (*33*), we analyzed microglia surface antigens. Polarization states of $CD11b^{+}CD45^{low}TMEM119^{+}$ microglia were evaluated by the expression of the major histocompatibility complex (MHC) II and CD86 surface markers. At T5, we observed a significant increase of MHC-II ($P = 0.0242$) and CD86 ($P = 0.0091$) on microglia of DSS-treated mice relative to control mice (fig. S5E). These data suggest that an acute intestinal inflammation quickly spreads extraintestinally, including to the brain, leading to a stepwise microglia activation most likely associated with CNS injury and neuroinflammation.

Gut–brain axis: Intestinal inflammation drives alteration of brain permeability

We thus surmised that intestinal inflammation and GVB disruption could have a direct impact on remodeling access to the brain. To examine this possibility, we analyzed the diffusion of high-molecular-weight (70-kDa) Cy7-conjugated dextran (Cy7-dextran) (Fig. 2A). We first analyzed dextran distribution in the brain circulation and found that as early as 1 hour (h) after injection, the maximum fluorescence intensity was homogeneously reached in all groups regardless of the treatment and time points (Fig. 2B in vivo, quantified in Fig. 2C). Mice were thus intracardially perfused with PBS at 1 h from dye injection to remove any circulating dextran and then analyzed ex vivo to evaluate the amount of extravasated dextran in the whole brain. We observed that at T0, untreated mice displayed brain accessibility to the 70-kDa dextran. By contrast, at T3, the access of 70-kDa dextran molecules to the brain was almost completely blocked, as indicated by the substantial reduction in intensity of the dextran-associated fluorescence and the dimension of the extravasation area (Fig. 2B ex vivo, quantified in Fig. 2D). This reduction at T3 was not observed in mice that were not perfused with PBS, confirming the comparable diffusion of circulating Cy7-dextran between untreated and T3 DSS-treated mice and further indicating that the difference found in ex vivo Cy7-dextran signal could be ascribed to the extravasated fraction (fig. S6). Brain closure was transient, because access to the 70-kDa dextran molecule was recovered after insult removal (T5) (Fig. 2B, quantified in Fig. 2D).

TJs of the BBB are remodeled upon intestinal inflammation

A role for the BBB in the diffusion of the 70-kDa dextran at steady state is unlikely because the BBB is permissive only to lipophilic molecules <500 Da (*3*, *34*). However, we could not exclude a concomitant or secondary remodeling of the BBB at the time of intestinal inflammation. We analyzed BBB integrity in multiple regions in the somatosensory cortical area (S1) and in the paraventricular nucleus of the thalamus (Pa) of DSS-treated mice (fig. S7, A and B). Analysis of the BBB-specific TJ protein claudin-5 (CLDN5) in both regions did not reveal any alteration with respect to control mice (fig. S7C, S1 and Pa, quantified in fig. S7, D and E). However, the S1 area in treated mice showed a decreased intensity of ZO-1 at T3. The staining of ZO-1 was altered in the Pa area from T1 until T3 of DSS treatment. In both regions, ZO-1 was restored to the levels of control mice at T5 (fig. S7F, S1 and Pa, quantified in fig. S7, G and H). Although these data suggest a region-specific remodeling of BBB adapter TJ proteins with a primary involvement of the Pa region, they fail to explain the access of the dye at T0 and even contrast with the reduction of brain accessibility observed at T3.

Alteration of the BCSFB within the CP during intestinal inflammation

To further address which brain regions were permissive to the high-molecular-weight dye and involved in the alteration of brain vascular permeability, mice were injected intravenously with 70-kDa Cy3-dextran and intracardially perfused with PBS. The accumulation of dextran was visible in the paraventricular region of the CP and meninges in controls, as previously described for lower-molecular-weight dyes (*35*) (Fig. 2E and fig. S8). However, the presence of the extravasated dextran was detected near the vascular unit of the CP, specifically in the stroma below the epithelial layer overlying the fenestrated endothelium (Fig. 2F, T0). After intracardiac perfusion with PBS, quantitative analysis of 70-kDa Cy3-dextran in the stroma of the CP highlighted a marked enrichment of extravasation at T1, suggesting an enhanced CP vascular permeability. This was followed by a complete absence of extravasated dextran at T3, indicating a robust vascular tightening (Fig. 2F, quantified in Fig. 2G) (movie S1). To evaluate whether the modulation of CP accessibility (endothelial and epithelial) also leads to changes in plasma and CSF exchange upon DSS treatment, we quantified the amount of the intravenously injected Cy3-dextran within the CSF. At the onset of intestinal inflammation (T1), we observed an increase of 70-kDa Cy3-dextran in the CSF, which was followed by a substantial reduction at T3 (Fig. 2H). Direct entry of dextran in the CSF after retro-orbital injections (*36*) is unlikely considering the observed dye reduction at T3. Thus, the most plausible route for plasma molecules as large as 70 kDa is to diffuse out of choroidal vessels into the stroma and,

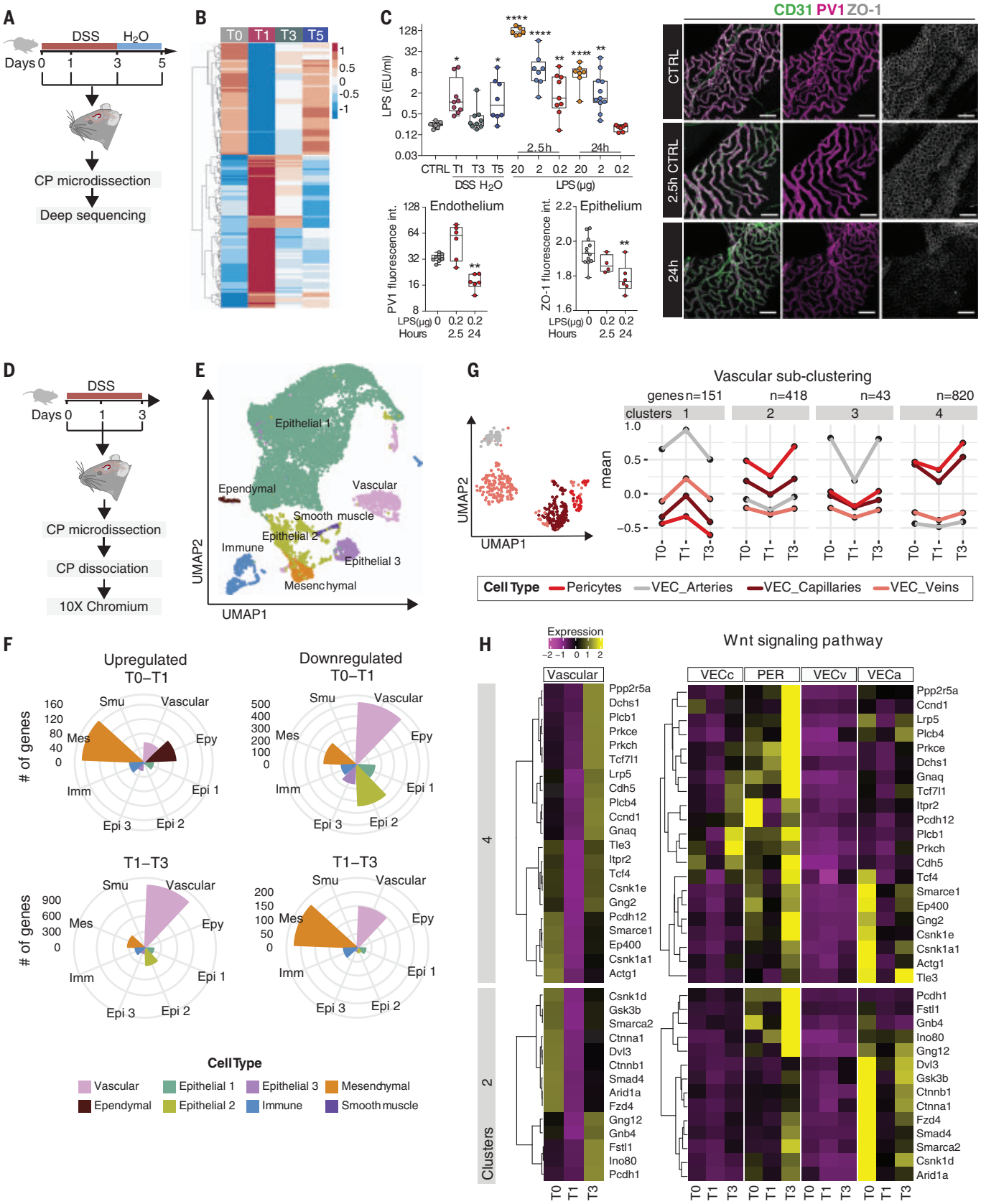


Fig. 4. Gene expression profiling of CP at tissue and single-cell resolution. (**A**) Time points chosen for RNA sequencing of CP. (**B**) Gene expression heatmap showing the mean of the 493 DEGs (FDR < 0.05 in the comparison T1 versus T0). The dendrogram highlights two gene expression clusters (225 up-regulated and 268 down-regulated genes). (**C**) (Top left) Endotoxin units per ml (EU/ml) of LPS in sera. *n* = 10 (T0), *n* = 9 (T1, T3), *n* = 8 (T5), and *n* = 6 and 7 for LPS 20 µg at 2.5 and

24 h, respectively; n = 8 and 11 for LPS 2 μg at 2.5 and 24 h, respectively; and n = 9 and 7 for LPS 0.2 μg at 2.5 and 24 h, respectively). (Right) Representative confocal images showing PV1 (magenta), CD31$^+$ (green), and ZO-1 (gray) expression in brain lateral CP. Scale bars, 50 μm. (Bottom left) Fluorescence intensity quantification of PV1 among CD31$^+$ vessels and ZO-1 outside CD31$^+$ blood vessels. For PV1, n = 8 and 6 for LPS 0 and 0.2 μg at 2.5 and 24, respectively. For ZO-1, n = 12 and 4 to 6 for LPS 0 and 0.2 μg at 2.5 and 24 h, respectively. Data are represented using boxplots. Comparisons were performed with nonparametric Kruskal-Wallis test. Brain sections were collected at the same rostrocaudal level according to bregma coordinates (between B = 0.02 and B = −1.94). (**D**) Time points of DSS treatment chosen for CP single-cell RNA sequencing. (**E**) Uniform manifold approximation and projection (UMAP) showing cell clustering. (**F**) Rose plots showing the number of DEGs at T0 versus T1 and at T1 versus T3 compared using the Wilcoxon statistic (average fold change = 0.25, adjusted P < 0.05). (**G**) UMAP of vascular cell subclustering (left) and line plots (right) showing the mean of DEGs. (**H**) Heatmap showing genes belonging to Wnt signaling pathway (clusters 2 and 4) deregulated across DSS treatment. *P < 0.05, **P < 0.01, ****P < 0.0001.

through TJ-expressing EpCs, reach the CSF. Conversely, during intestinal inflammation, the BCSFB within the CP is modified, with an initial increase of plasma molecule–CSF exchange (T1), followed by a rapid and prominent shutoff (T3), suggesting that the major gateway from systemic circulation to the brain is the CP, the accessibility of which is modulated during intestinal inflammation.

Intestinal inflammation leads to PVB closure

At T0, consistent with the homeostatic permeable nature of the CP vascular unit (*28*), we detected PV1 expression in CD31$^+$ ECs (movie S2 and Fig. 3A, quantified in Fig. 3B). At T1, we observed a significant increase in PV1 intensity on CD31$^+$ ECs (P = 0.0042), which was consistent with the increase in 70-kDa dextran accumulation within the stroma of CP and detection in the CSF. At T3, we observed a near-complete down-regulation of PV1, suggesting an endothelial sealing (Fig. 3A, quantified in Fig. 3B). Conversely, after 2 days of DSS deprivation (T5) (Fig. 1G), CP endothelial PV1 signal was normalized (Fig. 3A, quantified in Fig. 3B). Starting from T1, the signal of EpC-associated TJ ZO-1 was altered with respect to the control and at T3, when the fenestration of ECs was sealed, EpC- and EC-associated TJ ZO-1 was reduced (Fig. 3A, quantified in Fig. 3C, and fig. S9A). Despite the opening of the epithelial associated tight junction at T3, 70-kDa dextran was excluded from the CSF at T3. This suggests that during intestinal inflammation ECs control the passage of molecules from the blood to the stroma: the spike of dye at T1 and its absence at T3. The epithelium is permeable to 70-kDa molecules.

Hence, our data show the existence of an additional vascular barrier in the brain which we named the CP vascular barrier (PVB) that complements the epithelial barrier of the BCSFB and is revealed during inflammation to control extravasation of blood associated molecules into the stroma.

Inflammatory cells are recruited to the CP upon intestinal inflammation

Given the established role of CP as an immune gateway (*37*), in contrast to the BBB (*38*), we sought to further evaluate CP involvement in the recruitment of inflammatory cells. At T1, the dissected lateral CP was enriched in frequency and absolute number of neutrophils, whose presence was no longer evident by T3 (Fig. 3D and fig. S9, B and C). Monocytes peaked at T5, and the frequency of macrophages decreased at T3, inversely correlating with those in the brain parenchyma (Fig. 3D, fig. S9C, and Fig. 1I, brain). Neutrophils, by contrast, were low at steady state and increased in both the CP and brain tissue at T1 but were no longer able to enter the brain after PVB closure, suggesting a pivotal role of the PVB in limiting the propagation of proinflammatory leukocytes to the brain.

Transcriptional profiling of CP at tissue resolution

To evaluate, at the molecular level, potential pathways involved in changes of PVB permeability, we dissected lateral CP tissues at all time points (T0, T1, T3, and T5) and performed transcriptional profiling by deep RNA-sequencing (Fig. 4A). We used a regression model–based method for the classification of co-regulated genes based on their temporal expression profiles and the identification of differentially expressed genes (DEGs). The regression method identified 80 genes displaying a significant enrichment at T1 [false discovery rate (FDR) < 0.05] upon DSS treatment and gradually returning to their basal expression at T5 (cluster 1), whereas 31 genes showed the complementary profile, being significantly down-regulated at T1 (FDR < 0.05) and increasingly resuming expression at T5 (cluster 2, fig. S10A). Concordantly, differential expression analysis showed a robust gene expression regulation mainly at T1 compared with T0 (493 DEGs, of which 225 were up-regulated and 268 down-regulated), whereas only marginal differences were observed when comparing T5 with the baseline (17 DEGs) (Fig. 4B). We then performed gene set enrichment analysis (GSEA) to identify biological processes and/or pathways affected by acute DSS treatment. We found that gene ontology (GO) terms related to epithelium, endothelium, immune, and metabolism were significantly (FDR < 0.05) enriched in DEGs at T1 (fig. S10B and table S1). GSEA of DEGs at T1 revealed up-regulation of genes involved in the wingless-type, catenin-beta 1 (Wnt/β-catenin) signaling pathway (GO: 0060828, FDR = 0.04) (fig. S10B and table S1), a process regulating the maintenance of tissue homeostasis and the preservation of barrier properties in mature CNS vasculature (*39*) and in the GVB (*2*). In addition, we observed significant up-regulation of genes belonging to the inflammatory response (GO: 0006954, FDR = 0.002), immune response (GO: 0006955, FDR = 0.01), chemotaxis (GO: 0050920, FDR = 0.0001), and cytokine binding (GO: 0019955, FDR = 0.0003) (fig. S10B and table S1). Among differentially expressed genes, consistent with PV1 detection, transcriptional up-regulation of *Plvap* (the gene encoding the PV1 protein) was observed at T1, followed by down-regulation at T3 (fig. S10, C and D).

LPS modulates the PVB

Among the differentially expressed inflammatory pathways, we found activation of genes involved in the response to bacterial LPS (GO: 0032496, FDR = 0.02) in the CP tissue upon DSS treatment (fig. S10B and table S1). This led us to investigate whether, during DSS administration, LPS was detectable in the systemic circulation because this may parallel the presence of the LPS-binding protein in patients with UC (fig. S1). We found a spike of LPS at T1 (Fig. 4C, top panel), when the opening of both GVB and PVB occurred. After this first event, blood concentration of LPS decreased at T3. The reason for this is unknown, but an increased scavenging activity of circulating inflammatory cells might be hypothesized. To assess a direct involvement of LPS in the early alteration of CP accessibility, we injected LPS intraperitoneally. We then analyzed the concentration of circulating LPS 2.5 and 24 h after injection of 20, 2, and 0.2 μg LPS. We found that 0.2 μg of LPS after 2.5 and 24 h led to LPS blood concentrations similar to those measured in DSS-treated mice at T1 and T3, respectively; 2.5 h after LPS injection, we observed an up-regulation of PV1 intensity in ECs of the CP (Fig. 4C, bottom panel), suggesting that LPS could be a driver for the initial increased CP accessibility. By contrast, 24 h after LPS injection, we observed a significant decrease of endothelial PV1 staining (P = 0.0015) without any effect on EC-associated ZO-1, but there was a decrease of EpC-associated ZO-1, similarly to what was previously observed at T3 in DSS-treated mice (Fig. 4C, bottom panel, and fig. S11). These data suggest a correlation between circulating

LPS and the alteration of CP vasculature permeability.

The vascular compartment is involved in the modulation of CP accessibility

To study the remodeling of BCSFB and to understand the functions of individual cell types in the context of DSS treatment, we profiled the cellular components of the CP at single-cell resolution using 10× single-cell RNA sequencing (Fig. 4D). We profiled 17,705 CP cells from T0, T1, and T3 treated mice and, by unsupervised clustering of individual cell transcriptomes, we identified epithelial, mesenchymal, smooth muscle, immune, vascular, and ependymal cell types (*40*–*42*) (Fig. 4E; fig. S12, A to C; and table S2). Pairwise differential expression analysis across the different time points in each cell type confirmed a prominent transcriptional deregulation at T1 (Fig. 4F, fig. S12D, and tables S3 to S5), in agreement with our previous results on whole CP sequencing (Fig. 4B). In addition, we identified specific dynamics among the different cell types (Fig. 4F and fig. S12D). Despite gene expression changes at T1 being observed in all of the classes except smooth muscle cells (Fig. 4F and tables S3 to S5), a greater extent of transcriptional down-regulation was observed in vascular cells (508 genes of 561 deregulated in T0 versus T1 comparison, FDR < 0.05) (Fig. 4F and fig. S12D). At T3, a reverse pattern of regulation was evident for all populations as they returned to basal levels of expression (table S2), with the exception of vascular cells, which instead showed a sustained up-regulation of a large set of genes (1173 of 1321 deregulated genes in T1 versus T3 comparison, FDR < 0.05) (Fig. 4F; fig. S12, D and E; and tables S3 to S5), supporting the idea that the vascular cell population is critically involved in the changes observed in CP accessibility.

By population subclustering we identified pericytes (PER) and vascular ECs from capillaries, veins, and arteries (VEC-c, VEC-v, VEC-a) (Fig. 4G, fig. S12F, and table S6) (*43*, *44*). By far most of the genes altered upon DSS treatment were enriched and modulated in PER and VEC-c (Fig. 4G), with a high fraction of genes up-regulated at T3, supporting a potential direct role of PER and VEC-c in the CP vasculature closure observed at T3 (Fig. 4G and fig. S12G). In particular, functional annotation of gene clusters 2 and 4 showed a significant enrichment of pathways involved in vascular structure and physiology, including the Wnt/β-catenin signaling pathway (P00057; c4: $P < 0.01$; c2: $P < 0.01$), the epidermal growth factor receptor signaling pathway (P00018; c4: $P < 0.01$), and the platelet-derived growth factor signaling pathway (P00047; c4: $P < 0.01$; c2: $P < 0.01$) (fig. S12G and table S6) (*45*). We previously demonstrated that the Wnt/β-catenin pathway controls the permeability of the GVB and that β-catenin activation promotes GVB inaccessibility (*2*, *4*). We thus evaluated the expression dynamics of genes associated with Wnt signaling pathway and found a specific down-regulation at T1, followed by an up-regulation at T3 in PER, VEC-c, and VEC-a (Fig. 4H). In addition, we found a significant enrichment of inflammatory pathways in both clusters 2 and 4 (P00031; c4: $P < 0.01$; c2: $P < 0.05$), supporting the recent findings that capillary transcriptomes respond to inflammatory cues (*45*) (fig. S12G). The contribution of mesenchymal cell gene expression, which is strongly down-regulated at T1, remains to be elucidated (*46*).

Experimental colitis leads to an impairment of GVB and PVB and a sickness-like behavior

We analyzed whether opening or closure of the GVB and PVB was observed in the classical protocol of DSS-induced experimental colitis

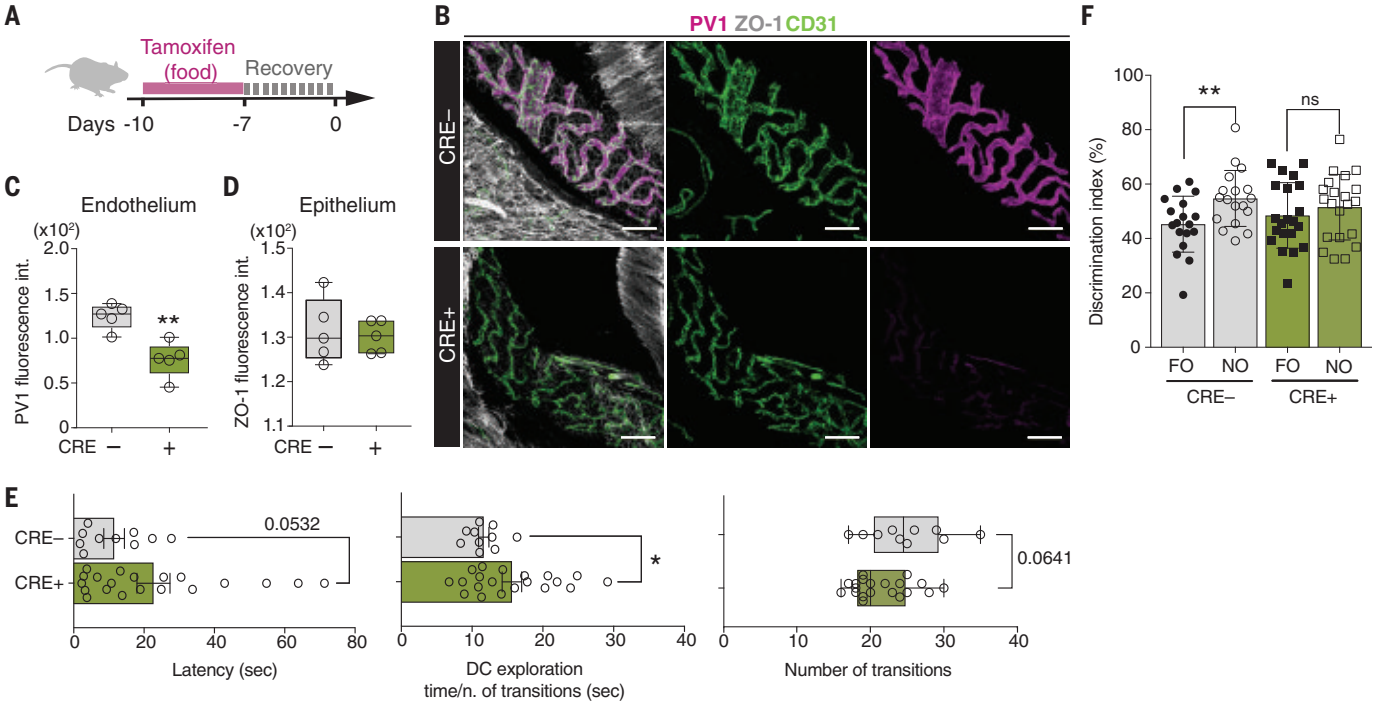


Fig. 5. Alteration of GVB-PVB vascular permeability leads to behavioral impairment. (**A**) Experimental design: β-catenin GOF (CRE^+) and $\beta\text{-}catenin^{lox(ex3)/lox(ex3)}$ (CRE^-) control mice were fed tamoxifen for 3 days followed by 1 week of recovery. (**B**) Representative confocal images showing PV1 (magenta) in $CD31^+$ (green) blood vessels and ZO-1 (gray) in CP. (**C** and **D**) Fluorescence intensity quantification of PV1 among the $CD31^+$ blood vessels (C) and ZO-1 outside $CD31^+$ blood vessels (D). Data from one experiment representative of three independent biological repetitions are presented using boxplots. $n = 5$ CRE^- and CRE^+. (**E**) Graphs showing latency (in seconds), time spent in a dark chamber (in seconds), and the number of transitions between the light and the dark chamber. $n = 10$ CRE^- and CRE^+. (**F**) Novel object recognition (NOR) test analyzed as discrimination index of time spent exploring each object [familiar object (FO) or novel object (NO)] over the total time spent exploring both objects. $n = 18$ CRE^-; $n = 21$ CRE^+. For (E) and (F), data from two pooled experiments are represented as means ± SEM using a scatter dot plot. For (C) to (F), comparisons between groups were performed using two-sided unpaired *t* test with Welch's correction. NS, $P > 0.05$, $^*P < 0.05$, $^{**}P < 0.01$.

(acute phase at 7 days of DSS; remission after 2 weeks of water) (fig. S13A) and if this correlated with behavioral disorders. We found a significant and persistent increase of PV1 detection on $CD34^+$ blood ECs ($P = 0.0002$, $P = 0.0003$) and TJ ZO-1 ($P = 0.0088$, $P = 0.0015$) on iECs in both the acute and the remission phase (fig. S13B, quantified in fig. S13, C and D) and a corresponding decrease of PV1 detection in CP ECs (on $CD31^+$ cells) only in the acute stage, without differences in ZO-1 (fig. S13E, quantified in fig. S13, F and G). Intestinal epithelial ZO-1 was up-regulated in the acute phase and normalized during remission (fig. S13H) without any modulation in the epithelium of the CP (fig. S13I). We next evaluated whether DSS treatment leads to behavioral disorders. We found that DSS-treated mice were characterized by hypomotility and lack of exploratory behavior. In the open-field assay, we detected a significant decrease in total distance traveled ($P = 0.0003$) and time spent in the periphery of the arena ($P = 0.0006$) in acutely DSS-treated mice compared with untreated controls. Also, DSS-treated mice displayed a reduced ambulatory rate at the periphery of the arena together with a reduced count of vertical (rearing up) movements (fig. S13J), all signs of hypomotility. In the dark-light test, acutely DSS-treated mice showed a reduced exploratory behavior of the bright area, as indicated by the increased latency to enter the dark chamber compared with controls, although they spent a comparable time exploring the dark box and performed a similar number of transitions between the light and the dark chambers (fig. S13K). All defects were reversed at remission (fig. S13, J and K). We could not determine whether hypomotility and reduced exploratory behavior were the result of CNS changes or pain caused by colitis. Analysis of microglial cells in CA1 brain region of DSS-treated mice revealed a decrease in the numbers of branches and junctions during the acute phase of the disease, which returned to normal after DSS cessation (fig. S14A, quantified in fig. S14B). These data suggest that intestinal inflammation promotes the reactive microglial shape and causes sickness in mice, thus hindering the investigation of possible effects consequent to the closure of the PVB.

Genetic inhibition of GVB-PVB permeability leads to behavioral impairment

To directly evaluate the importance of the GVB-PVB vascular axis inhibition independently of inflammation-driven effects or the overall discomfort induced by the DSS, we used a genetic mouse model in which the endothelial barriers are stabilized. We used a tamoxifen-inducible and endothelial-specific [Cadherin-5, Cdh5 (PAC)-CreERT2] β-catenin gain-of-function (GOF) mouse model in which β-catenin stabilization (mediated by the CRE expression) occurs only in endothelial-Cdh5–positive cells (*2*, *47*–*50*). We used this model because *Pvlap* deletion in mice or homozygous nonsense mutations of *Pvlap* leading to PV1 loss in patients with severe forms of protein-losing enteropathy result in disruption of fenestrated endothelial diaphragms and fatal sieving hypoproteinemia and enteropathy (*10*, *51*).

Because the CP single-cell database showed Cdh5 expression to be highly enriched in VEC-c and PER (fig. S15A), we used the GOF mouse model to test GVB and CP permeability and behavioral phenotypes (Fig. 5A). The increase of *β-catenin* expression analyzed by reverse transcriptase quantitative polymerase chain reaction in the gut and the brain of GOF CRE^+ mice with respect to CRE^- mice confirmed the activation of the gene in both the GVB and CP (fig. S15B). As expected from our previous studies, the increased β-catenin activation did not modify PV1 expression in $CD34^+$ ECs or EC- and/or EpC-associated ZO-1 in the colon (fig. S15, C to E) (*2*, *4*). By contrast, in the CP, β-catenin activation led to a decrease in PV1 detection on $CD31^+$ blood vessels and an increase in ZO-1 expression in ECs, confirming previous studies on partial CP EC reprogramming to a BBB-like state (*39*, *52*, *53*) (Fig. 5B, quantified in Fig. 5C, and fig. S15F) without any alteration of EpC-associated ZO-1 (Fig.5D).

We next examined the effects of β-catenin GOF-induced inhibition of GVB and PVB vascular permeability on the behavioral profile of the mice. Analysis of microglial cells in the CA1 brain region of CRE^+ mice did not reveal changes in morphology, confirming the lack of activation processes (fig. S15G). Also, consistent with a healthier profile of CRE^+ β-catenin GOF mice compared with DSS-treated mice, the general locomotor activity was not altered in CRE^+ mice (fig. S16A). In addition to the lack of locomotor defects, CRE^+ mice did not display depressive-like behaviors, as indicated by the forced swim test or the tail suspension test (fig. S16B). However, CRE^+ mice showed higher latency in entering the dark chamber, accompanied by a significant increase in the time spent in it ($P = 0.0169$), a sign of anxiety-like behavior, without a significant change in the number of transitions (Fig. 5E). CRE^+ mice also displayed a significant impairment in episodic memory ($P = 0.0094$), as indicated by the defective discrimination of a novel object relative to a familiar one in the novel object recognition test (Fig. 5F). Together, these data indicate that CRE^+ mice, in which the GVB-PVB axis is interrupted without concomitant inflammation, display impairment in episodic memory and some anxiety-like behavior. However, although we did not observe any modification in body weight, colon length, immune cell populations, or microbiota between CRE^+ and CRE^- mice (fig. S17), we cannot exclude that alterations of *β-catenin* in other ECs where CDH5 is active may occur in these mice and contribute to the observed phenotype.

Discussion

Our data indicate that the gut and the brain are connected by a vascular axis that is tightly regulated by the GVB and the PVB, respectively. The GVB connects our body with the external world, ensuring that nutrients and small molecules, but not bacteria or large bacterial molecules (such as LPS), gain access to the portal circulation and the liver (*2*–*4*, *54*). By contrast, the PVB (which is permissive at steady-state) ensures the connection between the brain and the rest of the body because it allows the passage of molecules as large as 70 kDa, which are filtered by epithelial cells and would be otherwise restricted by the BBB. During intestinal inflammation, the GVB is disrupted, which correlates with a state of bacterial product translocation and systemic inflammation, eventually leading to immune cell recruitment to the liver and propagation of inflammation to the brain. Whether the liver also plays an active role in propagating inflammation remains to be established. The presence of inflammatory and bacterial mediators of pathogen-associated molecular patterns such as LPS can be sensed by the cellular network of the CP capillaries and microglial scavenging, which in turn locally activate gene expression programs involved in the response to bacterial components. The circulation of gut-derived inflammatory mediators leads to temporary changes in the PVB and is likely to be the main driver of the activation of a brain protection program leading to CP shutdown.

Whereas in the gut, epithelial cells are at the forefront of changes in the intestinal environment, in the CP, ECs and PER are the first to sense signals coming from the systemic circulation to the brain, which are then likely transduced by the Wnt signaling pathway.

After an initial opening of the endothelial compartment of the PVB, a complete shutdown of the barrier was observed (fig. S18). We cannot exclude that modulations of PV1 expression, and/or changes in its conformation, could be involved in the alteration of the CP vascular compartment, including the endothelial caveolae structure (*55*), or that injury and repair mechanisms may have occurred in ECs in the time frame of the experiments and could not be detected.

Alterations of the BBB within the somatosensory cortical area and in the paraventricular nucleus of the thalamus are temporary and follow changes in the PVB, suggesting a secondary rather than a primary response, but this needs further investigation. In addition, it will be interesting to understand if during intestinal inflammation, the dural vessels could

also undergo alterations in permeability and if these changes could influence the response of the CP or the skull bone marrow (*56*). Thus, PVB regulation trigged by gut-derived inflammatory mediators may be a major contributor to the establishment of the gut–brain axis. Our data support the possibility that at least part of the behavioral and cognitive alterations that have been described in patients with IBD may result not from the enhanced inflammation, as generally hypothesized, but rather from the defense strategy activated by the organism to protect the brain from damage and guarantee its function. These data suggest that a deregulated GVB-PVB interaction could be a therapeutic target in several behavioral disorders.

REFERENCES AND NOTES

1. J. S. Levine, R. Burakoff, *Gastroenterol. Hepatol.* **7**, 235–241 (2011).
2. I. Spadoni *et al.*, *Science* **350**, 830–834 (2015).
3. I. Spadoni, G. Fornasa, M. Rescigno, *Nat. Rev. Immunol.* **17**, 761–773 (2017).
4. J. Mouries *et al.*, *J. Hepatol.* **71**, 1216–1228 (2019).
5. A. Bertocchi *et al.*, *Cancer Cell* **39**, 708–724.e11 (2021).
6. M. Sorribas *et al.*, *J. Hepatol.* **71**, 1126–1140 (2019).
7. Y. Li *et al.*, *Life Sci.* **261**, 118460 (2020).
8. F. Ciccia *et al.*, *Ann. Rheum. Dis.* **76**, 1123–1132 (2017).
9. B. J. Hamilton, D. Tse, R. V. Stan, *J. Cell. Mol. Med.* **23**, 920–933 (2019).
10. R. V. Stan *et al.*, *Dev. Cell* **23**, 1203–1218 (2012).
11. K. Auvinen *et al.*, *Sci. Rep.* **9**, 15698 (2019).
12. R. Hallmann, D. N. Mayer, E. L. Berg, R. Broermann, E. C. Butcher, *Dev. Dyn.* **202**, 325–332 (1995).
13. W. E. Cromer, J. M. Mathis, D. N. Granger, G. V. Chaitanya, J. S. Alexander, *World J. Gastroenterol.* **17**, 578–593 (2011).
14. G. Byrne *et al.*, *Can. J. Gastroenterol. Hepatol.* **2017**, 6496727 (2017).
15. D. Golan *et al.*, *Inflamm. Bowel Dis.* **22**, 364–371 (2016).
16. I. Tadin Hadjina *et al.*, *Sci. Rep.* **9**, 13740 (2019).
17. E. Dempsey, Á. Abautret-Daly, N. G. Docherty, C. Medina, A. Harkin, *Brain Behav. Immun.* **80**, 616–632 (2019).
18. K. D. Nyuyki, N. L. Cluny, M. G. Swain, K. A. Sharkey, Q. J. Pittman, *Front. Behav. Neurosci.* **12**, 58 (2018).
19. J. Do, J. Woo, *Clin. Psychopharmacol. Neurosci.* **16**, 422–433 (2018).
20. R. Neuendorf, A. Harding, N. Stello, D. Hanes, H. Wahbeh, *J. Psychosom. Res.* **87**, 70–80 (2016).
21. F. Marques *et al.*, *J. Cereb. Blood Flow Metab.* **29**, 921–932 (2009).
22. F. Marques *et al.*, *Neuroscience* **144**, 424–430 (2007).
23. E. A. Neuwelt *et al.*, *Nat. Rev. Neurosci.* **12**, 169–182 (2011).
24. R. Haddad-Tóvolli, N. R. V. Dragano, A. F. S. Ramalho, L. A. Velloso, *Front. Neurosci.* **11**, 224 (2017).
25. N. Weiss, F. Miller, S. Cazaubon, P. O. Couraud, *Biochim. Biophys. Acta* **1788**, 842–857 (2009).
26. C. E. Johanson, *J. Neurol. Neuromedicine* **3**, 39–58 (2018).
27. P. A. Johansson, *Front. Neurosci.* **8**, 340 (2014).
28. M. Schwartz, K. Baruch, *EMBO J.* **33**, 7–22 (2014).
29. J. Koster, "R2: Genomics analysis and visualization platform" (2021); http://r2.amc.nl.
30. O. P. Rojo *et al.*, *Inflamm. Bowel Dis.* **13**, 269–277 (2007).
31. S. A. Wolf, H. W. G. M. Boddeke, H. Kettenmann, *Annu. Rev. Physiol.* **79**, 619–643 (2017).
32. F. Filipello *et al.*, *Immunity* **48**, 979–991.e8 (2018).
33. H. Kettenmann, U.-K. Hanisch, M. Noda, A. Verkhratsky, *Physiol. Rev.* **91**, 461–553 (2011).
34. W. M. Pardridge, *NeuroRx* **2**, 3–14 (2005).
35. P. Mastorakos, D. McGavern, *Sci. Immunol.* **4**, eaav0492 (2019).
36. S. S. Erlich, J. G. McComb, S. Hyman, M. H. Weiss, *J. Neurosurg.* **70**, 926–931 (1989).
37. G. Kunis *et al.*, *Brain* **136**, 3427–3440 (2013).
38. B. Engelhardt, R. M. Ransohoff, *Trends Immunol.* **33**, 579–589 (2012).
39. Y. Zhou *et al.*, *J. Clin. Invest.* **124**, 3825–3846 (2014).
40. M. P. Lun *et al.*, *J. Neurosci.* **35**, 4903–4916 (2015).
41. V. Zywitza, A. Misios, L. Bunatyan, T. E. Willnow, N. Rajewsky, *Cell Rep.* **25**, 2457–2469.e8 (2018).
42. G. L. Lehmann *et al.*, *J. Exp. Med.* **217**, e20190730 (2020).
43. J. Kalucka *et al.*, *Cell* **180**, 764–779.e20 (2020).
44. L. He *et al.*, *Sci. Data* **5**, 180160 (2018).
45. M. B. Chen *et al.*, *Cell Rep.* **30**, 4418–4432.e4 (2020).
46. V. Koliaraki, A. Prados, M. Armaka, G. Kollias, *Nat. Immunol.* **21**, 974–982 (2020).
47. M. Corada *et al.*, *Dev. Cell* **18**, 938–949 (2010).
48. S. Liebner *et al.*, *J. Cell Biol.* **183**, 409–417 (2008).
49. Y. Wang *et al.*, *Nature* **465**, 483–486 (2010).
50. V. Brault *et al.*, *Development* **128**, 1253–1264 (2001).
51. A. Elkadri *et al.*, *Cell. Mol. Gastroenterol. Hepatol.* **1**, 381–394.e7 (2015).
52. Y. Wang *et al.*, *eLife* **8**, e43257 (2019).
53. F. Benz *et al.*, *eLife* **8**, e43818 (2019).
54. I. Spadoni, A. Pietrelli, G. Pesole, M. Rescigno, *Gut Microbes* **7**, 540–548 (2016).
55. B. W. Chow *et al.*, *Nature* **579**, 106–110 (2020).
56. A. Cugurra *et al.*, *Science* **373**, eabf7844 (2021).
57. Data for: S. Carloni, A. Bertocchi, S. Mancinelli, M. Bellini, M. Erreni, A. Borreca, D. Braga, S. Giugliano, A. M. Mozzarelli, D. Manganaro, D. Fernandez Perez, F. Colombo, A. Di Sabatino, D. Pasini, G. Penna, M. Matteoli, S. Lodato, M. Rescigno, Identification of a choroid plexus vascular barrier closing during intestinal inflammation, Zenodo (2021); https://doi.org/10.5281/zenodo.4723172.
58. R. Edgar, M. Domrachev, A. E. Lash, *Nucleic Acids Res.* **30**, 207–210 (2002).

ACKNOWLEDGMENTS

We thank T. Di Fiore, V. Ferrari, A. Melacarne, G. Fornasa, J. Mouries, and F. Spaggiari for experimental help; and A. Carloni for graphical support; and E. Dejana for providing the *Cdh5* (PAC)-CreERT2/*β-catenin*$^{lox(ex3)/lox(ex3)}$/VEC-PAC mice and useful collaboration. We are also grateful for a research donation in memory of G. Caligaris. **Funding:** This research was supported by the European Research Council (ERC HomeoGUT grant no. 615735 to M.R.); the Italian Foundation for Cancer Research (grant nos. AIRC IG 22026, AIRC 5x1000 2018 ID.21147, and AIRC 5x1000 2019 ID.22757 to M.R.); Fondazione CARIPLO (grant no. 2019-1785 to S.L.); PRIN (Ministero dell'Istruzione dell'Università e della Ricerca grant no. 2017A9MK4R to M.M.); and FISM (2019/R-Single/032 to M.M.). D.F.-P. was supported by a fellowship from AIRC/FIRC within the PhD program of the European School of Molecular Medicine (SEMM). **Author contributions:** Conceptualization: M.R.; Data curation: S.C.; Formal analysis: M.B., M.E., S.C., S.M., A.Bo., D.B., D.P.F.; Funding acquisition: M.R.; Investigation: S.C., A.B., S.G., A.M.M., D.M., F.C., M.B., S.M., M.E., A.Bo.; Resources: A.D.S., D.P., M.M., S.L., M.R.; Software: M.E.; Supervision: G.P., D.P., M.M., S.L., M.R.; Visualization: S.C., M.E.; Writing – original draft: S.C., M.M., S.L., M.R.; Writing – review & editing: S.C., M.M., S.L., M.R. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** *Cdh5* (PAC)-CreERT2 mice and *Cdh5* (PAC)-CreERT2/*β-catenin*$^{lox(ex3)/lox(ex3)}$/VEC-PAC mice were obtained under a material transfer agreement with R. H. Adams, Cancer Research UK London Research Institute and European Institute of Oncology (Milan, Italy). Original data and macros for Fiji generated in this study have been deposited to Zenodo (*57*). The total RNA-sequencing and single-cell RNA data discussed in this publication have been deposited in the National Center for Biotechnology Information's Gene Expression Omnibus (*58*) and are accessible through GEO series accession number GSE178752.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abc6108
Materials and Methods
Figs. S1 to S15
References (*59*–*75*)
Tables S1 to S6
Movies S1 and S2
MDAR Reproducibility Checklist

4 May 2020; resubmitted 10 April 2021
Accepted 9 July 2021
10.1126/science.abc6108

REPORTS

NANOPHOTONICS

Broadband electro-optic polarization conversion with atomically thin black phosphorus

Souvik Biswas[1]†, Meir Y. Grajower[1]†, Kenji Watanabe[2], Takashi Taniguchi[3], Harry A. Atwater[1]*

Active polarization control is highly desirable in photonic systems but has been limited mostly to discrete structures in bulky dielectric media and liquid crystal–based variable retarders. Here, we report electrically reconfigurable polarization conversion across telecommunication wavelengths (1410 to 1575 nanometers) in van der Waals layered materials using tri-layer black phosphorus (TLBP) integrated in a Fabry-Pérot cavity. The large electrical tunability of birefringence in TLBP enables spectrally broadband polarization control. We found that polarization states could be generated over a large fraction of the Poincaré sphere through spectral tuning, and that electrical tuning enables the state of polarization conversion to span nearly half the Poincaré sphere. We observed both linear to circular and cross-polarization conversion with voltage, demonstrating versatility with a high dynamic range.

Polarization is a fundamental property of light that plays a crucial role in classical and quantum optics. In tabletop optical experiments, it is usually tailored using polarizers, wave plates, and phase retarders. Generating arbitrary polarization states on demand is a requirement for a wide range of photonic processes such as circular dichroism sensing of chiral molecules and proteins, polarization-sensitive digital holography, imaging, polarization encoding for photonic qubits, and detecting material quasiparticle excitations such as phonons, spins, or excitons and their associated anisotropy (*1*–*6*). Active

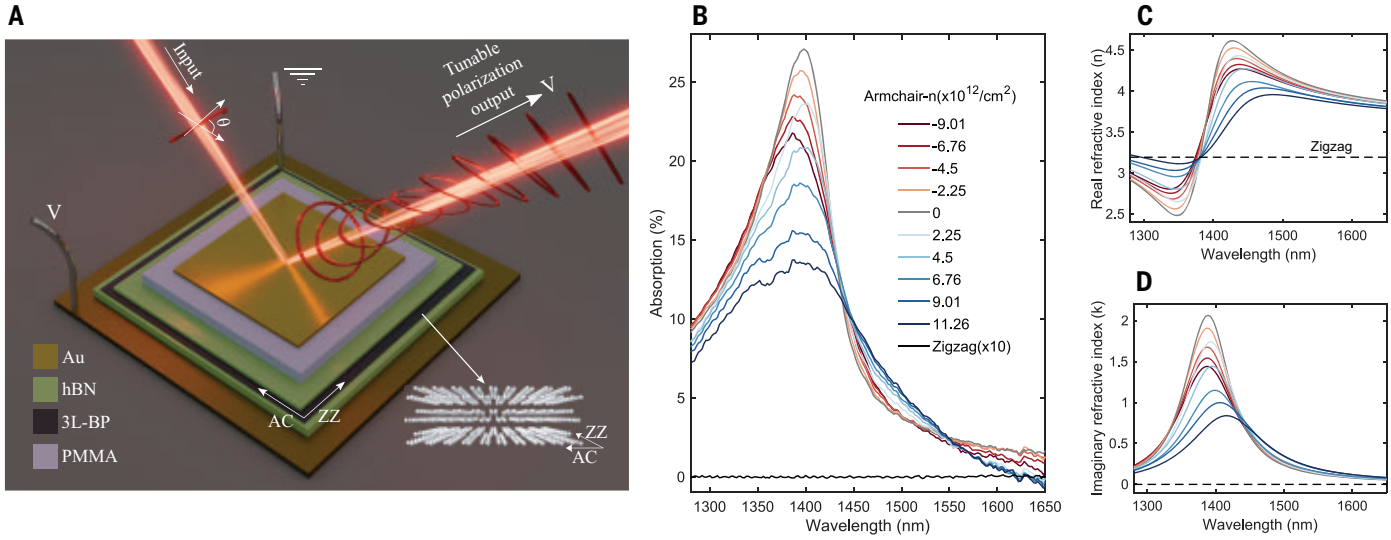


Fig. 1. Schematic of electrically tunable polarization conversion and TLBP birefringence. (**A**) Schematic of cavity design and polarization conversion. TLBP is incorporated in a dielectric environment between two mirrors [one partially reflective (top) and one highly reflective (bottom)]. The incoming beam is linearly polarized, and the output beam can be azimuthally rotated or converted between circular and linear polarization with applied voltage (between the TLBP and the back electrode or mirror) for a fixed wavelength. (**B**) Experimentally measured polarized absorption from a TLBP device (non-cavity-integrated) for different doping densities along the AC direction. The ZZ direction remains featureless for all conditions. (**C** and **D**) Extracted complex refractive index (real and imaginary part, respectively) for TLBP as a function of doping density for the AC and ZZ directions.

[1]Thomas J. Watson Laboratory of Applied Physics, California Institute of Technology, Pasadena, CA 91125, USA. [2]Research Center for Functional Materials, National Institute for Materials Science, Tsukuba 305-0044, Japan. [3]International Center for Materials, Nanoarchitectonics, National Institute for Materials Science, Tsukuba 305-0044, Japan.
***Corresponding author. Email: haa@caltech.edu**
†These authors contributed equally to this work.

control of polarization demands a platform with two key ingredients: a material with complex refractive indices that can be dynamically tuned (*7*, *8*) and in-plane symmetry breaking to generate birefringence either from an inherent material property (*9*) or through nanostructured metasurface elements (*10*, *11*). The former can be achieved using different effects, such as the Pockels effect (magneto-optical), the Kerr effect, Pauli blocking, the quantum-confined Stark effect, free carrier absorption, and optical pumping (*8*, *12*–*16*). Despite important advances in the exploration of these phenomena, electrically reconfigurable control of polarization remains largely elusive and suffers from limited dynamic range of polarization conversion. Furthermore, commercial polarization control devices such as variable phase retarders that use liquid crystal or lithium niobate as active media are bulky and difficult to integrate into integrated photonic platforms, restricting the miniaturization of photonic systems.

Two-dimensional (2D) van der Waals semiconductors are known to exhibit strongly bound excitons and are extremely polarizable in the presence of external electric fields, enabling unprecedented electric field–induced doping and refractive index modulation (*17*–*19*). The atomic-scale thickness of van der Waals materials and the lack of a lattice-matching requirement make them attractive candidates for future-generation optoelectronics operating in the visible and telecommunications frequency bands for applications such as coherent control of quantum light, free-space and fiber-based communications, and light detection and ranging. (*20*, *21*) However, most reports on electro-optic modulation in 2D materials thus far have largely been limited to graphene and transition metal dichalcogenides, which are optically isotropic in the layer plane and thus are not suitable for inherent active polarization control. Black phosphorus (BP), a 2D quantum well–like semiconductor, exhibits natural birefringence because of its in-plane anisotropic crystal structure (*22*–*24*). Its unique electrically tunable optical dichroism demonstrated at the few-layer limit combined with its thickness tunable bandgap (~750 nm to ~4 μm) opens possibilities for polarization-sensitive electro-optic modulation at infrared wavelengths in the telecommunication band with a well-chosen thickness of BP (*23*–*28*).

In the conceptual visualization of polarization conversion (Fig. 1A), an incoming linearly polarized light illuminates a Fabry-Pérot cavity incorporating tri-layer BP (TLBP). The polarization state of the reflected light can be electrically tuned to alter its ellipticity (circular to linear polarization) or its azimuthal angle (s- to p-polarization). Here, the two cavity mirrors are composed of a highly reflecting, thick back gold (Au) mirror and a partially transmitting, thin top Au mirror. The cavity medium is formed by hexagonal boron nitride (hBN)–encapsulated TLBP and polymethyl methacrylate (PMMA), which act to adjust the cavity resonance frequency to a critically coupled condition for maximal polarization conversion.

To characterize the electrically tunable complex refractive index of TLBP, gate-dependent polarized absorption measurements were performed on a sample with the following configuration: hBN/TLBP (with few-layer graphene contacts)/hBN/Au (Fig. 1B). The charge neutral response (0 V) is dominated by a strong excitonic feature at 1398 nm, arising from the optical transition between the lowest (highest) lying conduction (valence) bands, in the quantum well–like band structure of TLBP. For both positive and negative voltages (electron and hole doping, respectively), the peak absorption reduces along with an increase in the linewidth of the excitonic transition. A stronger change is observed in the hole-doping response with applied voltage compared with electron doping. A detailed discussion of the mechanism for the strong electro-optic response is provided in sections S8 and S9 of the materials and methods (*29*). Through a

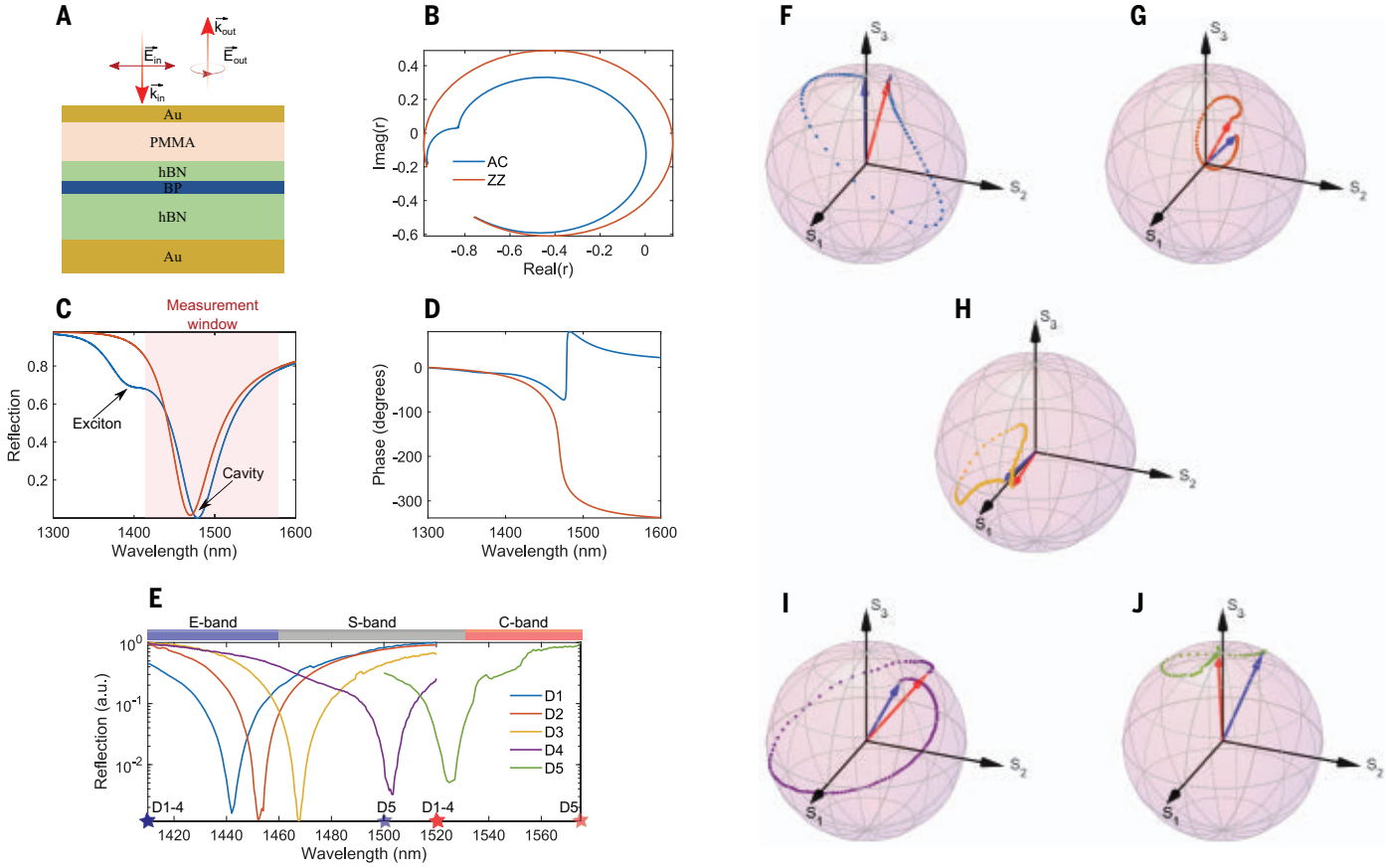


Fig. 2. Example cavity design for polarization conversion and large anisotropy bandwidth experimental demonstration. (**A**) Side view of a typical cavity structure adopted in this work. The top and bottom mirrors are formed by thin and thick Au films, respectively. The cavity is composed of hBN-encapsulated TLBP and PMMA, which act as the tunable part in determining the resonance wavelength. (**B** to **D**) Theoretically calculated complex reflection phasor, reflection amplitude, and phase spectrum, respectively, for such a typical cavity structure having resonance ~1480 nm, showing differences in both parameters along AC and ZZ and thus establishing polarization conversion. (**E**) Summary of reflection amplitude spectra from five representative devices fabricated as part of this study showing tunable cavity resonance. The PMMA thickness was tuned systematically to change the resonance over 90 nm across the telecommunication band (E, S, and C). (**F** to **J**) Experimentally measured spectral trajectories on the normalized Poincaré sphere corresponding to the five device resonances plotted in (E). All trajectories show strong spectral polarization conversion (either in the azimuthal orientation or the ellipticity or both). The differences in the trajectories are intimately related to the critical coupling between the cavity and the incoming polarization. For all presented trajectories, the azimuthal orientation was aligned nearly 45 degrees to the AC and ZZ direction of the TLBP flake. For each normalized Poincaré sphere, the blue arrows mark the beginning of the spectral scan (1410 nm for D1 to D4, 1500 for D5) and the red arrows mark the end (1520 nm for D1 to D4, 1575 for D5), also shown as stars in *x*-axis of (E).

Kramers-Kronig consistent transfer matrix analysis of the gate-dependent absorption, the complex refractive indices ($\tilde{n} = n + ik$) of TLBP for different doping densities were estimated (Fig. 1, C and D). Near-unity tuning of the complex refractive indices (for both n and k) is observed near the excitonic resonance along the armchair (AC) direction. No noticeable feature was seen along the orthogonal direction, zigzag (ZZ), for any voltages, rendering that polarization passive.

We then next designed a heterostructure for polarization conversion by integrating the TLBP in an optically resonant cavity geometry that enhances the degree of polarization conversion. A transfer matrix calculation of a typical Fabry-Pérot cavity design (side view in Fig. 2A) yields the complex reflection phasor, amplitude, and phase spectra (Fig. 2, B to D, respectively). From the phasor diagram in Fig. 2B, a prominently different complex reflectivity feature is seen along the two polarizations. A clear resonance from the cavity is seen at 1479 (1470) nm for the AC (ZZ) direction in the reflection amplitude in Fig. 2C, along with a weaker excitonic absorption feature at 1398 nm seen only along the AC direction. The reflected phase along the AC and ZZ in Fig. 2D shows strong differences near the cavity resonance. Taken together, these results indicate the potential for appreciable polarization conversion of the reflected light. The cavity parameters used in Fig. 2, B to D, are not a unique choice. In fact, the optical anisotropy in TLBP is broadband, enabling operation over the entire wavelength range of the telecommunication E, S, and C bands with appropriate changes in the cavity parameters, primarily through adjustment of the thicknesses of the dielectric medium (hBN or PMMA) and the top Au mirror. Both parameters are important in determining the resonance wavelength and the reflection extinction ratio [see sections S14 and S15 in the materials and methods (*29*)]. To experimentally demonstrate the broadband nature of the TLBP anisotropy, reflection intensities were measured from five representative heterostructures (Fig. 2E). The PMMA thickness was sequentially tuned to redshift the cavity resonance, spanning ~100 nm across the E, S, and C telecommunication bands. For

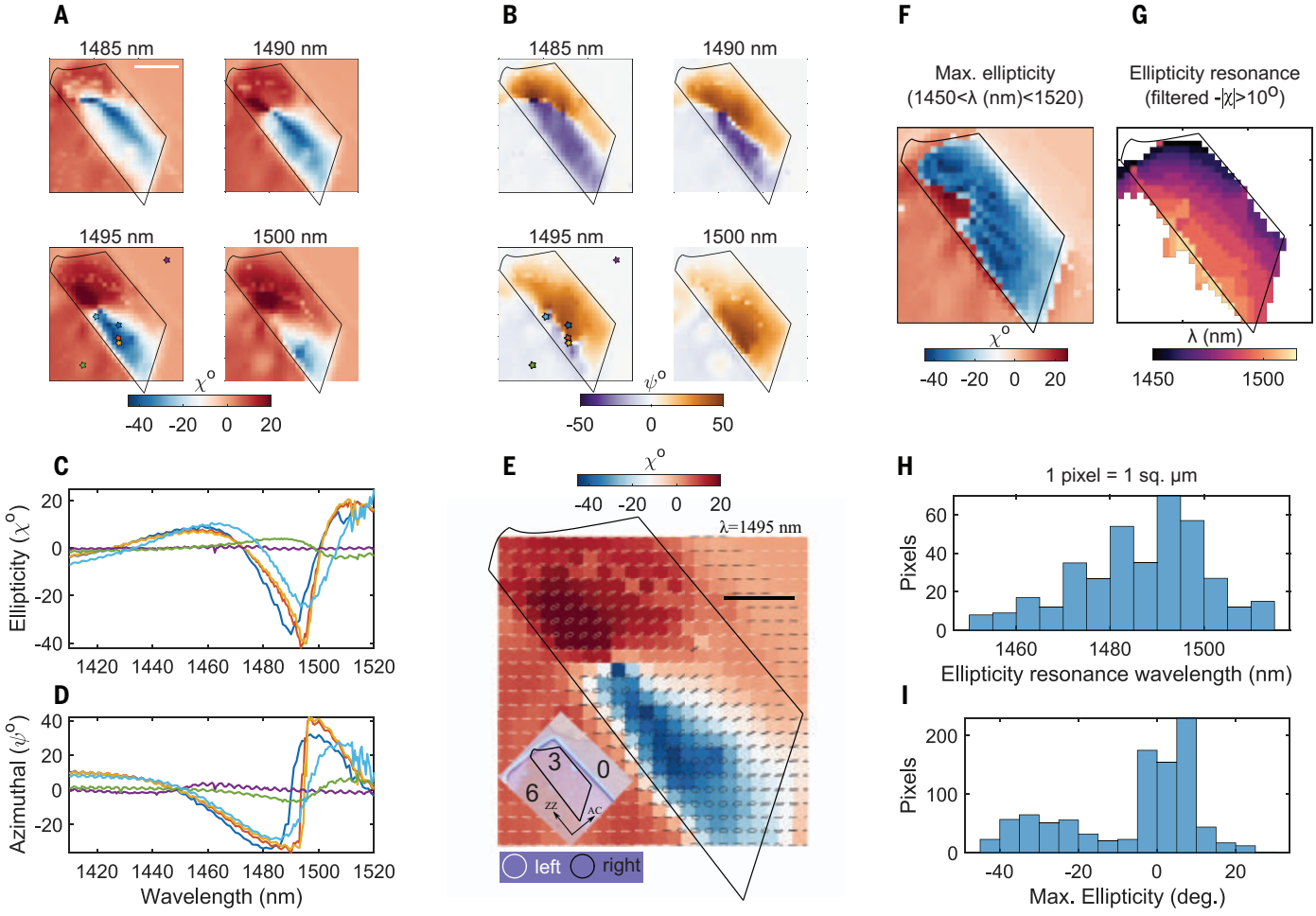


Fig. 3. Spatial mapping of optical anisotropy probed by polarization conversion. (**A**) Spatial maps of ellipticity angle (in degrees) of device D4 for four different wavelengths near the resonance (~1490 nm) of the cavity. Black lines indicate the extent of the tri-layer region [optical image of D4 is shown in the inset of (E)]. White scale bar, 10 μm. (**B**) Same as (A) but for azimuthal angle. (**C** and **D**) Ellipticity and azimuthal angle spectral scans for a few points [marked with appropriately colored stars in (A) and (B)], showing spatial variation of the resonance in the tri-layer region, as well as flat background response from the bare cavity and weak polarization conversion from the six-layer region. (**E**) Zoom-in spatial color map of ellipticity at 1495 nm, along with superimposed reflected polarization ellipses at each point, for better visualization of covariation of azimuthal and ellipticity angles. White lines correspond to right handedness and black lines correspond to left handedness. Black scale bar, 5 μm. Inset, Optical image of device D4 outlining the three-layer ("3") region. Also shown is the six-layer region ("6") and the bare cavity ("0"). (**F**) Spatial map of maximum ellipticity for each point within a spectral window between 1450 and 1520 nm. (**G**) Spatial map of ellipticity resonance wavelength (filtered for $|\chi| > 10°$ to only highlight the three-layer region). (**H** and **I**) Histograms of ellipticity resonance wavelength (filtered for $|\chi| > 10°$) and maximum ellipticity (in degrees), where one pixel on the map corresponds to 1 μm^2.

each heterostructure device (D1 to D5), a corresponding spectral trajectory on the Poincaré sphere is shown in Fig. 2, F to J, respectively. The blue (red) arrows mark the beginning (end) of the measured spectral trajectory, corresponding to 1410 (1520) nm for D1 to D4 and 1500 (1575) nm for D5. Efficient polarization conversion can be seen for all the devices, confirming the broadband nature of the anisotropy in TLBP coupled with the cavity mode. The differences in the trajectories arise where the cavity resonance wavelength is with respect to the beginning and ending point of the spectral scans. In addition, the arc length subtended by the trajectories on the Poincaré sphere is intimately related to how well the cavity critically couples to the incoming free-space electromagnetic field and is dominated strongly by the top mirror reflectivity.

Planar heterostructure–based optical cavities can act as sensitive probes to any spatially varying features such as dielectric disorder or thickness of BP, which are key to understanding polarization conversion dynamics. Spatial false-color maps of ellipticity (Fig. 3A) and azimuthal angles (Fig. 3B), measured for heterostructure device D4, are shown for four different wavelengths near the cavity resonance, showing that the polarization conversion has strong spatial variation. An optical image of the sample with appropriately outlined bare cavity (0), three-layer, and six-layer BP regions is shown in the inset of Fig. 3E. To gain further insight, spectral scans are shown in Fig. 3, C and D, for a select few points (appropriately labeled) chosen from the aforementioned regions. Strong polarization conversion is seen on the trilayer region. For example, a linearly polarized input (azimuth ~45° to the AC/ZZ axes) is converted to nearly circular (42° ellipticity) at the resonance. An associated feature is seen in the azimuth spectrum as well (with a

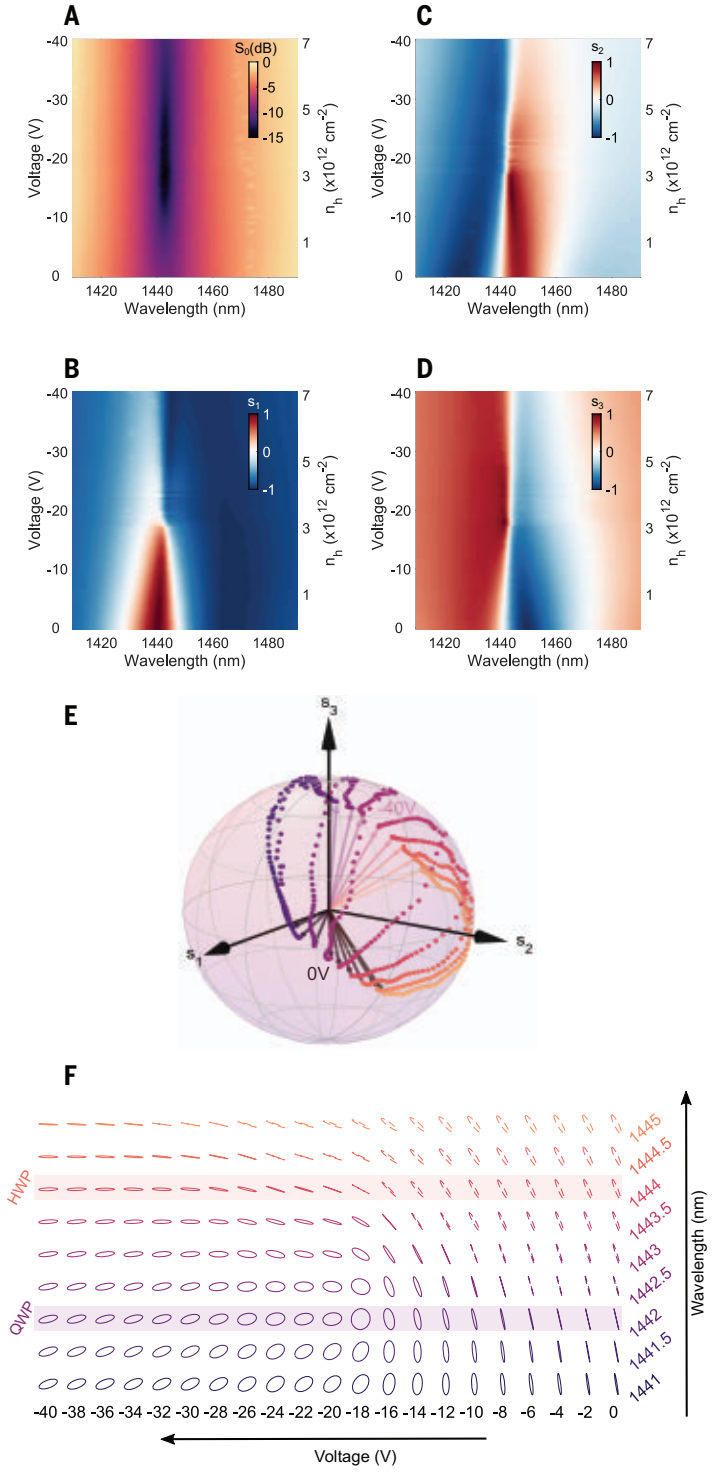


Fig. 4. Electrically tunable polarization dynamics. (**A** to **D**) False-color maps of the evolution of the intensity (S_0) and the three normalized Stokes parameters (s_1, s_2, s_3) determining the polarization state of the reflected light as a function of wavelength and negative voltages (for hole doping). The results are from device D1. Continuous tuning of all four parameters can be seen around the cavity resonance (~1440 nm) for the entire range of doping, illustrating efficient tuning of the polarization state with voltage. (**E**) Voltage-dependent trajectories on the normalized Poincaré sphere for nine different wavelengths showing large dynamic range in tunability of the reflected polarization state. Each color corresponds to a wavelength [same color code in (F)]. The dark arrows mark the beginning of the voltage scan (0 V), and the correspondingly colored arrows indicate the end of the voltage scan (−40 V), hole doping. (**F**) Visualization of the measured reflected polarization ellipse for selected voltages and the same nine wavelengths as in (E). At 1442 nm, a strong change in ellipticity is seen where the state becomes almost circular at −18 V and the ellipticity decreases for higher voltages, with the device acting as a quarter-wave plate. The change in ellipticity is associated with a change in the azimuthal orientation of the beam. At 1444 nm, however, minimal change in ellipticity is seen with a strong change in the azimuthal orientation, with the device effectively behaving like a half-wave plate. The solid (dashed) lines correspond to right (left) handedness.

derivate-like line shape), peaking at −37° and +42° around the ellipticity resonance. As expected, very weak (no) polarization conversion effects are seen spectrally in the six-layer (bare cavity) region. The weak effect for the six-layer region presumably originates from the higher quantum–well subband transitions leading to reduced birefringence. To aid the visualization of covariation in azimuthal and ellipticity spatially, a false-color map of ellipticity superimposed with measured reflection polarization ellipses is shown in Fig. 3E. The white (black) color corresponds to right (left) handedness. Finally, in Fig. 3, F and G, spatial maps of the maximum achieved ellipticity over a bandwidth of 70 nm (1450 to 1520 nm) and the corresponding resonance wavelength are plotted, respectively. The overall achieved maximum ellipticity is relatively homogeneous, although the spectral distribution is quite broad. The histogram in Fig. 3H (filtered for absolute ellipticity values >10°) confirms a strong centering of the ellipticity response around the resonance wavelength (~1490 nm). Figure 3I shows

the distribution of maximum ellipticity achieved for all pixels across the entire map presented in Fig. 3F. Two strong peaks are seen near +10° and −35°, corresponding to the peak ellipticity values observed in most of the trilayer region. Our observation of spatially varying complex refractive indices is consistent with dielectric disorder seen in typical 2D heterostructures (*30*). We speculate the origin of such behavior to be twofold: fabrication-induced trapped hydrocarbons and strain between constituent layers in the heterostructure and different stacking orders in TLBP (ABA, AAB, and ACA) (*31*–*33*).

The key result of this work, the electrical tunability of the different polarization parameters, is summarized for device D1 in Fig. 4. The Stokes intensity (S_0) and the three normalized Stokes parameters (s_1, s_2, s_3), which completely characterize the polarization conversion induced by the device, are shown as a function of wavelength and gate voltage in the form of false-color maps in Fig. 4, A to D, for hole-doping regime. It can be seen that all three s parameters tune efficiently around the resonance (~1440 nm) with increased hole doping because of the strong cavity and TLBP interaction. Our observations of the electrically driven changes in the s parameters are consistent with the measured complex refractive indices of TLBP, including a stronger hole-doped response compared with electron doping [see section S21 of the materials and methods (*29*)]. Spectral variation of S_0 (Fig. 4A) with doping indicates that the strongest coupling with the cavity was achieved at ~−18V. Furthermore, in all three s parameters (Fig. 4, B to D), a suppression of the overall magnitude of the resonance, which is caused by the reduction in excitonic anisotropy, can be seen. In addition, the linewidth of the aforementioned features decreases with increasing voltage because of the suppression of the losses along the AC direction. The polarization rotation at the resonance wavelength is primarily enabled by the higher absorption along the AC axis compared with the ZZ axis (Fig. 1B), and this is further enhanced by the cavity.

To better visualize the changes in the reflected polarization states, traces on the Poincaré sphere for different wavelengths around the resonance are plotted as functions of gate voltage. Figure 4E shows such traces on the Poincaré sphere for nine different wavelengths. The dynamic range corresponds to nearly half the Poincaré sphere in terms of solid angle subtended by the voltage-driven arcs. Two interesting traces can be identified. The first one is at 1442 nm, where the ellipticity changes from 0.3° to 43.7° to 16.8° at 0, −18, and −40 V, respectively, showing tunable quarter-wave plate operation. By contrast, at 1444 nm, the azimuthal angle is tuned from 24.6° to 89.4° between 0 and −40 V with suppressed ellipticity changes, demonstrating tunable half-wave plate operation. At other wavelengths, it is possible to demonstrate a wide variety of elliptically polarized states. A 2D map of different polarization ellipses measured as a function of nine different wavelengths (same wavelengths and color code as Fig. 4E) is shown in Fig. 4F for voltages between 0 and −40V (hole doping), where the strongest changes are noted. This map better illustrates the quarter-wave and half-wave plate–like operation, as well as other intermediate polarization conversion configurations. A similarly high dynamic range for polarization conversion is also seen for electron doping (fig. S24). A general trend observed in these voltage-dependent polarization conversion measurements is that upon doping the TLBP, the spectral trajectory on the Poincaré sphere can be collapsed to a point, a manifestation of electrically tunable anisotropy suppression. Thus, the larger the trajectory at nearly charge-neutral doping conditions, the higher the capability to access a wide range of polarization states by applying a voltage.

Our results demonstrate the versatility of BP as an active medium for electronically reconfigurable broadband polarization conversion. The Γ-point nature of the direct-band minima enables BP to have pronounced band-edge optical anisotropy spanning from visible (750 nm) to the mid-infrared (4 μm) (*23*, *24*), whereas free-carrier modulation provides access to mid to far-infrared wavelengths (>5 μm) (*26*). TLBP is of particular interest for polarization conversion at telecommunications wavelengths because of its near unity birefringence close to the excitonic resonance. These findings represent a route for active control of optical polarization at the nanoscale. The high dynamic range of polarization conversion may open an avenue for the realization of densely integrated arrays of nanoscale BP electro-optic polarization converters as a fundamental step beyond discrete dielectric polarization converters in lithium niobate or arrays based on micron-scale liquid crystal spatial light modulators. Previous reports on high-speed BP electrophotoresponse (*34*) and recent advances in centimeter-scale, layer-controlled growth of BP thin films (*35*) suggest that possibilities may emerge for large-area, broadband polarization-selective sensing, photodetection, and active electro-optic modulation.

REFERENCES AND NOTES

1. S. Fleischer, Y. Zhou, R. W. Field, K. A. Nelson, *Phys. Rev. Lett.* **107**, 163603 (2011).
2. G. P. Crawford, J. N. Eakin, M. D. Radcliffe, A. Callan-Jones, R. A. Pelcovits, *J. Appl. Phys.* **98**, 123102 (2005).
3. S. G. Demos, R. R. Alfano, *Appl. Opt.* **36**, 150–155 (1997).
4. M. Schubert, T. E. Tiwald, C. M. Herzinger, *Phys. Rev. B Condens. Matter Mater. Phys.* **61**, 8187–8201 (2000).
5. X. Ling *et al.*, *Nano Lett.* **16**, 2260–2267 (2016).
6. A. Crespi *et al.*, *Nat. Commun.* **2**, 566 (2011).
7. Y. Yu *et al.*, *Nano Lett.* **17**, 3613–3618 (2017).
8. E. Feigenbaum, K. Diest, H. A. Atwater, *Nano Lett.* **10**, 2111–2116 (2010).
9. S. Niu *et al.*, *Nat. Photonics* **12**, 392–396 (2018).
10. L. H. Nicholls *et al.*, *Nat. Photonics* **11**, 628–633 (2017).
11. A. H. Dorrah, N. A. Rubin, A. Zaidi, M. Tamagnone, F. Capasso, *Nat. Photonics* **15**, 287–296 (2021).
12. S. Abel *et al.*, *Nat. Mater.* **18**, 42–47 (2019).
13. R. R. Subkhangulov *et al.*, *Nat. Photonics* **10**, 111–114 (2016).
14. Y. Gao *et al.*, *Nano Lett.* **15**, 2001–2005 (2015).
15. Y.-H. Kuo *et al.*, *Nature* **437**, 1334–1336 (2005).
16. A. J. Sternbach *et al.*, *Science* **371**, 617–620 (2021).
17. A. Chernikov *et al.*, *Phys. Rev. Lett.* **115**, 126802 (2015).
18. J. S. Ross *et al.*, *Nat. Commun.* **4**, 1474 (2013).
19. S. Gao, Y. Liang, C. D. Spataru, L. Yang, *Nano Lett.* **16**, 5568–5573 (2016).
20. I. Datta *et al.*, *Nat. Photonics* **14**, 256–262 (2020).
21. J. van de Groep *et al.*, *Nat. Photonics* **14**, 426–430 (2020).
22. X. Ling, H. Wang, S. Huang, F. Xia, M. S. Dresselhaus, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 4523–4530 (2015).
23. G. Zhang *et al.*, *Nat. Commun.* **8**, 14071 (2017).
24. G. Zhang *et al.*, *Sci. Adv.* **4**, eaap9977 (2018).
25. W. S. Whitney *et al.*, *Nano Lett.* **17**, 78–84 (2017).
26. S. Biswas *et al.*, *Sci. Adv.* **7**, eabd4623 (2021).
27. C. Chen *et al.*, *Sci. Adv.* **6**, eaay6134 (2020).
28. Y. Liu *et al.*, *Nano Lett.* **17**, 1970–1977 (2017).
29. The materials and methods are available as supplementary materials.
30. A. Raja *et al.*, *Nat. Nanotechnol.* **14**, 832–837 (2019).
31. D. Y. Qiu, F. H. da Jornada, S. G. Louie, *Nano Lett.* **17**, 4706–4712 (2017).
32. S. Huang *et al.*, *Nat. Commun.* **10**, 2447 (2019).
33. D. Çakır, C. Sevik, F. M. Peeters, *Phys. Rev. B Condens. Matter Mater. Phys.* **92**, 165406 (2015).
34. N. Youngblood, C. Chen, S. J. Koester, M. Li, *Nat. Photonics* **9**, 247–252 (2015).
35. Z. Wu *et al.*, *Nat. Mater.* **20**, 1203–1209 (2021).

ACKNOWLEDGMENTS

We thank W. S. Whitney, M. Z. Alam, J. Wong, and P. C. Wu for fruitful discussions; G. R. Rossman for access to Raman and Fourier transform infrared spectroscopy, and A. Dutt for valuable feedback on the work. **Funding:** The authors gratefully acknowledge support from the Department of Energy Office of Science under grant DE-FG02-07ER46405. K.W. and T.T. acknowledge support from the Elemental Strategy Initiative conducted by Ministry of Education, Culture, Sports, Science and Technology (MEXT grant JPMXP0112101001), Japan Society for the Promotion of Science (JSPS KAKENHI grant JP20H00354), and Centers of Research Excellence in Science and Technology (CREST grant JPMJCR15F3), Japan Science and Technology Agency (JST). **Author contributions:** S.B., M.Y.G., and H.A.A conceived the project. S.B. fabricated and characterized the devices. S.B. and M.Y.G. performed numerical calculations and electrical and optical measurements and discussed the results. H.A.A. supervised the project. K.W. and T.T. provided hBN and BP crystals. S.B. wrote the manuscript with input from M.Y.G and H.A.A. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** All data required to evaluate the conclusions in the manuscript are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj7053
Materials and Methods
Supplementary Text
Figs. S1 to S36
References
Movies S1 to S4

27 May 2021; accepted 13 September 2021
10.1126/science.abj7053

WATER CAPTURE

Evolution of water structures in metal-organic frameworks for improved atmospheric water harvesting

Nikita Hanikel[1], Xiaokun Pei[1], Saumil Chheda[2], Hao Lyu[1], WooSeok Jeong[3], Joachim Sauer[4]*, Laura Gagliardi[5]*, Omar M. Yaghi[1]*

Although the positions of water guests in porous crystals can be identified, determination of their filling sequence remains challenging. We deciphered the water-filling mechanism for the state-of-the-art water-harvesting metal-organic framework MOF-303 by performing an extensive series of single-crystal x-ray diffraction measurements and density functional theory calculations. The first water molecules strongly bind to the polar organic linkers; they are followed by additional water molecules forming isolated clusters, then chains of clusters, and finally a water network. This evolution of water structures led us to modify the pores by the multivariate approach, thereby precisely modulating the binding strength of the first water molecules and deliberately shaping the water uptake behavior. This resulted in higher water productivity, as well as tunability of regeneration temperature and enthalpy, without compromising capacity and stability.

The recent discovery that porous metal-organic frameworks (MOFs) (*1*, *2*) can extract atmospheric water in the desert to produce potable water (*3*–*7*) raises the question of how such MOFs "pluck out" water from arid air and easily release it, especially on the molecular level. Indeed, the evolution of water structures in MOFs and synthetic crystals is sought after, but a full mechanistic understanding of water uptake behavior is still missing (*8*–*17*). Although the water positions in some of these structures have been determined with diffraction techniques (*11*, *13*, *16*), the mechanism of how water-binding sites are populated is much harder to decipher because it requires high-quality data and the ability to collect these data at each loading increment. Knowledge of the mechanism for water behavior in MOFs should enable the design of water-harvesting systems that can operate with greater energy efficiency and productivity.

We successfully determined the uptake mechanism of MOF-303 {[Al(OH)(PZDC)], where $PZDC^{2-}$ = 1-*H*-pyrazole-3,5-dicarboxylate}, the state-of-the-art water-harvesting MOF (Fig. 1A) (*5*, *6*), by using single-crystal x-ray diffraction (SCXRD) to locate all of the water molecules in its pores and identify the molecule-by-molecule sequence of filling these locations. Previously, the inorganic units of MOFs were observed as the strongest water-binding sites (*10*, *11*, *15*, *18*). Using experimental and computational methods, we found that the polar organic linkers play a primary role as adsorptive sites in MOF-303, where they are aligned to create hydrophilic pockets into which the first water molecules bind strongly and seed further uptake.

We used the multivariate MOF strategy (*19*, *20*), in which multiple functionalities line the pores across the crystal, to achieve accurate design of the geometry and strength of water interactions in a series of nine multivariate MOFs. This approach allowed us to precisely control the relative humidity (RH) levels at which these compounds extract water from arid air, as well as heat of adsorption, desorption temperature, and water productivity, without compromising the pore size, shape, or hydrolytic stability of the MOFs. On a fundamental level, this study transforms the development of water-harvesting materials from a trial-and-error activity to precision molecular design and deliberate shaping of the water behavior in the pores.

The structure of MOF-303 is based on infinite rod-like secondary building units (SBUs) consisting of alternating cis-trans corner-shared AlO_6 octahedra that are connected by $PZDC^{2-}$ linkers (Fig. 1). Topological reconstruction of this MOF provides the **xhh** topology (text S2). This arrangement has neighboring pyrazole functionalities that point toward each other and form a pocket defined by three μ_2-OH groups and two N(H) on each of the linkers (Fig. 1, A and B). Water adsorption into this environment resulted in an unusual water sorption isotherm in which the isotherm exhibits a small but functionally impacting step (labeled S; red segment in Fig. 1A) at low vapor pressures. Step S reduced the water-harvesting output per cycle (working capacity) for this MOF by about 20 wt%, and upon cycling extensively, this reduction would amount to large quantities of water. As additional water

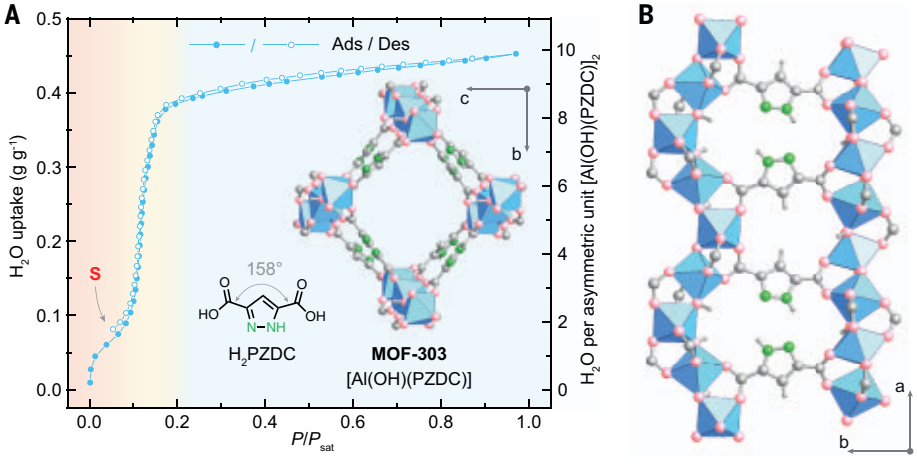


Fig. 1. Water sorption isotherm and crystal structure of MOF-303. (**A**) Water sorption isotherm at 25°C, where P is water vapor pressure and P_{sat} is the saturation water vapor pressure. The water uptake is displayed gravimetrically and with respect to the asymmetric unit $[Al(OH)(PZDC)]_2$ of MOF-303. The three segments of the isotherm are highlighted in red, yellow, and blue in the background. The linker 1*H*-3,5-pyrazoledicarboxalic acid (H_2PZDC) and MOF crystal structure are shown in the inset. The initial step is labeled with S. (**B**) Side view visualizing two hydrophilic pockets defined by a pair of pyrazole-based linkers with their nitrogen atoms pointing toward each other and flanked by two aluminum oxide rod-like SBUs. Here and in other figures, coordinate systems are given for guidance. Al, blue polyhedron; O, pink; C and H, gray; N, green.

[1]Department of Chemistry and Kavli Energy Nanoscience Institute, University of California, Berkeley, CA 94720, USA. [2]Department of Chemical Engineering and Materials Science, Chemical Theory Center, and Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, MN 55455, USA. [3]Department of Chemistry, Chemical Theory Center, and Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, MN 55455, USA. [4]Institut für Chemie, Humboldt-Universität zu Berlin, 10099 Berlin, Germany. [5]Department of Chemistry, Pritzker School of Molecular Engineering, James Franck Institute, and Chicago Center for Theoretical Chemistry, University of Chicago, Chicago, IL 60637, USA.
***Corresponding author. Email: js@chemie.hu-berlin.de (J.S.); lgagliardi@uchicago.edu (L.G.); yaghi@berkeley.edu (O.M.Y.)**

molecules filled the pore, a large water uptake within a small RH window was observed in the isotherm (yellow segment in Fig. 1A). Further water uptake by MOF-303 filled the pores entirely over a large RH range (blue segment in Fig. 1A). We sought to learn how to eliminate the initial step S by understanding the water uptake process through location of the adsorption sites and the mechanism with which they are populated.

Accordingly, we grew single crystals of sufficient size (15 μm × 15 μm × 20 μm; text S1) to be studied by synchrotron SCXRD. For the determination of the water structures at different loadings, we developed a procedure by which it was possible to slowly desorb water upon gradually ramping up the temperature of the dry protective gas stream. The process was designed to be sufficiently mild to avoid crystal cracking, which has often prevented SCXRD analysis. This procedure initiated a controlled release of water molecules from the MOF pores, which we monitored closely by collecting a series of SCXRD datasets, each acquired in a time frame of 3 to 12 min (text S3). The almost perfect overlay of the adsorption and desorption isotherm curves with minimal hysteresis observed for MOF-303 (Fig. 1A) indicated that the water sorption process was reversible and could be studied in either the adsorption or desorption mode. Thus, our SCXRD measurements allowed us to decipher the molecular water adsorption mechanism of MOF-303.

The SCXRD analysis revealed that the first and strongest water adsorption site of the framework (labeled I) was located between the pyrazoles, in which the water molecule formed three hydrogen bonds (H-bonds) to two pyrazole groups and one μ_2-OH group, with respective distances of 2.797(7) Å, 2.887(9) Å, and 2.798(6) Å between the heteroatoms (Fig. 2A). The second water molecule (II) was also situated between the pyrazoles and formed two H-bonds with N ⋮ O_{water} distances of 2.72(2) Å and 2.96(3) Å (Fig. 2B). Both of these sites were located within the hydrophilic pocket of MOF-303 and could be associated with the step S at low vapor pressures.

The next water molecule occupied site III and only interacted with the remaining μ_2-OH group at a distance of 2.89(3) Å (Fig. 2C). The fourth water molecule (IV) H-bonded to the water molecules at I and II but not to the framework, thus forming a trimeric cluster (I, II, and IV; Fig. 2D), with the water molecule at III remaining detached from it. At this loading stage, such clusters were isolated from others in neighboring symmetry equivalent pockets (Fig. 2E). Adsorption of water molecules at I to IV represented the seeding stage (red segment in Fig. 1A), which served as a nucleus for binding of other water molecules.

Using density functional theory (DFT) calculations (text S4), we predicted the seeding water adsorption sites in MOF-303. We found that the first adsorbed water molecule exhibited two short H-bonds to the pyrazole moieties (both 2.8 Å) and one short H-bond to the μ_2-OH group (2.7 Å), similar to our experimental observations. The second water molecule was determined to be H-bonded to the neighboring pyrazole pair (2.6 and 2.8 Å) and confirmed the role of the aligned pyrazole functionalities as primary adsorption sites. The third and fourth water molecules H-bonded to the second μ_2-OH group and to the initially adsorbed water molecules, respectively. This geometrical water arrangement closely resembled the obtained SCXRD structures, where the first three molecules adsorbed onto the framework and the fourth molecule H-bonded with these molecules. Furthermore, by applying a many-body decomposition scheme (text S4), we calculated the individual binding energy contributions of each water molecule at the seeding stage. A clear trend between H-bonding interaction

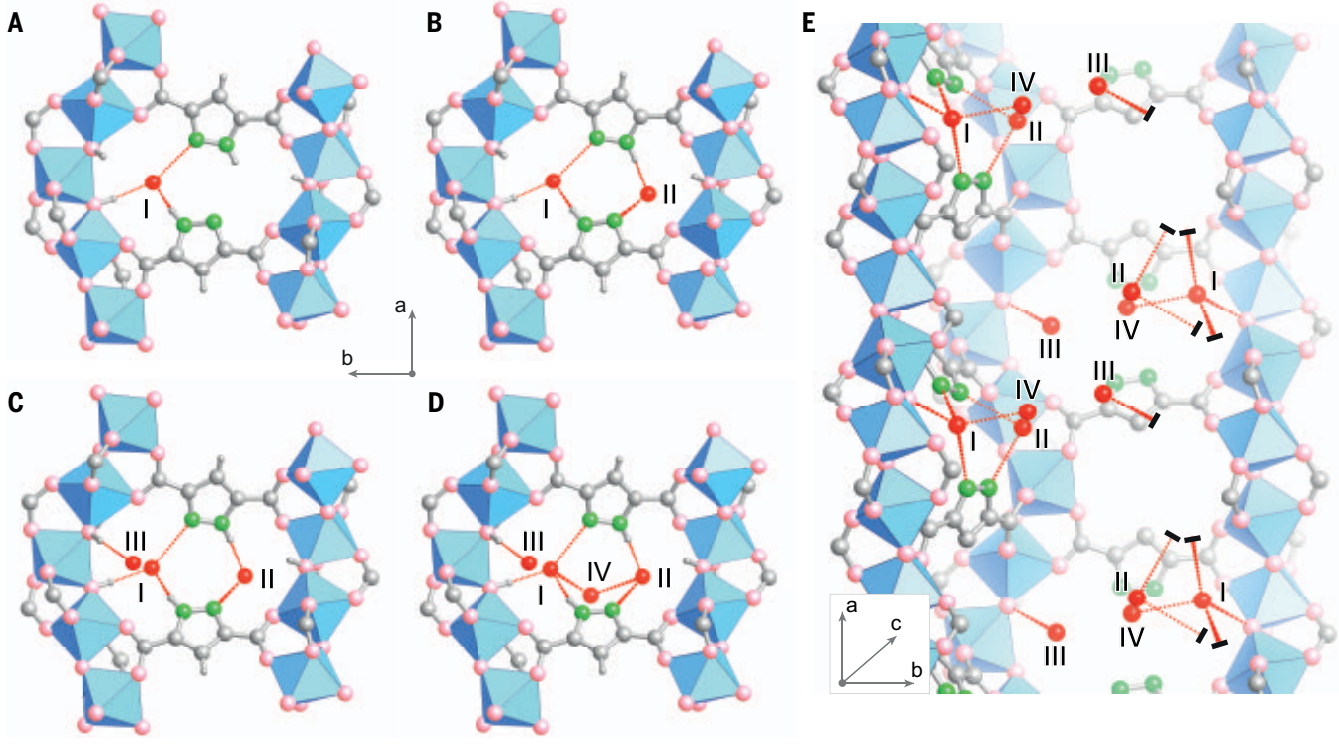


Fig. 2. Crystal structures of the seeding water adsorption sites in MOF-303. (**A** to **D**) Sequential adsorption of the first four water molecules (I to IV) per asymmetric building unit depicted in the hydrophilic pocket, as determined by SCXRD analysis. (**E**) Three-dimensional view of the first four water molecules in the framework pore. Shown crystallographic snapshots were captured at 0.04 g g^{-1} in (A), 0.06 g g^{-1} in (B), 0.11 g g^{-1} in (C), and 0.15 g g^{-1} in (D) and (E). H-bonds are depicted as red dashed lines. Black solid bars at the end of H-bonds in (E) represent binding interactions to the MOF, which is partially omitted for clarity. Al, blue polyhedron; O in the framework structure, pink; O in H_2O, red; C and H, gray; N, green. H atoms in (E) are omitted for clarity.

energies with the framework and the water-filling sequence was observed, thus further confirming the preferential siting of water molecules in MOF-303.

The crystallographic study revealed that additionally adsorbing water molecules filled the pores by interacting with other water molecules rather than with the framework itself (Fig. 3). Adsorption of two additional water molecules at V and VI yielded a structure in which the neighboring clusters (I, II, and IV) were connected with each other through new

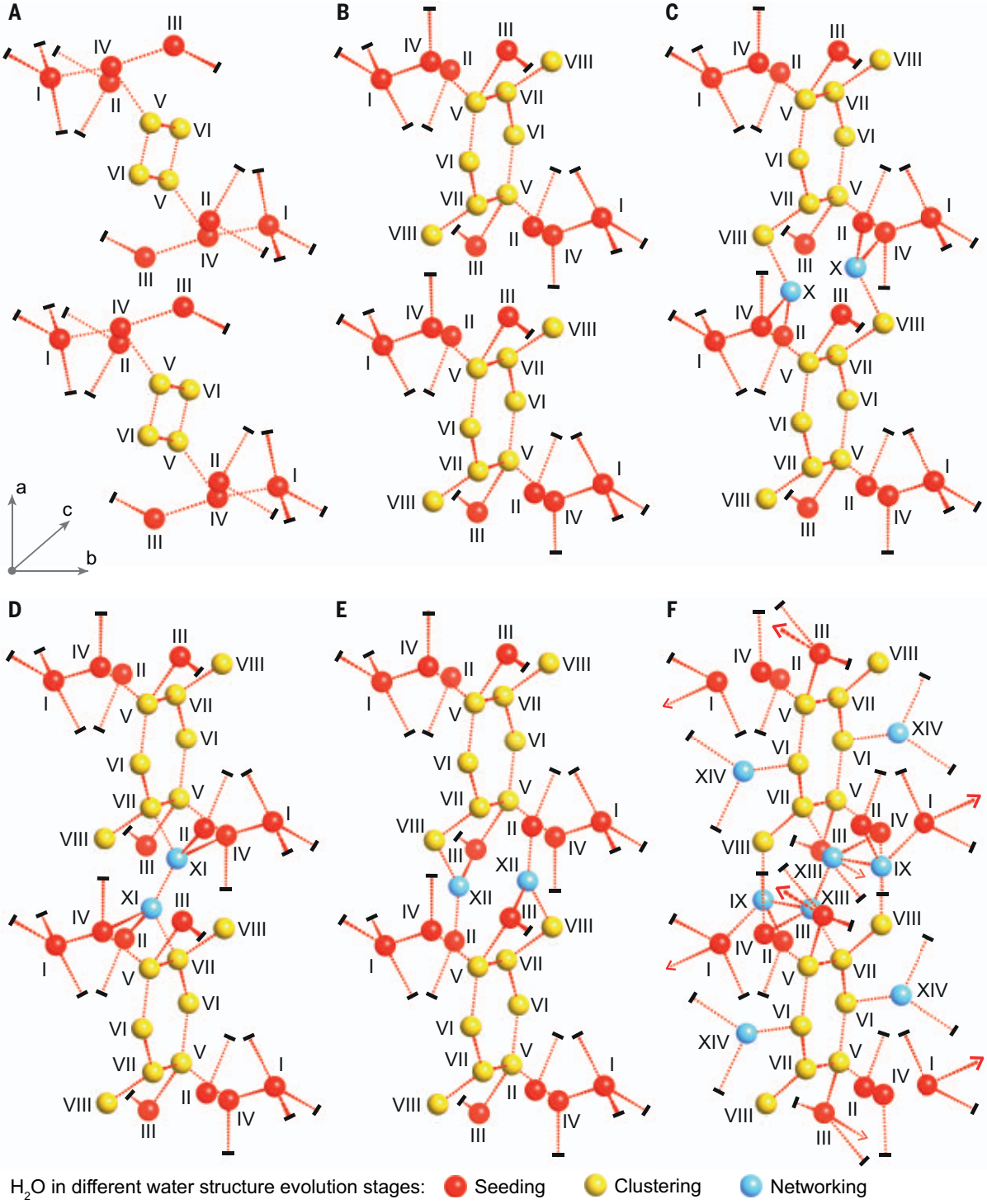


Fig. 3. Evolution of water structures in MOF-303 at increased loadings. (**A** to **F**) Water molecule positions at different adsorption states, as determined by SCXRD analysis. The water structures in (C) to (F) were consecutively co-populated and are depicted separately for clarity. Shown crystallographic snapshots were captured at 0.18 g g^{-1} in (A), 0.36 g g^{-1} in (B), 0.40 g g^{-1} in (C), 0.41 g g^{-1} in (D), 0.42 g g^{-1} in (E), and 0.45 g g^{-1} in (F). H-bonds are depicted as red dashed lines. Black solid bars at the end of H-bonds represent binding interactions to the MOF, which is omitted for clarity. The viewing direction for all panels is the same as in Fig. 2E. Arrows in (F) indicate periodic extension of water chains into a three-dimensional water network. Only the O atoms in H_2O are displayed, with colors indicating the pore-filling stages: seeding, red; clustering, yellow; networking, blue.

water tetramer clusters (Fig. 3A). Upon addition of water molecules at VII and VIII, the tetramers transformed into hexamers (V to VII) with one dangling water molecule at VIII connected to VII (Fig. 3B). The evolution of water structures from V to VIII, termed the clustering stage, fell in the yellow segment of the water sorption isotherm (Fig. 1A).

At higher water loadings, the sequential partial filling of sites IX to XIV together with the fully occupied sites I to VIII created infinite H-bonded water chains of clustering units (networking stage; blue segment in Fig. 1A). First, the mutually exclusive, disordered sites IX and X were populated simultaneously, with the associated water molecules H-bonding to the water molecules at II, IV, and VIII (Fig. 3C; only X is depicted for clarity). Second, site XI, which was connected to II, IV, and VII, was partly filled (Fig. 3D). Third, partial population of site XII connected II, III, and VIII (Fig. 3E). At the highest loading, the pyrazole functionalities became slightly displaced, whereas a water molecule inserted at XIII between sites IV, VII, and IX, consequently rearranging the water molecule at IX to H-bond to the species at I, II, and VIII.

These changes to the framework and restructuring of the water network resulted in partial filling of site XIV, in which the water molecule H-bonded to the O-atoms of the carboxy functional groups (Fig. 3F). At this loading, sites IX to XIV were co-populated and partly filled, and the occupational factors of these sites summed up to two water molecules in total (table S3). Ultimately, neighboring chain networks connected into a three-dimensional water network. The highest water load of 10 water molecules per asymmetric unit, whose crystallographic locations were presented above, corresponded to a total uptake of 0.45 g g^{-1} and was in agreement with the total uptake found in the water sorption isotherm (Fig. 1A).

Our series of SCXRD measurements allowed us to further elaborate the impact of water uptake on the framework. Throughout the water adsorption process, the MOF underwent substantial structural transformations and the unit cell parameter changes were tracked at different loadings (table S1). At its extreme, the MOF lattice changed from 14.5037(6) Å to 16.7259(7) Å in the *b* direction, and from 101.465(2)° to 105.091(2)° in β. We also conducted SCXRD measurements on the fully activated MOF at both 330 and 100 K and observed only incremental unit cell changes on the order of 0.01 Å. This result confirmed that these transitions were not caused by temperature variations but rather by water filling the pores.

In contrast, a larger difference was observed in the unit cell dimensions (up to 0.5 Å in the *b* direction) between the activated and fully loaded MOF at 100 K. Close examination of the hydrophilic pocket at different loadings revealed that the neighboring pyrazoles underwent adjustments to fit the water molecules (table S5). Specifically, the closest intermolecular N ⋮ N distance in the activated structure was 3.218(5) Å. Then, upon first and second water binding, this distance grew to 3.708(7) Å and 3.760(8) Å, respectively, as the pyrazole functionalities moved apart to accommodate the water molecules. At full water loading, the N ⋮ N distance peaked at 4.364(3) Å and allowed the formation of a three-dimensional water network. Complementary DFT calculations indicated an increase in framework strain at higher water loadings, which was, however, compensated by favorable H-bond interactions (text S4).

Having identified both experimentally and computationally that the step in the water sorption isotherm originated from water interactions with two neighboring pyrazole functionalities (sites I and II), we decided to control the H-bonding in the pocket by substituting $PZDC^{2-}$ with another linker that was less hydrophilic. We used 2,4-furandicarboxylic acid (H_2FDC) to synthesize a new MOF, MOF-*333* {[Al(OH)(FDC)], where FDC^{2-} = 2,4-furandicarboxylate; Fig. 4A}, isoreticular to MOF-*303* (*21*, *22*); that is, its rod-like SBUs consisted of cis-trans corner-shared AlO_6 octahedra. The sequence of the SBU was likely programmed by the similar angle between the carboxylic acid groups of the two linkers (158° for H_2PZDC and 156.5° for H_2FDC, as evidenced by SCXRD; text S3) and resulted in higher pore volumes, and consequently higher water uptakes, relative to other Al-MOFs built from rod-like SBUs (*23*).

In comparison to pyrazole, we expected furan to undergo weaker H-bonding with water molecules because it is less acidic and less basic, which is a consequence of the absence of acidic H atoms and the lower protonation propensity of the more electronegative O atom. Indeed, DFT calculations of the adsorption structures in MOF-*333* suggested a stronger interaction of the water molecules with the μ_2-OH group rather than the linker molecule, which was reflected in their respective interatomic distances (text S4). Furthermore, the computed binding energies indicated that water adsorption into the MOF-*303* pocket was substantially stronger than into the MOF-*333* pocket.

By following a SCXRD procedure similar to that described above for MOF-*303*, we identified the primary adsorption sites in MOF-*333* (text S3). The initially bound two water molecules (I′ and II′) could be distinguished by SCXRD analysis. They exhibited one strong H-bond to an individual μ_2-OH group with O_{OH} ⋮ O_{water} distances of 2.770(14) Å and 2.779(15) Å. In addition, water molecule I′ had only a very weak interaction with the furan linker with an O_{furan} ⋮ O_{water} distance of 3.01(2) Å, thus further confirming that this linker choice moderated the water sorption properties of MOF-*303* (Fig. 4A).

Next, we conducted a water sorption analysis on MOF-*333* and observed an ideally shaped water sorption isotherm with a steep step at 22% RH and without the step S observed for MOF-*303* (Fig. 4A). This value set a low RH cutoff at which our new MOF could operate. To extend its working range to more arid conditions, we prepared a series of multivariate MOFs consisting of both $PZDC^{2-}$ and FDC^{2-}. The nine synthesized MOFs covered the entire range of linker mixing ratios, including the two single-linker frameworks MOF-*303* and MOF-*333*. We use the nomenclature '*n/m*,' which describes the input ratio of $PZDC^{2-}$ to FDC^{2-} (*n* to *m*); for example, '*5/3*' denotes a 5-to-3 input ratio of $PZDC^{2-}$ to FDC^{2-} and '0/8' refers to the single-linker MOF-*333*.

Powder x-ray diffraction (PXRD) analysis revealed that all nine products were isostructural (text S5). The linker ratio for each MOF in the multivariate series, determined by nuclear magnetic resonance (NMR) analysis on completely base-hydrolyzed MOF crystals and elemental microanalysis, was nearly proportional to the input ratio (Fig. 4B and text S6). The close correspondence between the input and output ratios and the reproducible nature of this reaction chemistry indicated a robust reaction across all ratios—a behavior that may be called formulaic. Additionally, the presence and homogeneous distribution of $PZDC^{2-}$ in the multivariate MOF crystals and the absence of single-linker MOFs were verified by scanning electron microscopy coupled with energy-dispersive x-ray spectroscopy (text S7). Mapping of Al, C, N, and O showed homogeneous elemental distribution within the crystal outlines, and the overall intensity of the N signal continuously decreased with increasing substitution of $PZDC^{2-}$ by FDC^{2-} (fig. S34).

The thermal stability and porosity of our products were assessed by thermogravimetric analysis and N_2 sorption analysis, respectively. The MOF series exhibited no weight loss up to 375°C under N_2 and up to 325°C under air (text S8). The BET areas (1280 to 1360 m^2 g^{-1}), pore volumes (0.48 to 0.51 cm^3 g^{-1}), and diameters (~9.4 Å) extracted from N_2 sorption measurements on all nine compounds were comparable and consistent with an isostructural series of MOFs (text S9). The small variation in pore volumes stands in contrast to the conventional method of tuning MOF hydrophilicity involving the addition of functional groups to the framework backbone, which invariably reduces the free pore volume and subsequently the water uptake capacity (*11*, *24*, *25*).

We observed that the substitution of $PZDC^{2-}$ with increasing amounts of FDC^{2-} continuously shifted the water isotherm step toward

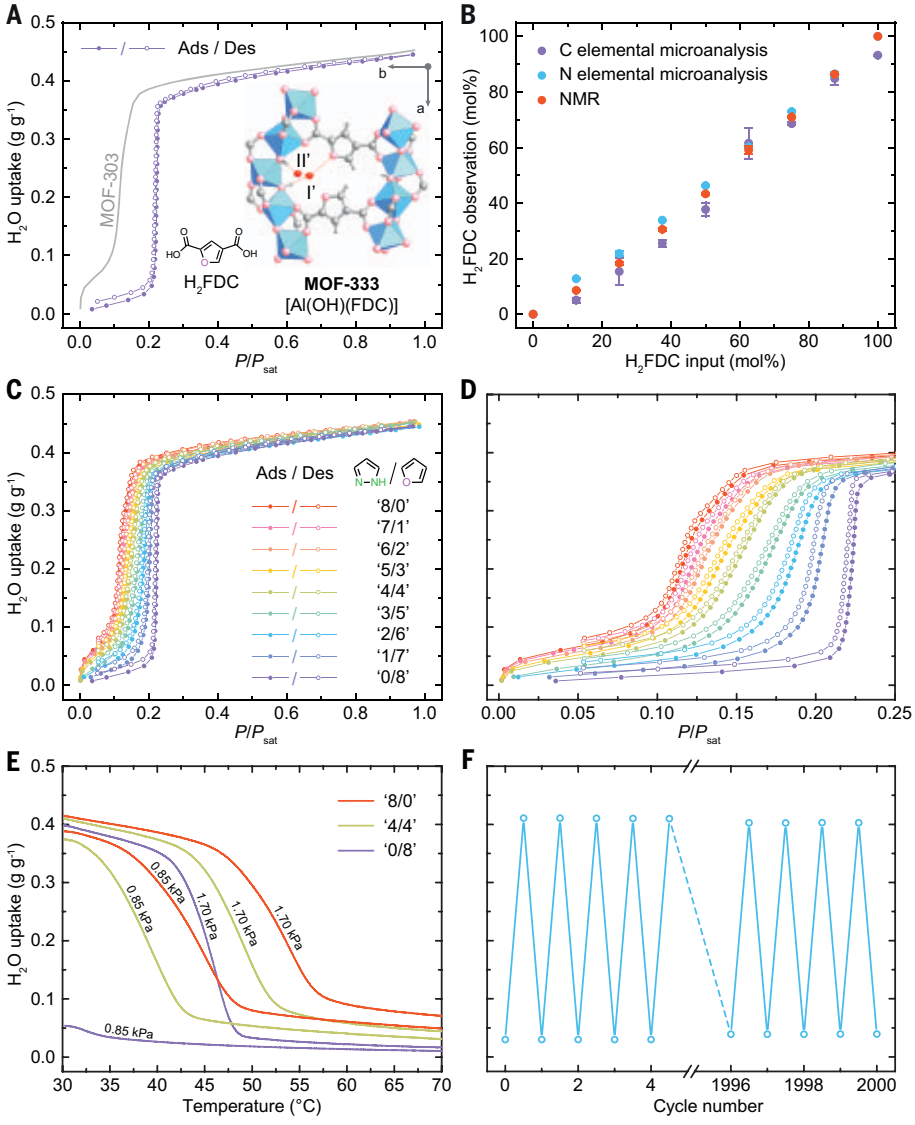


Fig. 4. Characterization of the multivariate MOF series. (**A**) Water sorption isotherm of MOF-333 at 25°C. The water sorption isotherm of MOF-303 is shown as a gray line for reference. The linker 2,4-furandicarboxylic acid (H_2FDC) and the pocket of MOF-333 with the first two adsorption sites (I′ and II′) are shown in the inset. Al, blue polyhedron; O in the framework structure, pink; O in H_2O, red; C and H, gray. For simplicity, only one orientation of the C/O positional disorder of FDC^{2-} in the MOF structure is shown. (**B**) H_2FDC linker ratios determined by NMR and C/N elemental microanalysis are plotted against the respective input linker ratios. Standard deviations (estimated from duplicate measurements on the same sample for elemental microanalysis and analysis of three separate samples for NMR) are depicted as error bars. (**C** and **D**) Full-range and low-pressure region of the water sorption analyses on the multivariate MOF series at 25°C. (**E**) Water vapor isobar measurements at 0.85 to 1.70 kPa for '8/0,' '4/4,' and '0/8.' (**F**) Water adsorption–desorption cycling of '4/4' between 30° and 85°C at 1.70 kPa for 2000 cycles of ~75 min each.

higher vapor pressures and covered the entire range between the water sorption isotherms of MOF-303 and MOF-333 (Fig. 4C). Furthermore, in accord with the analogous pore volumes of all MOF compounds, the maximal water uptake capacity was not compromised upon linker substitution. The initial step in the isotherm of MOF-303 was successfully mitigated with increasing $PZDC^{2-}$ substitution and its prominence disappeared at FDC^{2-} contents > 50%. All multivariate MOFs exhibited negligible hysteresis between the adsorption and desorption curves (Fig. 4D). This feature is desirable because hysteresis tends to reduce the working capacity and generally increases the energetic requirement for sorbent regeneration. The formulaic character of this multivariate series (Fig. 4B) and the continuous variation of the water sorption behavior allowed us to accurately design water sorption isotherms by simply choosing an appropriate input linker ratio.

Additionally, we assessed the impact of the multivariate strategy on the water adsorption enthalpies by extending our water sorption analysis to other temperatures (15°, 35°, and 45°C). Similar to the measurements at 25°C, the water sorption isotherms exhibited continuity within the multivariate series, lack of hysteresis, and similar maximal uptakes, which indicated consistent performance across different temperatures (text S10). From these isotherms, we estimated the differential adsorption enthalpy Δh_{ads} for all nine compounds using the Clausius-Clapeyron equation (text S10). With an increasing degree of FDC^{2-} incorporation, the average Δh_{ads} increased continuously from –53 to –50 kJ mol^{-1}. We note that the heat of condensation of water at 25°C is –44 kJ mol^{-1} and indicates the maximal possible enthalpy value for Δh_{ads}. Thus, the multivariate method allowed us to effectively decrease the adsorption heat penalty by up to 35%.

Moreover, our approach could be used to design water sorbents with ultralow desorption temperatures, as demonstrated with isobaric desorption curves at water vapor pressures between 0.85 and 1.70 kPa (corresponding to 20 to 40% RH at 30°C) for '8/0,' '4/4,' and '0/8' (text S10). The isobars exhibited a steep step that increased with higher vapor pressures. Notably, the minimal desorption temperature could be decreased up to 10°C by substituting $PZDC^{2-}$ with FDC^{2-} in the MOF structure (Fig. 4E). Furthermore, and in accordance with the isotherms, the working capacity increased with higher FDC^{2-} contents up to 15%. Both of these factors have a fundamental impact on energetic requirements and output per cycle of a water-harvesting system. Additionally, higher desorption temperatures restrict the number of cycles the water harvester can perform daily because longer heating and subsequent cooling times will be necessary to drive the higher temperature gradients.

Finally, we confirmed that mixing of $PZDC^{2-}$ and FDC^{2-} within one MOF crystal does not compromise the hydrolytic stability of the single-linker framework by exposing '4/4' to a water vapor pressure of 1.7 kPa and cycling the temperature between 30° and 85°C (Fig. 4F and text S10)—a test that reliably verifies the MOF's longevity (*7*, *26*). After 2000 uptake and release cycles, the sorbent retained ~97% of its working capacity.

REFERENCES AND NOTES

1. O. M. Yaghi, M. J. Kalmutzki, C. S. Diercks, *Introduction to Reticular Chemistry: Metal-Organic Frameworks and Covalent Organic Frameworks* (Wiley-VCH, 2019).
2. L. Öhrström, F. M. Amombo Noa, *Metal-Organic Frameworks* (American Chemical Society, 2021).
3. H. Kim *et al.*, *Science* **356**, 430–434 (2017).
4. H. Kim *et al.*, *Nat. Commun.* **9**, 1191 (2018).
5. F. Fathieh *et al.*, *Sci. Adv.* **4**, eaat3198 (2018).
6. N. Hanikel *et al.*, *ACS Cent. Sci.* **5**, 1699–1706 (2019).

7. N. Hanikel, M. S. Prévot, O. M. Yaghi, *Nat. Nanotechnol.* **15**, 348–355 (2020).
8. A. I. Kolesnikov *et al.*, *Phys. Rev. Lett.* **93**, 035503 (2004).
9. X. Lin *et al.*, *J. Am. Chem. Soc.* **128**, 10745–10753 (2006).
10. Y.-K. Seo *et al.*, *Adv. Mater.* **24**, 806–810 (2012).
11. H. Furukawa *et al.*, *J. Am. Chem. Soc.* **136**, 4369–4381 (2014).
12. J. Varlec *et al.*, *New J. Chem.* **40**, 4178–4186 (2016).
13. S. M. Towsif Abtab *et al.*, *Chem* **4**, 94–105 (2018).
14. M. Wahiduzzaman, D. Lenzen, G. Maurin, N. Stock, M. T. Wharmby, *Eur. J. Inorg. Chem.* **2018**, 3626–3632 (2018).
15. A. J. Rieth, K. M. Hunter, M. Dincă, F. Paesani, *Nat. Commun.* **10**, 4771 (2019).
16. V. Bon *et al.*, *J. Mater. Chem. A* **7**, 12681–12690 (2019).
17. N. C. Burtch *et al.*, *Nat. Chem.* **12**, 186–192 (2020).
18. D. F. Sava *et al.*, *Chem. Mater.* **25**, 2591–2596 (2013).
19. H. Deng *et al.*, *Science* **327**, 846–850 (2010).
20. L. M. Aguirre-Díaz *et al.*, *J. Am. Chem. Soc.* **137**, 6132–6135 (2015).
21. M. Eddaoudi *et al.*, *Science* **295**, 469–472 (2002).
22. V. Colombo *et al.*, *J. Am. Chem. Soc.* **134**, 12830–12843 (2012).
23. X. Liu, X. Wang, F. Kapteijn, *Chem. Rev.* **120**, 8303–8377 (2020).
24. H. Reinsch *et al.*, *Chem. Mater.* **25**, 17–26 (2013).
25. N. Ko *et al.*, *J. Mater. Chem. A* **3**, 2057–2064 (2015).
26. J. E. Mondloch *et al.*, *Chem. Commun.* **50**, 8944–8946 (2014).
27. N. Hanikel *et al.*, Computational results for the publication "Evolution of water structures in metal-organic frameworks for improved atmospheric water harvesting". Zenodo (2021); https://doi.org/10.5281/zenodo.5294977.

ACKNOWLEDGMENTS

N.H. thanks C. Gropp (Yaghi laboratory, UC Berkeley) and E. A. Kapustin (Water Harvesting Inc.) for fruitful discussions. S.C. and W.J. thank the Minnesota Supercomputer Institute for providing computational resources. L.G. thanks the Alexander von Humboldt Foundation for a Humboldt Research Award. We thank Z. Shi, T. Xu, and Y.-B. Zhang (ShanghaiTech University) for conducting preliminary SCXRD measurements; D. Moore and colleagues (GE Research) for invaluable discussions and support; and King Abdulaziz City for Science and Technology and UAE University for collaborations with UC Berkeley. **Funding:** This material is based on work supported by the Defense Advanced Research Projects Agency (DARPA) under contract HR001-21-C-0020. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of DARPA. This research used resources of the Advanced Light Source (beamline 12.2.1) at Lawrence Berkeley National Laboratory, which is a DOE Office of Science user facility under contract DE-AC02-05CH11231. Instruments in UC Berkeley's NMR facility in the College of Chemistry are supported in part by NIH grant S10OD024998. This study is also supported by the Studienstiftung des deutschen Volkes and a Kavli ENSI Philomathia Graduate Student Fellowship (N.H.) and the German Research Foundation (DFG) within a Reinhart Koselleck grant (J.S.). **Author contributions:** N.H. and O.M.Y. conceived the idea and led the experimental efforts; J.S. and L.G. led the computational analysis. N.H. synthesized the compounds, conducted sorption measurements as well as PXRD, NMR, and thermogravimetric analyses; N.H. and X.P. performed SCXRD experiments and solved and refined the crystal structures; S.C. and W.J. designed the models and conducted the DFT calculations; H.L. carried out SEM-EDS measurements; N.H. and O.M.Y. prepared the initial draft; and all authors contributed to revising the manuscript. **Competing interests:** O.M.Y. is co-founder of Water Harvesting Inc., aiming at commercializing related technologies. O.M.Y., N.H., and H.L. are inventors on a patent application (PCT/US2019/063442) describing analogous materials. **Data and materials availability:** The crystallographic data are archived at the Cambridge Crystallographic Data Centre under reference numbers CCDC 2078716 to 2078739, and the computational results are available on Zenodo (*27*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj0890
Materials and Methods
Supplementary Text
Figs. S1 to S88
Tables S1 to S61
References (*28*–*54*)

20 April 2021; accepted 1 September 2021
10.1126/science.abj0890

NANOCATALYSTS

Sulfur-anchoring synthesis of platinum intermetallic nanoparticle catalysts for fuel cells

Cheng-Long Yang[1]†, Li-Na Wang[2]†, Peng Yin[1]†, Jieyuan Liu[2]†, Ming-Xi Chen[1], Qiang-Qiang Yan[1], Zheng-Shu Wang[1], Shi-Long Xu[1], Sheng-Qi Chu[3], Chunhua Cui[4], Huanxin Ju[5], Junfa Zhu[5], Yue Lin[1]*, Jianglan Shui[2]*, Hai-Wei Liang[1]*

Atomically ordered intermetallic nanoparticles are promising for catalytic applications but are difficult to produce because the high-temperature annealing required for atom ordering inevitably accelerates metal sintering that leads to larger crystallites. We prepared platinum intermetallics with an average particle size of <5 nanometers on porous sulfur-doped carbon supports, on which the strong interaction between platinum and sulfur suppresses metal sintering up to 1000°C. We synthesized intermetallic libraries of small nanoparticles consisting of 46 combinations of platinum with 16 other metal elements and used them to study the dependence of electrocatalytic oxygen-reduction reaction activity on alloy composition and platinum skin strain. The intermetallic libraries are highly mass efficient in proton-exchange-membrane fuel cells and could achieve high activities of 1.3 to 1.8 amperes per milligram of platinum at 0.9 volts.

For catalysis applications, the defined surface and near-surface atomic arrangement of intermetallic nanoparticles (i-NPs) affords geometric and electronic properties that can enhance activity, selectivity, and stability (*1*–*5*). The regular structure of i-NPs also ensures homogeneity of active sites, which aids the investigation of structure-activity relations (*1*, *6*). Although ordered intermetallics are thermodynamically stable relative to disordered solid solutions, the realization of the disorder-to-order transition must overcome the kinetic energy barrier of atom ordering (Fig. 1A) (*2*).

Thus, preparing i-NP catalysts generally requires high temperatures to promote atomic diffusion and ordering, but high-temperature annealing also accelerates interparticle sintering (Fig. 1B), which creates larger i-NPs (>5 nm) with broader size distributions (*7*, *8*), lower specific surface areas, and low mass-based activities. Recent efforts aimed at preparing small i-NPs—particularly Pt-based materials for fuel cell application (*9*)—include coating with polymer or metal oxide protective shells before annealing (*10*, *11*), KCl matrix-assisted annealing (*12*), metal-organic framework–confined coreduction (*13*), and chemical vapor deposition with organometallic precursors (*14*). However, there is still no reported method for synthesizing libraries of supported sub–5 nm i-NPs catalysts, in which the composition and activity enhancement mechanism can be explored systematically.

We present a sulfur-anchoring approach for the synthesis of small Pt-based i-NP libraries (Fig. 1C). Strong chemical interaction between Pt and the sulfur atoms doped in the carbon matrix largely suppressed NPs sintering up to 1000°C, which enabled the formation of atomically ordered i-NPs with an average size of <5 nm at high temperatures. The synthetic capability of the sulfur-anchoring method allowed comprehensive exploration of structure-activity relations in i-NP libraries.

The porous S-doped carbon (S-C) supports were prepared by cobalt-assisted carbonization of molecular precursors with silica NPs as templates (*15*, *16*). The prepared S-C supports were composed of carbon (76.0 wt %), sulfur (14.0 wt %), and oxygen (9.0 wt %), accompanied by negligible cobalt and nitrogen (table S1). Sulfur was bonded in this matrix primarily as C-S-C, along with a small fraction of S-O_x species (fig. S1, A and B). The high surface area (1490 $m^2 g^{-1}$) and mesoporous structure would endow the S-C supports with abundant homogeneously distributed sulfur sites for metal NP anchoring (fig. S1, C and D, fig. S2, and table S1).

We tested the antisintering ability of S-C for monometallic Pt against high-temperature annealing. After annealing at high temperatures (up to 1000°C), the average diameter of Pt NPs was still <5 nm (fig. S3). Antisintering at such high temperatures has not been reported for carbon-supported metals (*17*), and we observed severe Pt sintering on commercial carbon black supports for the same annealing process (fig. S3). Sulfur-doped carbons

[1]Hefei National Laboratory for Physical Sciences at the Microscale, Department of Chemistry, University of Science and Technology of China, Hefei 230026, China. [2]School of Materials Science and Engineering, Beihang University, Beijing 100191, China. [3]Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China. [4]Molecular Electrochemistry Laboratory, Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu 610054, China. [5]National Synchrotron Radiation Laboratory, University of Science and Technology of China, Hefei 230026, China.
†These authors contributed equally to this work.
***Corresponding author. Email: hwliang@ustc.edu.cn (H.-W.L.); shuijianglan@buaa.edu.cn (J.S.); linyue@ustc.edu.cn (Y.L.)**

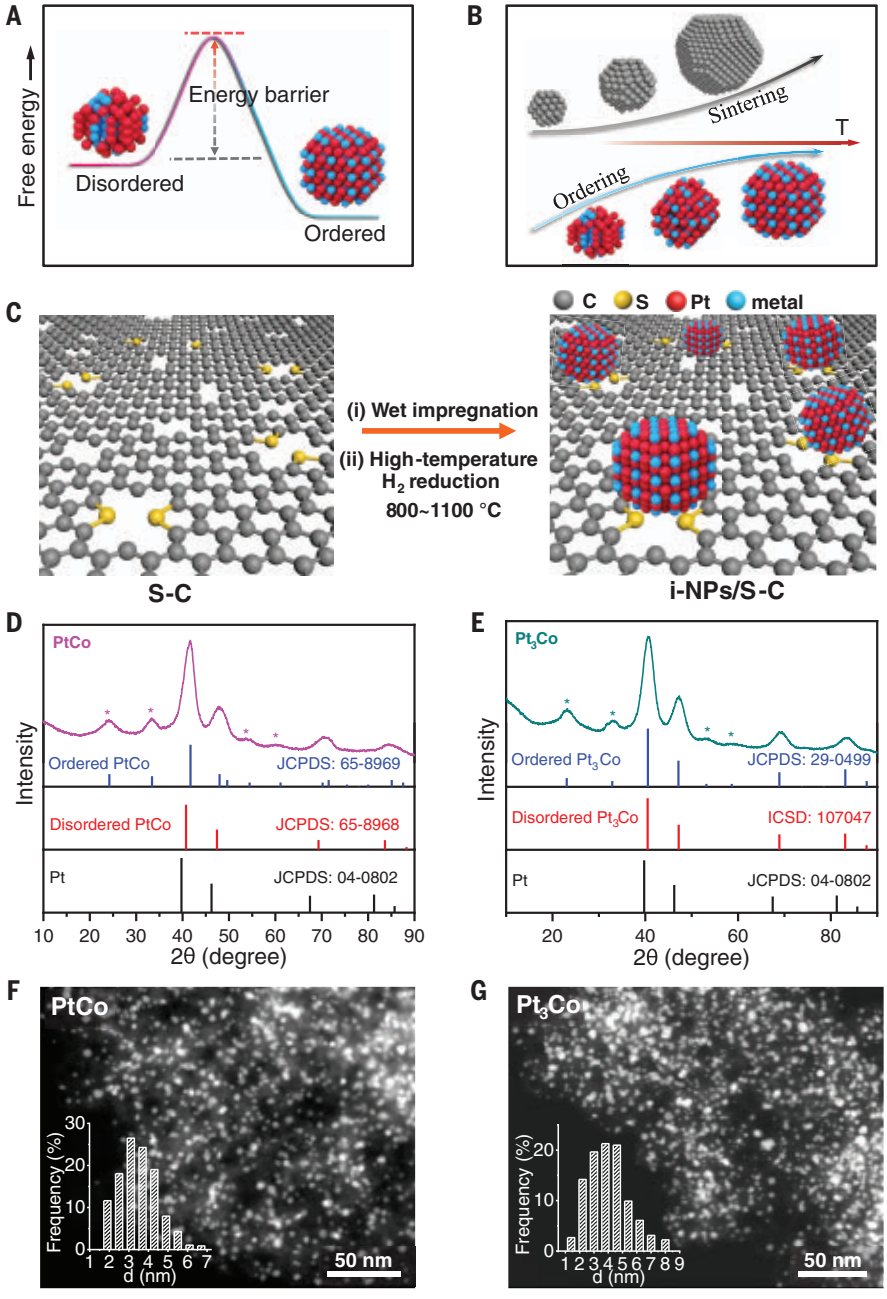


Fig. 1. Synthesis and structural characterization of Pt-based i-NPs on S-C. (**A**) Schematic illustration showing the kinetic energy barrier for atom ordering in the disorder-to-order transition. (**B**) Schematic illustration showing the dilemma of simultaneously accelerated sintering kinetics and atom ordering kinetics with temperature. (**C**) Schematic illustration of the high-temperature sulfur-anchoring synthetic approach. (**D** and **E**) XRD patterns of PtCo and Pt_3Co i-NPs, respectively. The standard peaks for Pt and ordered and disordered PtCo/Pt_3Co are also shown. Asterisks mark the characteristic superlattice peaks of ordered intermetallic structures. (**F** and **G**) HAADF-STEM images of PtCo and Pt_3Co i-NPs, respectively. The inserted histograms are the statistics of particle size distribution of corresponding i-NPs. d, diameter.

have been previously reported as supports for preparing atomically dispersed metal catalysts (*18*, *19*) and nanocluster catalysts (*20*), or for enhancing the electrochemical stability of Pt (*21*) and Pt-Co disordered alloys (*22*).

This antisintering property allowed us to prepare small intermetallic i-NPs through high-temperature annealing on S-C supports. We wet-impregnated metal salt precursors onto the S-C supports with total metal contents 20 wt%; the supported precursors were then subjected to one- or two-step high-temperature H_2-reduction treatment at 600° to 1100°C. We prepared 20 binary Pt-based i-NPs on the S-C support, including Pt alloyed with early transition metals (Pt_3Sc, Pt_3Ti, Pt_3V, Pt_3Cr, and Pt_3Zr), late transition metals (Pt_3Mn, Pt_3Fe, PtFe, Pt_3Co, PtCo, PtNi, PtCu, $PtCu_3$, Pt_3Zn, and PtZn), and p-block metals (Pt_3Al, Pt_3Ga, Pt_3Ge, Pt_3In, and Pt_3Sn) (fig. S4). These catalysts have broad applications in heterogeneous catalysis and electrocatalysis (*1*, *2*, *5*).

The x-ray diffraction (XRD) patterns indicated that the prepared i-NPs matched well with the Joint Committee on Powder Diffraction Standards (JCPDS) or the Inorganic Crystal Structure Database (ICSD) cards of corresponding ordered intermetallic phase, with the characteristic superlattice peaks of ordered intermetallic structures marked by asterisks (Fig. 1, D and E, and figs. S5 to S22). The crystalline structures of all 20 i-NPs were classified into four types, with M representing base metals (table S2): (i) all Pt_3M- and PtM_3-type i-NPs adopted a cubic structure in space group P*m-3m* (Cu_3Au structure), except for Pt_3Ge; (ii) all PtM type i-NPs adopted an ordered tetragonal structure in space group P4/*mmm*, except for PtCu; (iii) Pt_3Ge adopted a tetragonal structure in space group I4/*mcm*; and (iv) PtCu adopted a hexagonal structure in space group *R-3m*.

We quantitatively analyzed the ordering degree of each sample by comparing the intensity of superlattice reflections with the standard

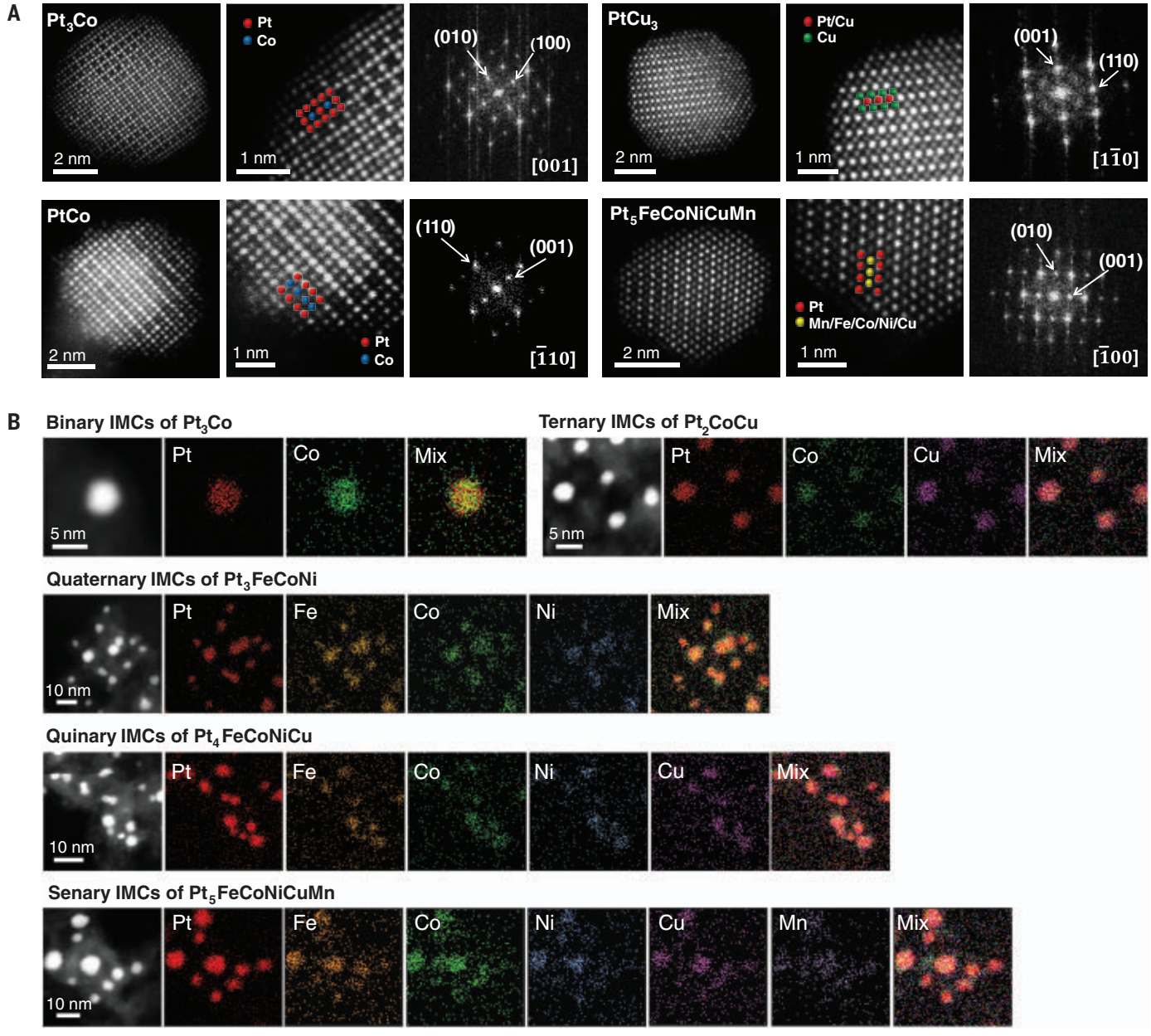


Fig. 2. Atomic-resolution HAADF-STEM images and EDS elemental mappings of representative i-NPs. (**A**) Atomic-resolution HAADF-STEM images and FFT patterns of Pt_3Co, PtCo, $PtCu_3$, and $Pt_5FeCoNiCuMn$. (**B**) EDS elemental mappings of Pt_3Co, Pt_2CoCu, $Pt_3FeCoNi$, $Pt_4FeCoNiCu$, and $Pt_5FeCoNiCuMn$.

powder XRD file card. The ordering degree of most samples (18 cases) was >50%, with only two cases showing relatively low ordering degree (35.9% for Pt_3V and 46.1% for PtNi) (table S2). The average NP size estimated by the Debye-Scherer equation, based on the full width at half maximum of XRD patterns, was <5 nm for all 20 samples. This result was consistent with statistical analysis of observations from high-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) (Fig. 1, F and G, figs. S5 to S22, and table S2). This route appears to be more effective for the synthesis of small and highly ordered intermetallic catalysts compared with the reported methods (table S3).

We performed a control experiment by depositing PtFe and Pt_3Co i-NPs on a desulfurized S-C support obtained by annealing a pristine S-C support at 1100°C in 5% H_2/Ar. After desulfurization treatment, the sulfur content decreased to 0.2 wt %, but the porous structure and high surface area remained almost unchanged (fig. S23 and table S1). The desulfurized S-C support showed little suppression of sintering of PtFe and Pt_3Co particles at high temperatures (figs. S24 and S25). We also observed large NPs (3 to 50 nm in diameter) and broad size distributions for PtFe and Pt_3Co i-NPs on commercial carbon black supports (figs. S26 to S29).

We used HAADF-STEM imaging to analyze the ordered structure of the i-NPs on an atomic scale. Atomic number (Z) contrast imaging of HAADF images reflects the Z-contrast of materials, so Pt columns will have a higher intensity than columns of metals with lower Z. For example, for Pt_3Co, we observed a structure of a periodic square array of Co (Z_{Co} = 27) columns surrounded by Pt (Z_{Pt} = 78) columns at the edges and corners of each unit cell along the [001]zone axis, corresponding to the $L1_2$ ordered structure (Fig. 2A and fig. S30). $PtCu_3$ also displayed an ordered $L1_2$ structure with alternating stacking of bright mixed Pt/Cu columns and dark Cu (Z_{Cu} = 29) columns along the $[1\bar{1}0]$ zone axis (Fig. 2A).

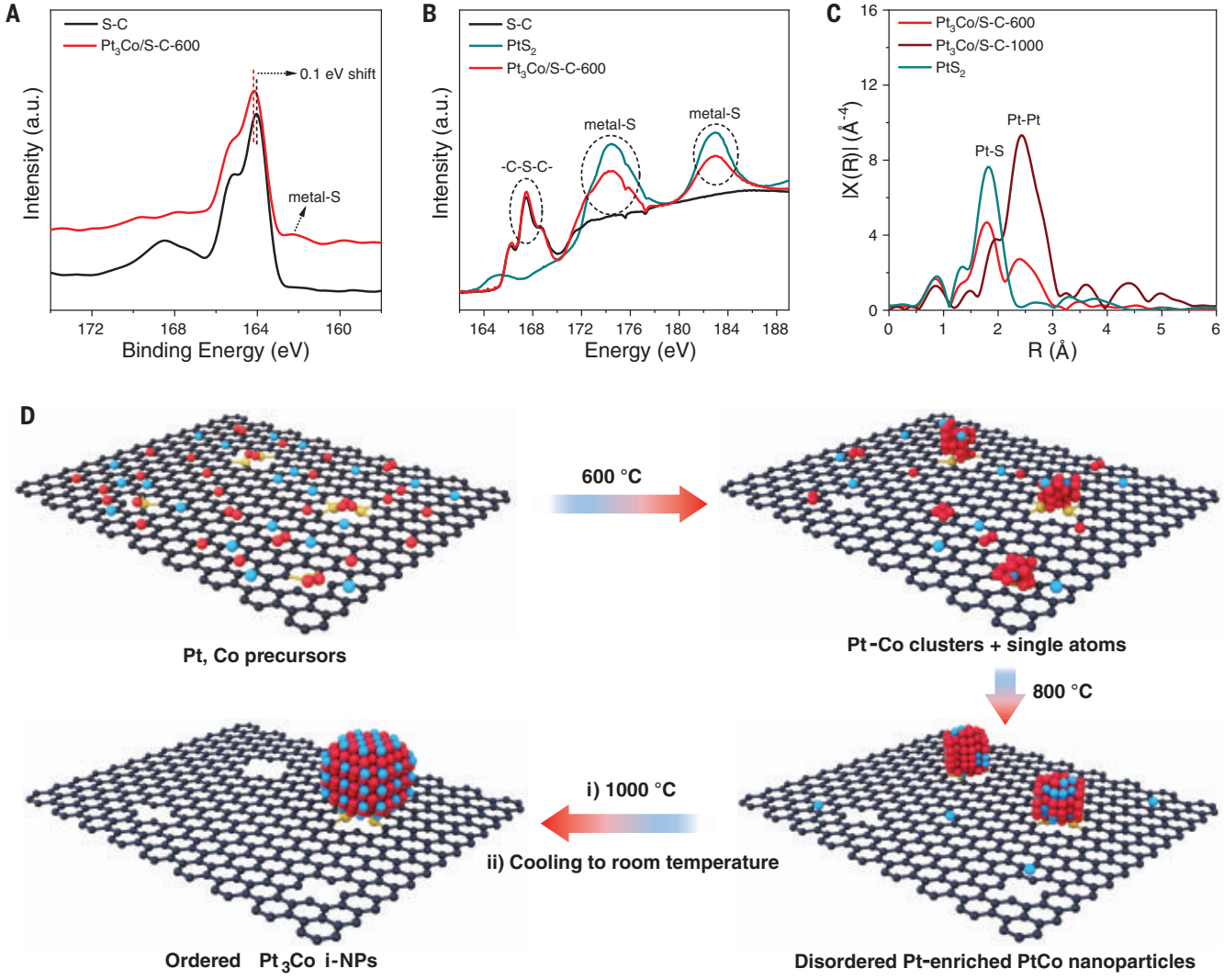


Fig. 3. Strong chemical Pt-S interaction and formation mechanism of i-NPs on S-C. (**A**) XPS spectra of Pt_3Co/S-C-600 and S-C, indicating the electronic interaction between metals and S-C. a.u., arbitrary units. (**B**) XANES spectra at the S L-edge of Pt_3Co/S-C-600, PtS_2, and S-C, indicating the formation of metal-S bonds. (**C**) Fourier-transformed EXAFS data at the Pt L_3-edge of Pt_3CoS-C-600, Pt_3Co/S-C-1000, and PtS_2, indicating the presence of Pt-S bonds in Pt_3Co/S-C. (**D**) Schematic illustration of the formation mechanism of small-sized ordered Pt_3Co i-NPs on S-C.

Alternating stacking of Pt and Co (or Fe) columns was observed for PtCo (or PtFe) i-NPs along the [$\bar{1}$10] or [010] zone axis, which we assigned to the $L1_0$ ordered structure (Fig. 2A and figs. S31 and S32). Periodic square array structures were also observed for $L1_2$-ordered Pt_3Al, Pt_3Sc, Pt_3Ti, and Pt_3V i-NPs (figs. S33 to S36). Fast Fourier transform (FFT) patterns and alternating intensity profiles further confirmed the ordered structures of these i-NPs (Fig. 2A and figs. S30 to S36). Moreover, elemental mapping with energy-dispersive x-ray spectroscopy (EDS) showed that Pt and non-Pt elements were distributed uniformly in individual i-NPs with an atomic percent approximately equivalent to the theoretical value (Fig. 2B and figs. S37 to S41).

The antisintering ability of S-C enabled synthesis of high total metal loading (60 wt %) of small PtFe i-NPs on S-C supports (fig. S42). In addition, we also synthesized small i-NPs on the sulfur-doped carbon black supports (fig. S43), demonstrating the universality of the sulfur-anchoring synthesis strategy. We further investigated the combinatorial synthesis of ordered Pt-based multicomponent i-NPs consisting of every combination of Pt with five metallic elements (Mn, Fe, Co, Ni, and Cu), including 10 ternary, 10 quaternary, 5 quinary, and 1 senary i-NPs. These multicomponent i-NPs all adopted an ordered tetragonal structure in space group P4/*mmm* with average NP size of <5 nm and high ordering degree (figs. S44 to S69 and table S4). *Z*-contrast HAADF images and FFT patterns revealed alternating stacking of Pt and non-Pt metal columns along the [$\bar{1}$00] or [010] zone axis (Fig. 2A and figs. S70 to S73), which we assigned to the $L1_0$ ordered structure, and EDS elemental mapping confirmed the homogeneous distribution of Pt and other metallic elements in individual multimetallic i-NPs (Fig. 2B, figs. S74 to S86). A density functional theory (DFT) simulation revealed that the quinary alloy with 50 at % Pt content also has a thermodynamic tendency toward formation of ordered intermetallic structures rather than a disordered solid solution (fig. S87), which is distinct from the reported high-entropy–alloy NPs prepared through high-temperature thermal shock (*23*).

We performed x-ray photoelectron spectroscopy (XPS), x-ray absorption near edge structure (XANES), and extended x-ray absorption fine structure (EXAFS) characterizations to investigate the chemical interaction between sulfur and Pt. Because of the low

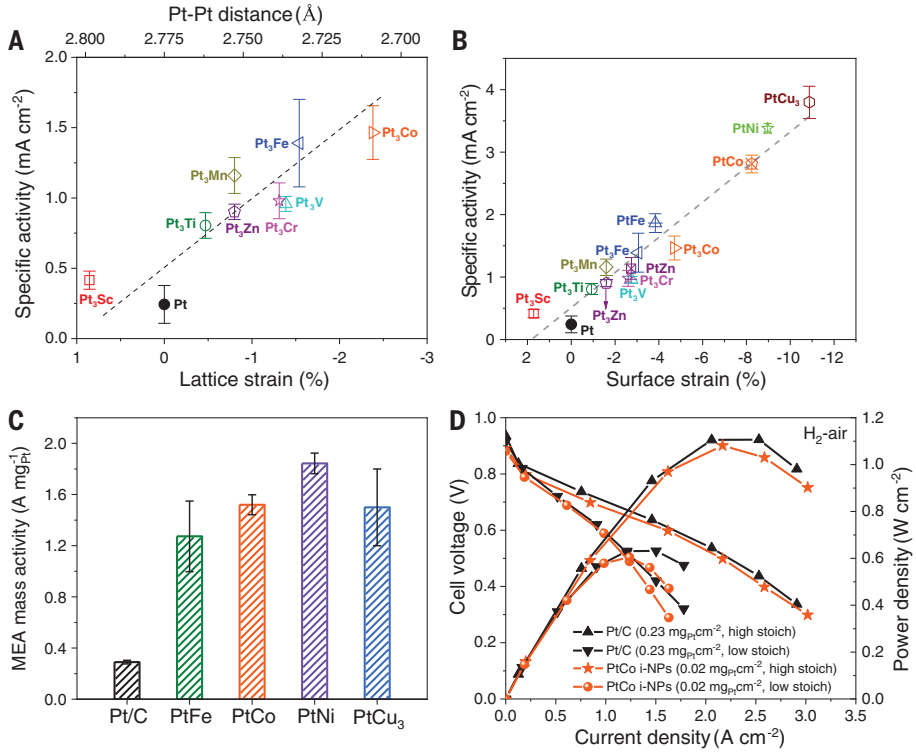


Fig. 4. Electrocatalytic performance and structure-activity relationship. (**A**) Specific activity at 0.9 V on various fcc $L1_2$ Pt_3M-i-NPs/Pt catalysts versus the in-plane Pt-Pt distance of Pt(111) and corresponding lattice strain. The negative value of strain indicates a compressive strain. (**B**) Specific activity at 0.9 V on the i-NPs/Pt catalysts covering all 10 3d transition metal elements versus surface strain (the area change of intermetallic structure relative to bulk Pt). (**C**) Cathodic mass activity of the membrane electrode assemblies (MEAs) made with the cathode catalysts of Pt/C and i-NP catalysts. The anode catalyst is Pt/C. Test conditions: 80°C, 100% relative humidity, 150 $kPa_{abs,\ outlet}$ H_2–O_2; H_2 and O_2 flow rates were fixed at 0.2 and 0.5 liters min^{-1}, respectively. (**D**) H_2-air polarization curve and power density curve of the MEAs made with the cathode catalysts of Pt/C and PtCo i-NPs. Cathode loadings (milligrams of Pt per square centimeter) are indicated. The anode catalyst is Pt/C, with a loading of 0.02 mg_{Pt} cm^{-2} for all MEAs. Test conditions: 80°C, 100% relative humidity, 250 $kPa_{abs,\ outlet}$ H_2–air at high stoichiometries (H_2 and air flow rates were fixed at 0.5 and 2.0 liters min^{-1}, respectively) and low stoichiometries of 1.5 and 1.8. The error bars in (A to C) show the standard deviation in at least two independent electrochemical measurements for each datum.

sulfur content (<1.0 wt %) of the ordered Pt_3Co/S-C prepared at 1000°C (table S1 and fig. S88), we measured the S XPS and L-edge XANES of Pt_3Co/S-C prepared at 600°C. In XPS spectra (Fig. 3A), the S 2p peak shifted to higher binding energy by 0.1 eV compared with the pristine S-C support, indicating the electron transfer from sulfur to metal. A new peak emerged at around 162 eV that we associated with a metal-S bond (*24*). For the S L-edge XANES (Fig. 3B), we assigned two peaks at 174.3 and 182.9 eV of Pt_3Co/S-C to the metal-S bond through comparison with the spectrum of PtS_2. Pt L_3-edge EXAFS results indicated the coexistence of Pt-S and Pt-Pt (Co) bonds in both Pt_3Co/S-C samples (600° and 1000°C). Metallic Pt elements dominated the ordered Pt_3Co/S-C sample prepared at 1000°C (Fig. 3C), which was confirmed through EXAFS fitting (figs. S89 and S90). These characterization data verified the presence of Pt-S bonds in Pt_3Co/S-C, which could increase the adhesion energy of i-NPs on the S-C support and suppressed interparticle sintering by retarding the Ostwald ripening (surface atom diffusion between different particles), particle migration and coalescence, or both (*17*, *25*).

We systematically characterized the Pt_3Co samples obtained at annealing temperatures that were gradually increased from 300° to 1000°C with ex situ HAADF-STEM, EDS elemental mapping, and XRD to understand the formation process of i-NPs on S-C. At 300°C, numerous single metal atoms and some clusters consisting of loosely assembled atoms observed in HAADF-STEM evolved into ~1-nm metal clusters with poor crystallinity at 600°C, accompanied by a notable decrease in the number of single atoms (fig. S91). At 800°C, metal NPs formed, which gradually became substantially crystalline, along with the emergence of a few partially ordered NPs. For annealing at 1000°C, well-ordered $L1_2$-type Pt_3Co i-NPs were formed. The XRD analyses confirmed the evolutionary process of the Pt_3Co structure with annealing temperatures (fig. S92), from poorly crystalline Pt-enriched alloy (600°C), to substantially crystalline but disordered alloy (800°C), and finally to highly ordered Pt_3Co intermetallic structures (1000°C).

Through EDS mapping, we found that both Pt and Co were homogeneously distributed throughout the S-C support at 300°C, and that most of the Pt—but much less of the Co—was concentrated on the cluster regions for the 600°C sample (fig. S93). With higher annealing temperatures (800° and 1000°C), the amount of Co distributed on the particle-free regions gradually decreased, and Co was concentrated on the NPs. Quantitative EDS analyses further confirmed the increased Co/Pt atomic ratios in individual NPs with annealing temperatures, gradually approaching the feed ratio. These EDS results demonstrate that high-temperature annealing is crucial for generating PtCo alloys with a targeted stoichiometric Co/Pt ratio (1/3 for $L1_2$ Pt_3Co i-NPs). The sulfur content decreased with the rise in annealing temperature, and by 1000°C almost all of the residuary sulfur element (~0.5 wt %) was enriched in the NP regions (fig. S93). Similar sulfur enrichment was also observed in other i-NPs samples (figs. S39 to S41 and figs. S74 to S86). This spatial relationship between sulfur and metals was consistent with the doped sulfur atoms serving as anchoring sites for i-NPs even at high temperatures (*21*).

On the basis of the above additional HAADF-STEM, EDS elemental mapping, and XRD analyses, combined with the XPS, XANES, and EXAFS data, we propose a sulfur-anchoring formation mechanism of small-ordered Pt_3Co i-NPs on S-C (Fig. 3D). First, H_2PtCl_6 was reduced by H_2 at a low temperature (<400°C) to form Pt clusters anchored by doped sulfur in S-C through Pt-S bonds. At an elevated temperature (600° to 800°C), Co was reduced and alloyed with Pt to form disordered Pt-Co NPs. However, the content of alloyed Co atoms in most of the individual NPs was not high enough to meet the targeted stoichiometric Co/Pt ratio (that is, 1/3) for the $L1_2$ Pt_3Co intermetallic structures. After high-temperature annealing at 1000°C, the Co/Pt ratio of alloy NPs finally approached the required stoichiometric ratio and formed the atomically ordered Pt_3Co intermetallic structure after cooling below the phase-transition temperature through the thermodynamic-driven disorder-order transition (see supplementary text and figs. S94 and S95). Note that the doped sulfur atoms gradually departed from the carbon matrix as the annealing temperature increased (fig. S96), which led to slight sintering as a result of NP

detachment from sulfur-leaving sites. The average size of ordered Pt_3Co i-NPs was still <5 nm, which could be ascribed to the primarily chemical confinement of residual sulfur sites enriched around NPs on the S-C support.

The combinatorial synthesis of i-NPs catalysts with high composition diversity enabled a comprehensive exploration of a structure and performance relationship in a systematic manner. We systematically studied the electrocatalytic performance of the i-NPs libraries for an oxygen reduction reaction (ORR) in proton-exchange membrane fuel cells (PEMFCs) (*26*), including the 14 binary Pt 3d transition metal i-NPs catalysts spanning all 10 of the 3d transition metal elements. The as-prepared i-NPs catalysts were treated first with acid then with an H_2-annealing treatment at 400°C to form the highly stable and active i-NP/Pt core/shell structures, with an intermetallic core and two to three atomic layers of Pt shell (fig. S97) (*27*, *28*).

These i-NP/Pt catalysts exhibited high mass activity and specific activity at 0.9 V of up to 4.18 A mg_{Pt}^{-1} and 3.8 mA cm_{Pt}^{-2}, respectively, as revealed by the rotating disk electrode measurements (figs. S98 and S99). We plotted the measured specific activities of all the Pt_3M $L1_2$-type face-centered cubic (fcc) intermetallic catalysts (including Pt_3Sc, Pt_3Ti, Pt_3V, Pt_3Cr, Pt_3Mn, Pt_3Fe, Pt_3Co, and Pt_3Zn) as a function of the in-plane Pt-Pt distance of the (111) plane and corresponding compressive lattice strain (Fig. 4A). We used DFT to calculate the adsorption energy of oxygen [ΔE(*O)] of the (111) surface modeled by a slab composed of two intermetallic bottom layers, which were fixed, and three top Pt layers, which were relaxed. The DFT simulation showed good agreement between the calculated adsorption energy and the experimentally measured ORR activity for all the $L1_2$-type Pt_3M catalysts (fig. S100), with a widely accepted optimal oxygen adsorption energy, which is weaker than that on pure Pt by ~0.2 eV (*29*).

We further recognized that all $L1_0$-type PtM face-centered tetragonal (fct) intermetallics showed higher ORR activity than the corresponding $L1_2$-type Pt_3M intermetallics, such as PtFe versus Pt_3Fe, PtCo versus Pt_3Co, and PtZn versus Pt_3Zn (fig. S101). Different from the fcc $L1_2$ intermetallics with isotropic strain, the fct $L1_0$ structures had anisotropic strains and changed the in-plane Pt-Pt distance along different directions (fig. S102). We attempted to correlate the structure-activity of all of the fcc $L1_2$ and fct $L1_0$ intermetallics by calculating the two-dimensional surface strain (the area change of intermetallic structure relative to bulk Pt; table S5). Rather than the theoretically predicted trend of a volcano relationship, we observed a monotonic relationship between ORR activity and surface strain for all of the fcc and fct intermetallic catalysts in a very wide range, up to ~11% compressive surface strain for $PtCu_3$ (Fig. 4B). The discrepancy between experimental results and theoretical prediction is likely related to the compressive-strain relaxation in the Pt shells (*30*). The Pt layers of the shell should exhibit gradually increased relaxation and decreased strain from inside to outside toward the lattice constant of bulk Pt.

These systematic studies suggested that all of the studied i-NPs catalysts still did not exhibit sufficiently high true surface strain to access the maximum of the volcano relationship. Further promotion of the surface compressive strain by shortening the lattice parameter of the intermetallic core may achieve record-breaking catalytic activity, although more relaxation probably occurs when the strain surpasses a critical point (*6*).

We screened the libraries of i-NPs for PEMFCs applications. Four i-NPs/Pt catalysts were selected as the cathode catalysts for single-cell tests. The PtFe, PtCo, PtNi, and $PtCu_3$ i-NPs catalysts exhibited mass activities of 1.27, 1.52, 1.84, and 1.50 A mg_{Pt}^{-1} at 0.9 volts (internal resistance corrected) in H_2-O_2 cell tests (Fig. 4C), which were much higher than Pt/C (0.29 A mg_{Pt}^{-1}) and exceeded the US Department of Energy 2025 target of 0.44 A mg_{Pt}^{-1}, and versus those reported for intermetallic catalysts such as PtFe (0.21 A mg_{Pt}^{-1}) (*27*), PtCo (0.56 A mg_{Pt}^{-1}) (*28*), Pt_3Co (0.52 A mg_{Pt}^{-1}) (*31*), and Pt_3Co/Co-N-C composite catalyst (1.77 A mg_{Pt}^{-1}) (*32*). The durability of the i-NPs catalysts was evaluated through an accelerated stress test (AST) in a PEMFC. After 30,000 AST cycles between 0.6 and 1.0 V, PtCo and PtFe i-NP catalysts retained 77 and 79% of their initial mass activity, respectively (fig. S103). The intermetallic core-shell structure of the catalyst was well retained after AST (fig. S104).

In an H_2-air PEMFC, the i-NP catalyst cathodes with an ultralow Pt loading of 0.02 mg_{Pt} cm^{-2} performed similarly to the Pt/C cathode whose Pt loading (0.23 mg_{Pt} cm^{-2}) was higher by a factor of 11.5 times under both high- and low-stoichiometry conditions (Fig. 4D and fig. S105). Specifically, the PtCo i-NPs cathode with ultralow Pt loading achieved a maximum power density of 1.08 W cm^{-2} at high stoichiometries (Fig. 4D). The performance at low stoichiometries could be enhanced by improving the local oxygen transport properties; for example, through engineering of porous structures and surface functionality of carbon supports (*33*, *34*).

REFERENCES AND NOTES

1. E. Antolini, *Appl. Catal. B* **217**, 201–213 (2017).
2. Y. Yan *et al*., *Adv. Mater.* **29**, 1605997 (2017).
3. M. Armbrüster *et al*., *Nat. Mater.* **11**, 690–693 (2012).
4. M. Miyazaki, S. Furukawa, T. Komatsu, *J. Am. Chem. Soc.* **139**, 18231–18239 (2017).
5. S. Furukawa, T. Komatsu, *ACS Catal.* **7**, 735–765 (2016).
6. M. Escudero-Escribano *et al*., *Science* **352**, 73–76 (2016).
7. J. T. L. Gamler, H. M. Ashberry, S. E. Skrabalak, K. M. Koczkur, *Adv. Mater.* **30**, e1801563 (2018).
8. D. Wang *et al*., *Nat. Mater.* **12**, 81–87 (2013).
9. X. X. Wang, M. T. Swihart, G. Wu, *Nat. Catal.* **2**, 578–589 (2019).
10. J. Kim, Y. Lee, S. Sun, *J. Am. Chem. Soc.* **132**, 4996–4997 (2010).
11. D. Y. Chung *et al*., *J. Am. Chem. Soc.* **137**, 15478–15485 (2015).
12. H. Chen *et al*., *J. Am. Chem. Soc.* **134**, 18453–18459 (2012).
13. X. X. Wang *et al*., *Nano Lett.* **18**, 4163–4171 (2018).
14. D. S. Choi, A. W. Robertson, J. H. Warner, S. O. Kim, H. Kim, *Adv. Mater.* **28**, 7115–7122 (2016).
15. H. W. Liang *et al*., *Nat. Commun.* **6**, 7992 (2015).
16. Z. Y. Wu *et al*., *Sci. Adv.* **4**, eaat0788 (2018).
17. Y. Dai, P. Lu, Z. Cao, C. T. Campbell, Y. Xia, *Chem. Soc. Rev.* **47**, 4314–4331 (2018).
18. C. H. Choi *et al*., *Nat. Commun.* **7**, 10922 (2016).
19. L. Wang *et al*., *Sci. Adv.* **5**, eaax6322 (2019).
20. P. Yin *et al*., *Nat. Commun.* **12**, 3135 (2021).
21. D. Higgins *et al*., *Adv. Funct. Mater.* **24**, 4325–4336 (2014).
22. R. Ahmadi, M. Amini, J. Bennett, *J. Catal.* **292**, 81–89 (2012).
23. Y. Yao *et al*., *Science* **359**, 1489–1494 (2018).
24. C. Kim *et al*., *J. Am. Chem. Soc.* **137**, 13844–13850 (2015).
25. S. L. Hemmingson, C. T. Campbell, *ACS Nano* **11**, 1196–1203 (2017).
26. M. K. Debe, *Nature* **486**, 43–51 (2012).
27. J. Li *et al*., *J. Am. Chem. Soc.* **140**, 2926–2932 (2018).
28. J. Li *et al*., *Joule* **3**, 124–135 (2019).
29. J. Greeley *et al*., *Nat. Chem.* **1**, 552–556 (2009).
30. P. Strasser *et al*., *Nat. Chem.* **2**, 454–460 (2010).
31. Y. Xiong *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **116**, 1974–1983 (2019).
32. L. Chong *et al*., *Science* **362**, 1276–1281 (2018).
33. V. Yarlagadda *et al*., *ACS Energy Lett.* **3**, 618–621 (2018).
34. S. Ott *et al*., *Nat. Mater.* **19**, 77–85 (2020).

ACKNOWLEDGMENTS

We thank X.-J. Wu, H.-F. Lv, H.-X. Ji, F.-J. Fan, and H.-B. Yao from USTC for the useful discussions, X. Han from USTC for drawing the structural model of i-NPs, and J. Liu from Anhui Contango New Energy Technology Co., Ltd. for assistance with MEA fabrication. **Funding:** H.-W.L. acknowledges support from the National Key Research and Development Program of China (2018YFA0702001), the National Natural Science Foundation of China (22071225), and the Fundamental Research Funds for the Central Universities (WK2060190103). J.S. acknowledges support from Natural Science Foundation of Beijing Municipality (Z200012). Y.L. acknowledges support from the Youth Innovation Promotion Association of the CAS (2020458) and the National Natural Science Foundation of China Excellent Young Scientists Fund (52122212). **Author contributions:** H.-W.L. conceived and designed the experiments. C.-L.Y., P.Y., M.-X.C., Q.-Q.Y., Z.-S.W., and S.-L.X. synthesized and characterized the i-NPs. Y.L. performed the aberration-corrected TEM characterization. L.-N.W., J.L., P.Y., and J.S. performed the PEMFC tests. J.L. and J.S. performed the DFT simulation. S.-Q.C., H.J. and J.Z. performed the x-ray absorption spectra measurements. H.-W.L., J.S., C.-L.Y., L.-N.W., P.Y., J.L., and C.C. cowrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. Information requests should be directed to the corresponding authors.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj9980
Materials and Methods
Supplementary Text
Figs. S1 to S105
Tables S1 to S5
References (*35*–*68*)

15 June 2021; accepted 9 September 2021
10.1126/science.abj9980

STRUCTURAL MATERIALS

Lightweight, strong, moldable wood via cell wall engineering as a sustainable structural material

Shaoliang Xiao[1]†, Chaoji Chen[1]†, Qinqin Xia[1]†, Yu Liu[1]†, Yuan Yao[2], Qiongyu Chen[3], Matt Hartsfield[4], Alexandra Brozena[1], Kunkun Tu[5,6], Stephen J. Eichhorn[7], Yonggang Yao[1], Jianguo Li[1], Wentao Gan[1], Sheldon Q. Shi[8], Vina W. Yang[9], Marco Lo Ricco[9], J.Y. Zhu[9], Ingo Burgert[5,6], Alan Luo[4], Teng Li[3], Liangbing Hu[1,10]*

Wood is a sustainable structural material, but it cannot be easily shaped while maintaining its mechanical properties. We report a processing strategy that uses cell wall engineering to shape flat sheets of hardwood into versatile three-dimensional (3D) structures. After breaking down wood's lignin component and closing the vessels and fibers by evaporating water, we partially re-swell the wood in a rapid water-shock process that selectively opens the vessels. This forms a distinct wrinkled cell wall structure that allows the material to be folded and molded into desired shapes. The resulting 3D-molded wood is six times stronger than the starting wood and comparable to widely used lightweight materials such as aluminum alloys. This approach widens wood's potential as a structural material, with lower environmental impact for buildings and transportation applications.

The shape of a material can be as important as its intrinsic properties (*1*–*4*). For example, structural components must be made of materials that can be physically formed to fit specific needs without sacrificing mechanical strength. Furthermore, materials that are also lightweight are particularly valuable for vehicle-based applications (e.g., cars, trains, and aircraft) (*5*), as weight-saving measures may be the most direct way to improve fuel-efficiency (*6*). For these reasons, polymers and some metals, such as aluminum (Al), are used for mechanical support, as they feature both low density and facile processability into lightweight structural components of various shapes and sizes by different means such as extrusion, casting, and injection molding (*5*, *7*). However, the development of more-sustainable materials is needed to mitigate the environmental costs of petrochemical-based plastics and the energy costs of metals (*8*–*11*).

Wood is one potential candidate for replacing some of these lightweight structural materials, especially given its advantage of being a renewable resource (*12*, *13*). Wood products can help mitigate climate change by replacing carbon-intensive, fossil fuel–based materials (*14*, *15*), as well as providing extended carbon storage (*16*). Wood is also mechanically strong, lightweight, and potentially low cost. Various approaches have been demonstrated to improve the properties and functionality of wood for a wider range of applications, including by delignification (*17*–*19*), densification (*20*–*22*), and other modifications [e.g., thermal treatments (*23*, *24*), organic or inorganic salt impregnation (*25*–*27*), etc.]. That said, its formability is generally poor compared with metals and plastics, which makes processing it into complex shapes difficult.

Different methods of forming wood for shape-specific needs have been explored with varying success. For example, wood can be sculpted into complex three-dimensional (3D) shapes using conventional subtractive manufacturing (e.g., carving, sculpting, turning) (*28*, *29*) and conventional carpentry, wherein pieces of wood are joined to make a more complex structure (*30*). However, these "physical approaches" generally engineer wood at the bulk scale, which does not change the intrinsic microstructure or material properties, thus failing to simultaneously achieve high mechanical strength and good formability, and limiting the practical applications of wood in advanced engineering fields that require both. In recent years, various bottom-up approaches have also been investigated in which wood is broken down into its constituent components and then refabricated into desired shapes and uses. For example, wood can be defibrillated into cellulose nanofibers (CNFs) with exceptional mechanical strength (up to 3 GPa) (*31*), which can then be processed into 3D forms using high-water-content slurries (up to 98 wt %) (*32*). However, CNFs are expensive, and the energy required for water removal is too intensive for the fabrication of large structures. Alternatively, adding polymers to CNFs can improve their processability by enabling casting or injection molding (*33*). But the high petrochemical polymer content of these CNF composites diminishes their advantage as a sustainable material. Furthermore, such bottom-up approaches come at the sacrifice of wood's naturally hierarchical and anisotropic structure (i.e., the highly aligned channels and fibers that run along the longitudinal direction of the stem) from where much of the material's natural strength and functionality derives.

We report a top-down approach by which wood can be processed into various shapes while also substantially increasing its mechanical strength. The principal concept is based on partially delignifying and softening natural wood, then shrinking its vessels and fibers by drying, followed by "shocking" the material in water to selectively open the vessels (Fig. 1A). This rapid water-shock process forms a distinct partially open, wrinkled cell wall structure that provides space for compression as well as the ability to support high strain, allowing the material to be easily folded and molded. The different shapes and structures that can be achieved with this moldable wood can then be set with air drying to remove the remaining water to form the final 3D-molded wood product (Fig. 1B).

This cell wall engineering process maintains the intrinsic anisotropic wood structure and enhances the interactions among the wood fibers to further increase the mechanical strength. For example, we demonstrated the fabrication of a honeycomb core material from wood veneers (produced using roll-to-roll rotary cutting) that has a tensile strength of ~300 MPa, similar to Al alloys but with a density of just ~0.75 g/cm^3 and at potentially lower cost (Fig. 1, B and C). This cell wall engineering approach is applicable to various hardwood species and is enabled by the water-shock process that expands the capabilities of wood as a lightweight structural material beyond conventional planar structures to complex 3D designs and components with greater versatility. Manipulating wood's naturally anisotropic structure will improve its potential as a sustainable and practical alternative for applications that have been traditionally limited to polymers, metals, and polymer composites (Fig. 1C) (*34*).

We first used a common aqueous-based delignification process (*35*, *36*) to remove ~55% of the lignin and ~67% of the hemicellulose from the lignocellulosic cell walls of basswood (Fig. 1A and figs. S1 to S3). The partial and selective removal of this hydrophobic component results in softening and mild expansion in the wood sample size due to the absorption of water by the remaining cell walls (fig. S1), which feature a higher fraction

[1]Department of Materials Science and Engineering, University of Maryland, College Park, MD 20742, USA. [2]Center for Industrial Ecology, Yale School of the Environment, Yale University, New Haven, CT 06511, USA. [3]Department of Mechanical Engineering, University of Maryland, College Park, MD 20742, USA. [4]Department of Materials Science and Engineering, The Ohio State University, Columbus, OH 43210, USA. [5]Wood Materials Science, Institute for Building Materials, ETH Zürich, 8093 Zürich, Switzerland. [6]WoodTec Group, Cellulose & Wood Materials, Empa, 8600 Dübendorf, Switzerland. [7]Bristol Composites Institute, CAME School of Engineering, University of Bristol, University Walk, Bristol BS8 1TR, UK. [8]Department of Mechanical Engineering, University of North Texas, Denton, TX 76203, USA. [9]US Department of Agriculture (USDA) Forest Products Laboratory, Madison, WI 53726, USA. [10]Center for Materials Innovation, University of Maryland, College Park, MD 20742, USA.
*Corresponding author. Email: binghu@umd.edu
†These authors contributed equally to this work.

of hydrophilic cellulose (fig. S2). As a result, the water content of the partially delignified wood is ~300 wt % (fig. S4). We then air dried the partially delignified wood under ambient conditions for ~30 hours to remove water and form the shrunken wood intermediate (~12 wt % water content) (Fig. 1A and fig. S4). Next, we immersed the shrunken wood in water for 3 min in what we called a water-shock process that partially re-swells the cell walls and leads to some expansion of the sample dimensions in the resulting product, which we call moldable wood (~100 wt % water content) (figs. S1 and S4 and movie S1). Whereas the wet natural wood (Fig. 2A) and shrunken wood (Fig. 2B) cannot be bent without breaking, the moldable wood is highly foldable (Fig. 2C and movie S2). The wood fibers in these samples run parallel to the folding direction. Wood sheets made with this fiber orientation can be scalably produced by rotary cutting, circumventing the size limitations of the length and width of the raw tree material (*37*).

Using scanning electron microscopy (SEM), we examined the microstructures of these wood samples to better understand their process-structure-property relationships. The

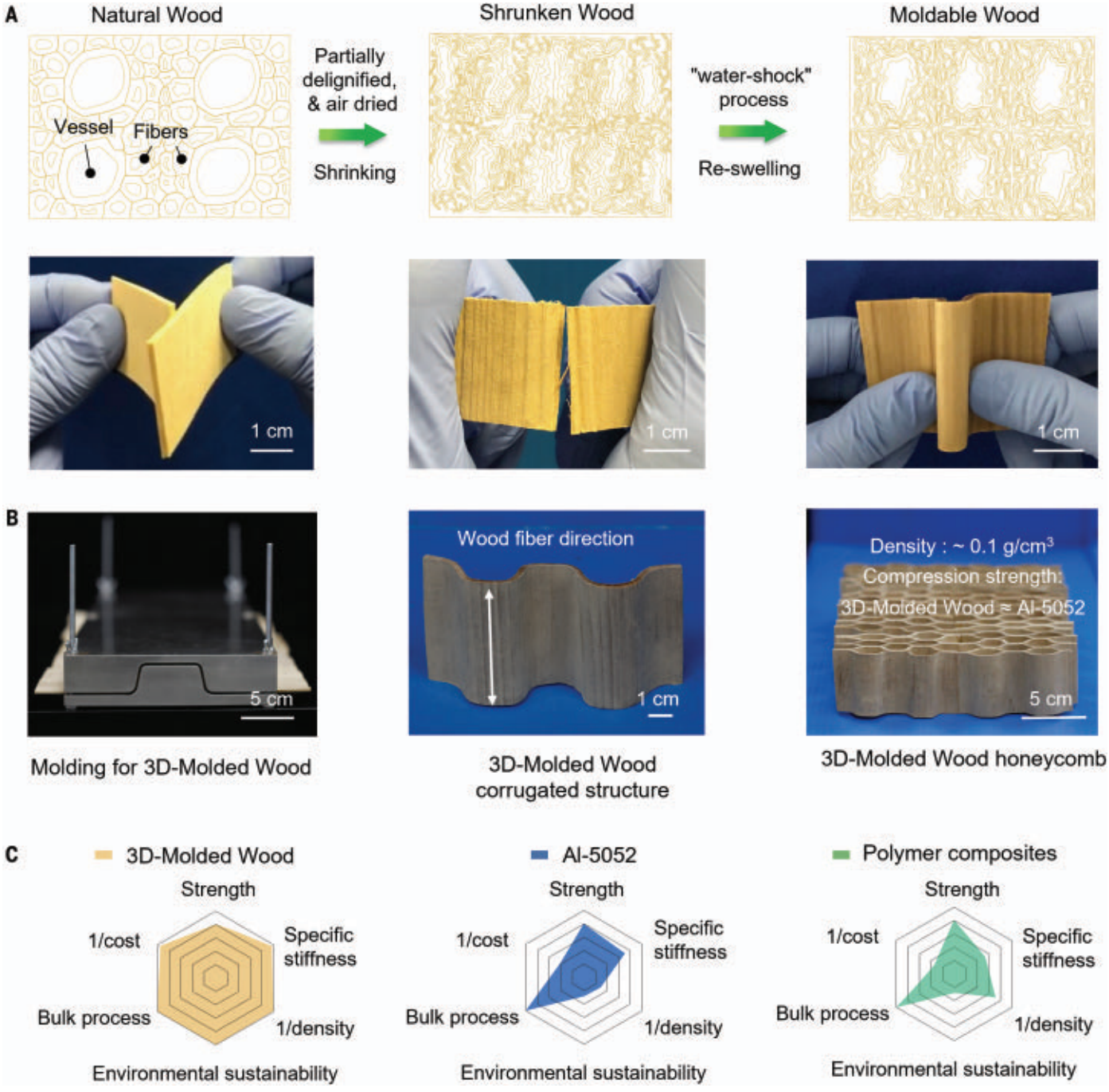


Fig. 1. A wrinkling cell wall engineering strategy through partial delignification combined with a "water-shock" process to make wood stronger and moldable. (**A**) Schematic illustration of the wood cell wall structures during the fabrication of the moldable wood, highlighting the critical role of the water-shock process, which forms wrinkled cell walls. (**B**) Corrugated and honeycomb 3D-molded wood structures made from wood veneers cut parallel to the fiber direction. (**C**) Radar plots comparing the performance of 3D-molded wood, Al-5052 alloy, and polymer composites (*34*), in which the results are normalized by the maximum value of each characteristic.

natural wood starting material has a 3D hierarchically porous cellular structure with many hollow vessels and fibers (Fig. 2D). These open cells become almost completely closed in the shrunken wood owing to the removal of lignin and water, forming a highly densified structure (Fig. 2E). However, the water-shock treatment that forms the moldable wood results in a distinctive wrinkled cell wall structure in which the vessels are partially open while the fibers are almost completely closed (Fig. 2F). During this treatment, we observed that the reopening of the vessels occurs extremely rapidly (3 s), while the morphology of the smaller fibers remains nearly unchanged (movie S3 and fig. S5). Such selective opening of the cell wall structure is noteworthy, as it may provide two simultaneous effects. The partially open vessels create space within the moldable wood that may accommodate both compressive and tensile deformation in an accordion-like manner, allowing the material to undergo severe compression and tension while being folded, even up to 180° without cracking (Fig. 2C and movie S2). Meanwhile, the densely packed closed fibers may provide mechanical support for enhanced strength.

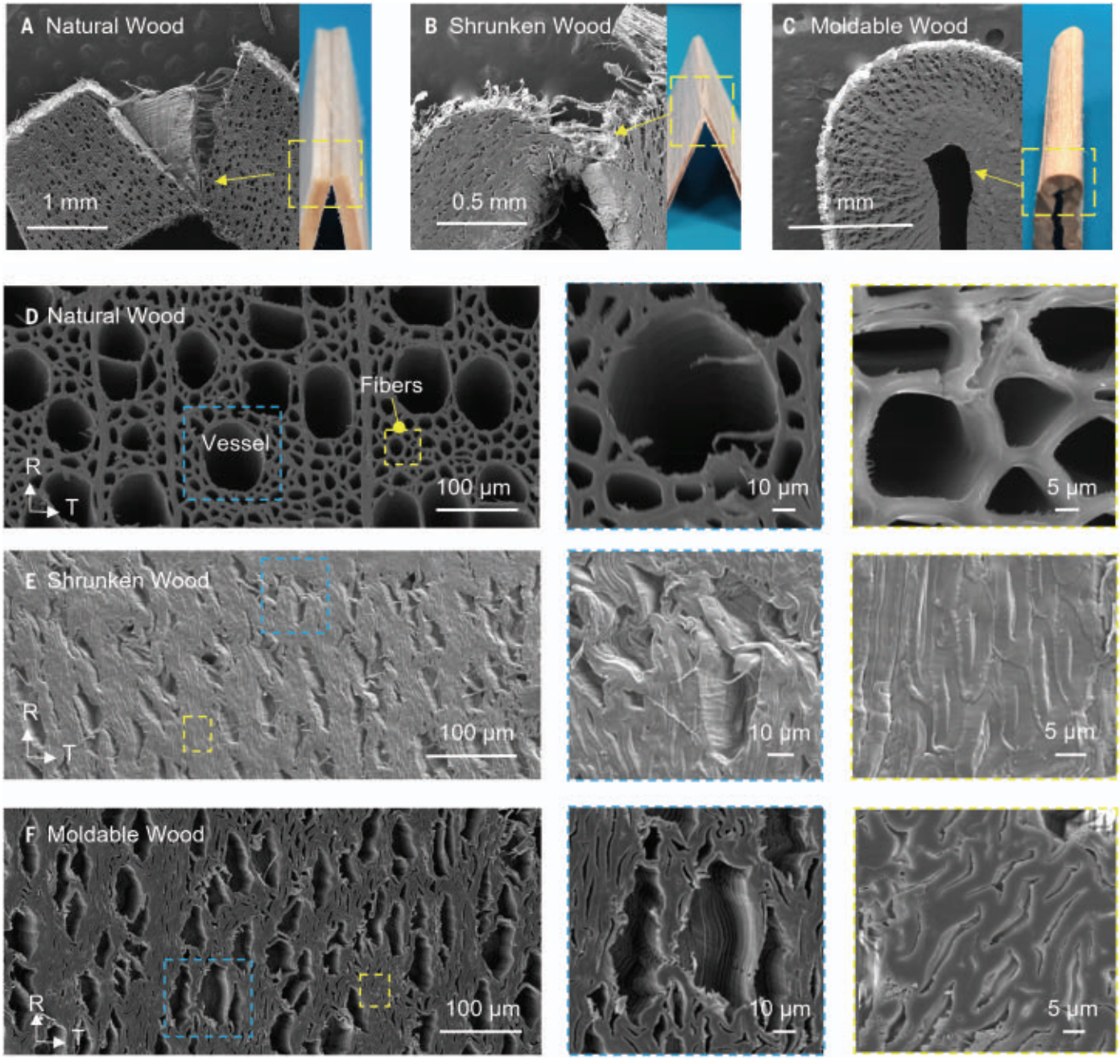


Fig. 2. Microstructures of the natural wood, shrunken wood, and moldable wood. (**A** to **C**) Photographs and SEM images demonstrate the (A) natural wood and (B) shrunken wood undergo deformation and fracture when bent, whereas (C) the moldable wood can be easily folded. (**D**) SEM images of the natural wood, which contains many hollow rigid vessels and fibers. (**E**) SEM images of the shrunken wood, which show the severely shrunken vessels and fibers after water removal. (**F**) SEM images of the moldable wood, which features a distinctive wrinkled cell wall structure with partially open vessels and almost completely closed fibers. R, radial direction; T, tangential direction.

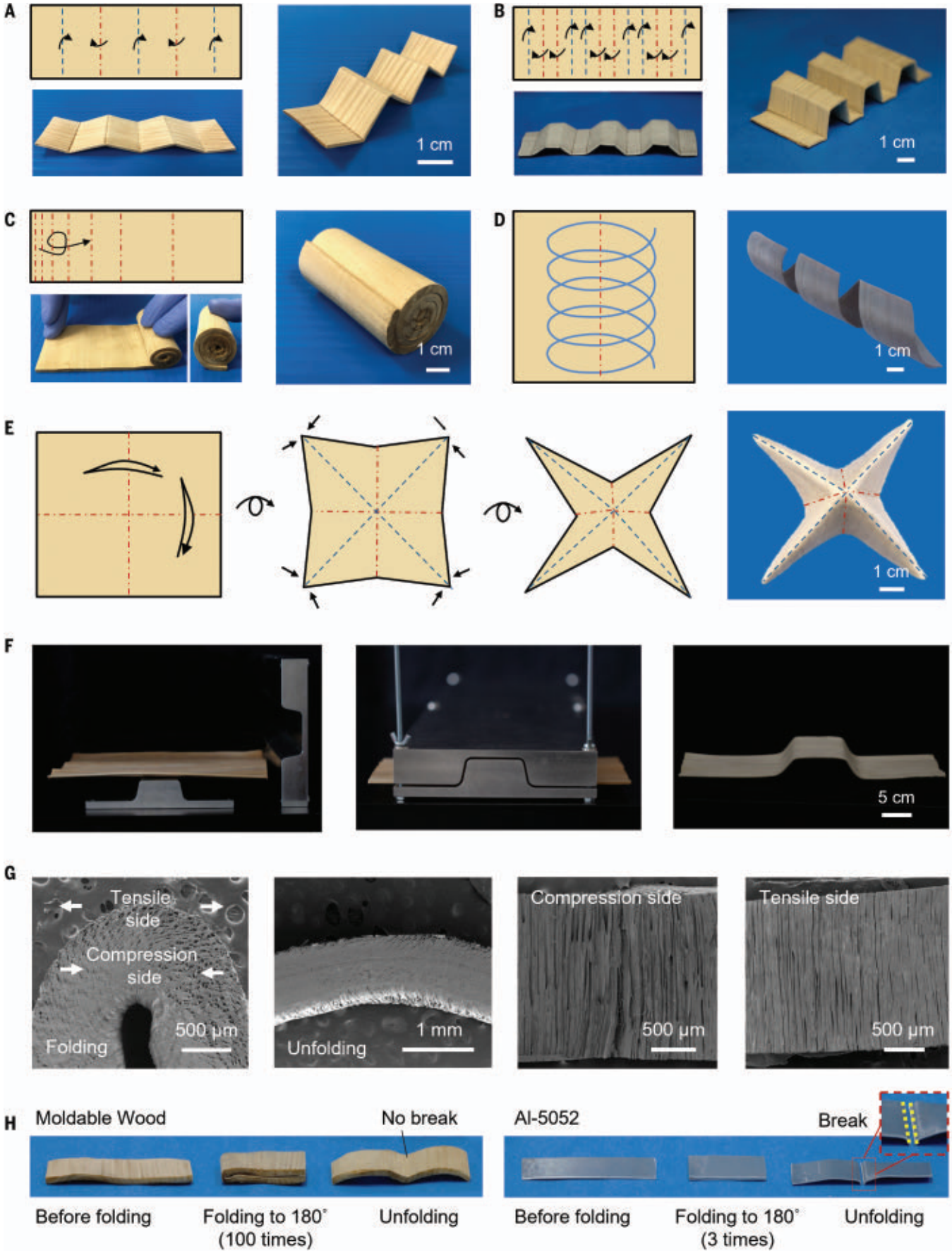

A
1 cm
B
1 cm
C
1 cm
D
1 cm
E
1 cm
F
5 cm
G
Tensile side
Compression side
Folding
500 μm
Unfolding
1 mm
Compression side
500 μm
Tensile side
500 μm
H
Moldable Wood
No break
Before folding
Folding to 180°
(100 times)
Unfolding
Al-5052
Break
Before folding
Folding to 180°
(3 times)
Unfolding

Fig. 3. Fabricating 3D-molded wood into various shapes using the cell wall engineering process. Schematics and corresponding photographs of the 3D-molded wood featuring (**A**) zigzag, (**B**) corrugated, (**C**) rolled, (**D**) helical, and (**E**) star-shaped structures. (**F**) The molding process of the 3D-molded wood corrugated structure (25 cm by 12 cm by 0.12 cm). (**G**) SEM images of the folded and unfolded 3D-molded wood, produced after air drying the moldable wood after 100 folding and unfolding cycles. We observed no crease in the compression or tensile sides in the resulting 3D-molded wood, demonstrating the excellent foldability and durability of the material when folded parallel to the wood fiber direction. (**H**) Comparison of the Al-5052 alloy and moldable wood after undergoing multiple cycles of folding and unfolding. The moldable wood remains intact after 100 cycles of folding and unfolding, whereas the Al alloy breaks after just three cycles.

To further investigate the impact of this wrinkled cell wall structure, we prepared a control in which we air dried the partially delignified wood for just 6 hours and did not apply the water-shock treatment in order to obtain the same water (~100 wt % water content) and lignocellulosic content as that of the moldable wood (fig. S4). The method we used to fabricate this control is similar to that of a previously reported high-strength wood material (*22*). The resulting control did not feature a wrinkled cell wall structure, instead demonstrating a similar open cell microstructure as the natural wood but with thinner and more separated cell walls owing to the partial removal of the lignin (fig. S6). Additionally, the material could not be bent without breaking (fig. S4), therefore we called it nonmoldable wood. This control clearly demonstrates the importance of the partially opened, wrinkled cell wall structure in enabling the moldable wood's mechanical flexibility.

We conducted simulations to investigate how the wrinkled cell wall structure enables the release of mechanical stress during folding to prevent the material from breaking. Fiber-scale mechanics modeling shows that the strain level in all cell walls of the moldable wood is extremely low (with a maximum principal tensile strain of 0.23% and compressive strain of 0.31%) even when the moldable wood is subjected to a 60% nominal strain deformation (tensile or compressive) (figs. S7 and S8). In contrast, the maximum principal tensile strain in the cell walls of the nonmoldable wood is as high as 2.3% under an overall elongation of 12.5%, substantially higher than that in the moldable wood (fig. S8). We explain this difference in the mechanical properties of the moldable and nonmoldable wood by their distinct microstructures. Despite both materials undergoing the delignification process, the fibers and vessels are open and only in loose contact in the nonmoldable wood, while the cell structures are more closed and in greater contact in the wrinkled cell walls of the moldable wood because of the drying and water-shock process. As a result, the moldable wood features sufficient hydrogen bonding among the cell walls to resist delamination during folding, while the nonmoldable wood readily fractures (fig. S9).

Using this cell wall engineering method, we can process the moldable wood into various shapes by mechanical bending, folding, and twisting. When the targeted structure is achieved, we can then dry the wood to fix its shape. For example, we fabricated zigzag and corrugated wood structures by repeatedly folding pieces of moldable wood and then drying the material to set the rigid forms (Fig. 3, A and B). Additionally, we can roll the moldable wood (Fig. 3C) and twist it (Fig. 3D, fig. S10, and movie S4) like metals and plastics. The outstanding foldability and excellent stability after drying allows us to design and fabricate complex 3D structures, such as star shapes (Fig. 3E). Other shapes include a corrugated structure (25 cm by 12 cm by 0.12 cm) using a homemade die to mold the material (Fig. 3F).

SEM morphology studies reveal the dense and integrated structure of the resulting 3D-molded wood. We did not observe any fibers peeling away from the folded corners of these structures, even after 100 folding and unfolding cycles (Fig. 3G). The surface finish remains smooth when folded parallel to the wood fiber direction. This excellent surface finish is likely due to the outstanding flexibility of the wood fibers and the hydrogen bonding between them, allowing the fibers to follow the folded shapes during mechanical deformation in a conformal manner (Fig. 3G). In terms of foldability and durability, the moldable wood also outperforms Al-5052, a widely used Al alloy for lightweight engineering structures, in which the moldable wood can be folded and unfolded 100 times without fracture, whereas the Al alloy broke after only three folding and unfolding cycles (Fig. 3H).

We found that the 3D-molded wood also has improved mechanical properties for lightweight structural applications, including a tensile strength of ~300 MPa and compressive strength of 60 MPa along the wood fiber direction. These values are nearly six and two times higher than that of the raw natural wood, respectively (fig. S11). The improved mechanical properties of the 3D-molded wood are due to its denser structure, which features highly packed intertwined cell walls at the microscale and well-aligned cellulose nanofibrils inside the cell walls at the nanoscale (fig. S12). The 3D-molded wood's low density of 0.75 g/cm^3 gives it a high specific tensile strength of 386.8 MPa per g/cm^3, which is about five times greater than that of Al-5052 (84.4 MPa per g/cm^3) (fig. S13). Furthermore, the stiffness of 3D-molded wood exceeds that of a range of hardwoods (*38*), softwoods (*38*), and polymers (*39*) (Fig. 4A).

The low density, high mechanical strength, and excellent formability of the 3D-molded wood offers broad versatility in designing and manufacturing large, lightweight, load-bearing designs. One example is a honeycomb structure, which is usually made of polymers or metals such as the Al-5052 alloy (*40*, *41*). We fabricated 3D-molded wood honeycomb structures by molding the material along the wood fiber direction (fig. S14). The 3D-molded wood honeycomb unit demonstrates comparable specific compressive strength (51.6 MPa per g/cm^3) as the Al-5052 honeycomb unit (46.8 MPa per g/cm^3) (Fig. 4B), although the 3D-molded wood features the added advantages of potentially being lower cost (table S1) and with potentially enhanced environmental sustainability. To evaluate the compressive and bending properties of the assembled 3D-molded wood honeycomb core, we sandwiched the structure between two Al plates (fig. S15). The Al–3D-molded wood–Al honeycomb core sandwich demonstrated a compressive strength of 9.1 MPa (Fig. 4C) and a specific compressive strength of 91.0 MPa per g/cm^3, which is higher than that of the Al-5052 honeycomb structure (specific compressive strength of 70.9 MPa per g/cm^3) (fig. S16). This sandwich structure was able to support the weight of a 1588-kg car, which corresponds to 1526 times the weight of the 3D-molded wood honeycomb core itself (Fig. 4D).

Scaling the manufacturing of corrugated 3D-molded wood structures could be accomplished using the mature roll-to-roll manufacturing processes from the wood and pulp industries (Fig. 4E) (*42*). For example, rotary cutting produces large wood sheets, which can be treated by the partial delignification, drying, and water re-swelling steps (fig. S17). All of these processes are common in the wood and pulp industries. Additionally, the 30-hour ambient drying step can be shorted to ~2 min using hot air (~80°C) without affecting the material's moldability. Rolling or die forming can then complete the fabrication of 3D-molded wood corrugated structures. Additional gluing and assembling steps can create the desired form, a 80 cm by 6 cm by 1.5 cm 3D-molded wood honeycomb core being one example (Fig. 4F).

In order to replace certain traditional structural materials, such as Al alloys (*40*), the long-

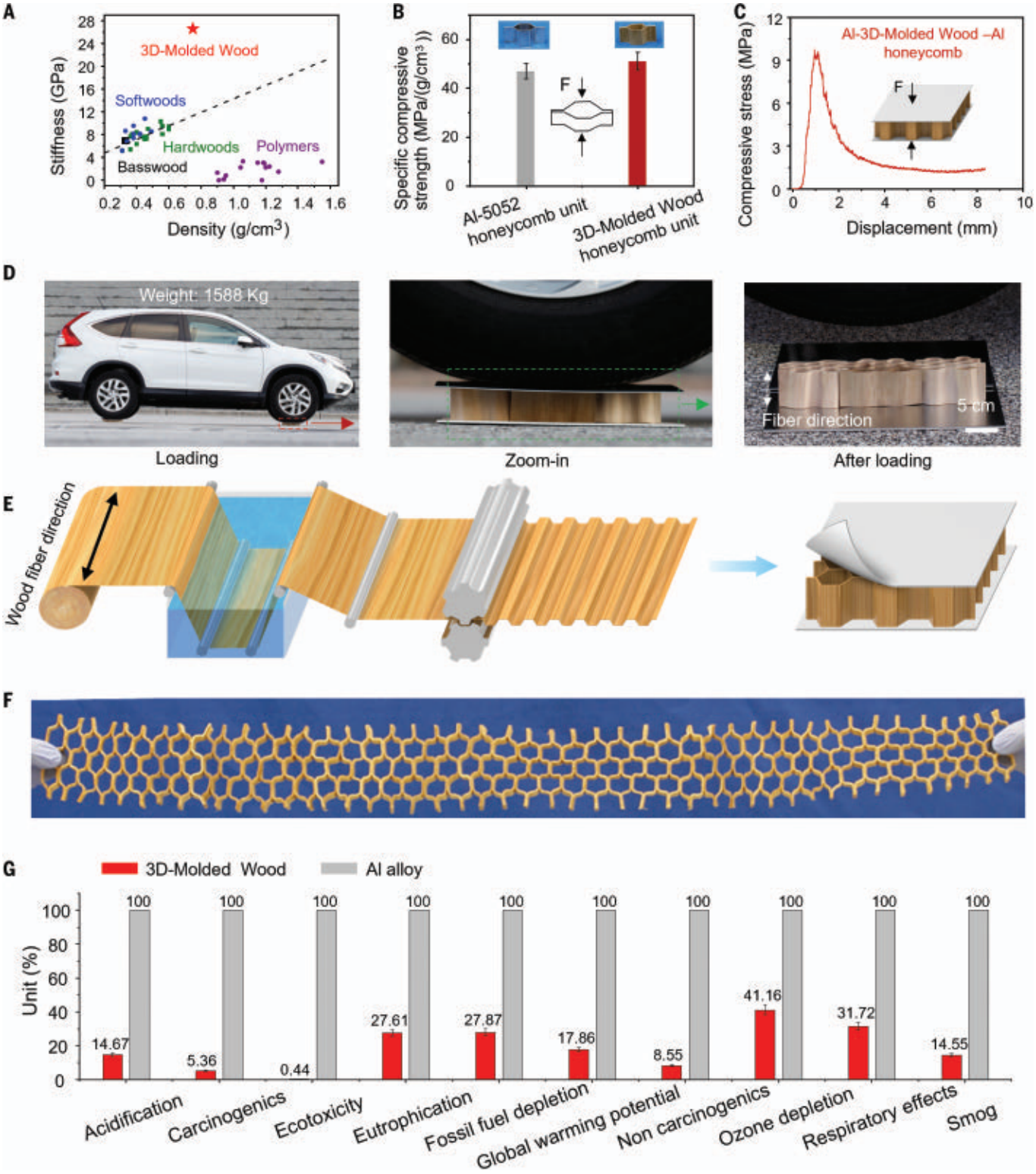


Fig. 4. Mechanical properties and LCA of the 3D-molded wood for strong, lightweight structural designs. (**A**) An Ashby diagram showing the stiffness versus density of 3D-molded wood, as well as various wood species (*38*) and polymers (*39*). (**B**) Comparison of the specific compressive strength of the 3D-molded wood and Al-5052 honeycomb units (1.5 cm in height by 2 cm in width and 4.4 cm in length). (**C**) The compressive stress versus displacement plot of the Al–3D-molded wood–Al honeycomb core sandwich structure. The inset shows the compression test schematic. The 3D-molded wood honeycomb core was 1.5 cm in height by 6.5 cm in length and 6 cm in width. (**D**) Photographs of the sandwich structure supporting the weight of a car, in which the core structure remained intact after loading. (**E**) Manufacturing process schematic for scaling the fabrication of the 3D-molded wood honeycomb core and its potential applications. (**F**) Photograph of a 3D-molded wood honeycomb core, 80 cm in length. (**G**) The environmental impacts of 3D-molded wood and Al alloy (high recycling content of 73.9%) per cubic centimeter per megapascal based on the specific tensile strength, normalized to the higher-impact material for each environmental impact category.

term durability of 3D-molded wood is necessary. Applying a surface coating of polyurethane to the 3D-molded wood can help stabilize it against moisture (fig. S18). Moreover, both the uncoated and coated 3D-molded wood demonstrate antifungal properties superior to those of natural wood (table S2 and fig. S19). However, more weathering tests are required to determine the exact types of modifications needed to ensure the long-term durability of the 3D-molded wood product. Ensuring uniform production during scale-up and specific performance tests when integrated into various system applications will also help determine the best use for the 3D-molded wood.

In addition to its excellent functionality, the 3D-molded wood is directly fabricated from a bio-sourced, renewable material (i.e., hardwood species), potentially providing enhanced environmental sustainability (e.g., a lower carbon footprint) compared with traditional metal and polymer-based structural materials. To investigate this potential, we performed a cradle-to-gate life cycle assessment (LCA) to understand the comparative environmental impacts of 3D-molded wood and Al alloy (table S3 and fig. S20). We calculated the ratio of the environmental impacts per cubic centimeter of the materials to their tensile strength (megapascals) and stiffness (gigapascals) (tables S4 and S5). Our study included different environmental impact categories, such as those related to ecosystem (acidification, eutrophication, and ecotoxicity), global warming potential, fossil fuel depletion, environmental issues (ozone depletion and smog formation), and human health impacts (including carcinogenic and noncarcinogenic health impacts, and respiratory effects). The 3D-molded wood shows substantial reductions (59 to 99%) across all environmental impact categories, even compared with Al alloy with a high recycled content of 73.9% (Fig. 4G). Such reductions are even larger (74 to 99% reduction) (fig. S21) compared with Al alloy with lower recycled content (35%). Additionally, on the basis of the specific tensile strength measured in this study, 3D-molded wood only needs 21 to 23% of the mass of Al alloy to fulfill the same strength requirements (equations S3 to S7). LCA results based on the material stiffness, another important mechanical property for a structural material, also show that compared with Al alloy, 3D-molded wood has lower environmental impacts in most categories, with a few exceptions (figs. S22 and S23). The contribution analysis of the LCA identifies the energy consumption of the wood treatment as the major contributor to the environmental impacts of the 3D-molded wood (table S6). However, by reducing the delignification time (table S7; see the supplementary materials for more details), the environmental impacts of the 3D-molded wood are drastically reduced and lower than Al alloy across all impact categories (figs. S24 to S28).

In this work, we demonstrate how cell wall engineering can render wood foldable and moldable while simultaneously improving its mechanical properties—endowing wood with a structural versatility previously limited to plastics and metals. The 3D-molded wood also has several advantages as a sustainable material, with a potentially lower environmental impact than Al alloy. 3D-molded wood may also enable substantial fuel savings and corresponding environmental benefits when used as a lightweight structural material for vehicles and aircraft. Additionally, wood products have the added benefit, compared to plastics and metals, of large carbon storage capabilities, which is recognized by the Intergovernmental Panel on Climate Change for mitigating global warming (*43*). Furthermore, advances in wood processability and functionality could motivate better forest management practices. We note that in addition to basswood, other hardwood species and woods of different tree ages and growth locations can be used to fabricate the 3D-molded wood by the same treatment, suggesting the universality of this approach to hardwoods and its potential for wide-scale application (figs. S29 to S31). Wood cell wall engineering can substantially expand the versatility of this sustainable and high-performance material, making wood a potential alternative to plastics and metals in structural applications.

REFERENCES AND NOTES

1. S. Xu *et al*., *Science* **347**, 154–159 (2015).
2. Y. Kim, H. Yuk, R. Zhao, S. A. Chester, X. Zhao, *Nature* **558**, 274–279 (2018).
3. C. Wang, C. Wang, Z. Huang, S. Xu, *Adv. Mater.* **30**, e1801368 (2018).
4. Y. Tang *et al*., *Adv. Mater.* **29**, 1604262 (2017).
5. A. K. Naskar, J. K. Keum, R. G. Boeman, *Nat. Nanotechnol.* **11**, 1026–1030 (2016).
6. A. I. Taub, *MRS Bull.* **31**, 336–343 (2006).
7. K. Zheng, D. J. Politis, L. Wang, J. Lin, *Int. J. Lightweight Mater. Manuf.* **1**, 55–80 (2018).
8. X. Liao *et al*., *Science* **366**, 1376–1379 (2019).
9. P. Fratzl, R. Weinkamer, *Prog. Mater. Sci.* **52**, 1263–1334 (2007).
10. Q.-F. Guan *et al*., *Sci. Adv.* **6**, eaaz1114 (2020).
11. *Nature* **557**, 280 (2018).
12. C. Chen *et al*., *Nat. Rev. Mater.* **5**, 642–666 (2020).
13. E. Toumpanaki, D. U. Shah, S. J. Eichhorn, *Adv. Mater.* **33**, e2001613 (2021).
14. A. Geng, H. Yang, J. Chen, Y. Hong, *For. Policy Econ.* **85**, 192–200 (2017).
15. G. Churkina *et al*., *Nat. Sustain.* **3**, 269–276 (2020).
16. Y. Pan *et al*., *Science* **333**, 988–993 (2011).
17. S. Fink, *Holzforschung* **46**, 403–408 (1992).
18. M. I. Shams, H. Yano, *J. Trop. For. Sci.* **21**, 175–180 (2009).
19. J. Gierer, *Wood Sci. Technol.* **19**, 289–312 (1985).
20. R. M. Seborg, M. A. Millett, A. J. Stamm, "Heat-stabilized compressed wood (Staypak)," Report 1580 (USDA Forest Service, Forest Products Laboratory, 1956).
21. H. Yano, A. Hirose, P. J. Collins, Y. Yazaki, *J. Mater. Sci. Lett.* **20**, 1125–1126 (2001).
22. J. Song *et al*., *Nature* **554**, 224–228 (2018).
23. H. Sivonen, S. L. Maunu, F. Sundholm, S. Jämsä, P. Viitaniemi, *Holzforschung* **56**, 648–654 (2002).
24. B. F. Tjeerdsma, M. Boonstra, A. Pizzi, P. Tekely, H. Militz, *Holz als Roh-und Werkstoff* **56**, 149–153 (1998).
25. S. Trey, S. Jafarzadeh, M. Johansson, *ACS Appl. Mater. Interfaces* **4**, 1760–1769 (2012).
26. V. Merk, M. Chanana, N. Gierlinger, A. M. Hirt, I. Burgert, *ACS Appl. Mater. Interfaces* **6**, 9760–9767 (2014).
27. C. Mai, H. Militz, *Wood Sci. Technol.* **37**, 339–348 (2004).
28. P. N. Hasluck, *Manual of Traditional Wood Carving* (Courier Corporation, 1977).
29. M. J. Golden, *A Laboratory Course in Wood-Turning* (Harper & Brothers, 1897).
30. B. Hahn, T.-E. Werner, P. Haller, *Constr. Build. Mater.* **229**, 116829 (2019).
31. T. Saito, R. Kuramae, J. Wohlert, L. A. Berglund, A. Isogai, *Biomacromolecules* **14**, 248–253 (2013).
32. A. Walther, J. V. I. Timonen, I. Díez, A. Laukkanen, O. Ikkala, *Adv. Mater.* **23**, 2924–2928 (2011).
33. X. Yang *et al*., *Adv. Mater.* **33**, 2002264 (2021).
34. I. A. Kinloch, J. Suhr, J. Lou, R. J. Young, P. M. Ajayan, *Science* **362**, 547–553 (2018).
35. C. Chen *et al*., *ACS Nano* **14**, 16723–16734 (2020).
36. Z. Fang *et al*., *Matter* **2**, 1000–1014 (2020).
37. A. Dupleix, L. Denaud, L. Bléron, R. Marchal, M. Hughes, *Eur. J. Wood Prod.* **71**, 163–171 (2013).
38. D. W. Green, J. E. Winandy, D. E. Kretschmann, in *Wood Handbook: Wood as an Engineering Material*, General Technical Report FPL–GTR–113 (USDA Forest Service, Forest Products Laboratory, 1999), pp. 4.1–4.45.
39. D. W. A. Rees, *Mechanics of Optimal Structural Design: Minimum Weight Structures* (John Wiley & Sons, 2009).
40. T. Bitzer, *Honeycomb Technology* (Springer Netherlands, 1997).
41. C. Moes, G. Hibbard, *Mater. Des.* **141**, 67–80 (2018).
42. M. Peresin *et al*., "Large-scale nanofibrillated cellulose film: An overview on its production, properties, and potential applications," *4th International Conference of Pulping, Papermaking and Biotechnology (ICPPB 2012)*, Nanjing, China, 7 to 9 November 2012.
43. B. Netz, O. R. Davidson, P. R. Bosch, R. Dave, L. A. Meyer, in *Climate Change 2007: Mitigation. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2007), pp. 1–23.

ACKNOWLEDGMENTS

We acknowledge Z. Li, S. He, and Z. Lin for their assistance in the car-lifting experiment. **Funding:** The information, data, or work presented herein was funded in part by the Advanced Research Projects Agency–Energy (ARPA-E), US Department of Energy, under award DE-AR0001025. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. L.H. also acknowledges support from the University of Maryland A. James Clark School of Engineering. We acknowledge the support of the Maryland Nanocenter, its Surface Analysis Center, and AIMLab. **Author contributions:** S.X., C.C., Q.X., and Y.L. contributed equally to this work. L.H. conceived of the concept and supervised the project. L.H., S.X., and C.C. designed the experiments. S.X., Q.X., and Y.L. carried out the wood shaping experiments. S.X., J.L., Yo. Yao, and W.G. measured the mechanical properties. Q.X. and Y.L. provided characterization via SEM. A.L and M.H. performed the compression and bending tests of the honeycomb structures. T.L. and Q.C. contributed to the mechanistic modeling of the foldability of the 3D-molded wood. K.T. and I.B. contributed to the Raman spectroscopy measurements and analysis. V.W.Y. performed the antifungal measurements. Yu. Yao conducted the life cycle assessment. L.H., S.X., C.C., and Q.X. analyzed the data of the manuscript. L.H., S.X., C.C., and A.B. wrote the manuscript. All authors commented on the final manuscript. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. **Competing interests:** The authors declare no competing interests.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abg9556
Materials and Methods
Supplementary Text
Figs. S1 to S31
Tables S1 to S7
References (*44*–*66*)
Movies S1 to S4

7 February 2021; accepted 14 September 2021
10.1126/science.abg9556

CORONAVIRUS

Defining variant-resistant epitopes targeted by SARS-CoV-2 antibodies: A global consortium study

Kathryn M. Hastie[1]†, Haoyang Li[1]†, Daniel Bedinger[2], Sharon L. Schendel[1], S. Moses Dennison[3], Kan Li[3], Vamseedhar Rayaprolu[1], Xiaoying Yu[1], Colin Mann[1], Michelle Zandonatti[1], Ruben Diaz Avalos[1], Dawid Zyla[1], Tierra Buck[1], Sean Hui[1], Kelly Shaffer[1], Chitra Hariharan[1], Jieyun Yin[1], Eduardo Olmedillas[1], Adrian Enriquez[1], Diptiben Parekh[1], Milite Abraha[3], Elizabeth Feeney[3], Gillian Q. Horn[3], CoVIC-DB team[1], Yoann Aldon[4], Hanif Ali[5], Sanja Aracic[6], Ronald R. Cobb[7], Ross S. Federman[8], Joseph M. Fernandez[9], Jacob Glanville[10], Robin Green[8], Gevorg Grigoryan[8], Ana G. Lujan Hernandez[11], David D. Ho[12], Kuan-Ying A. Huang[13], John Ingraham[8], Weidong Jiang[14], Paul Kellam[15,16], Cheolmin Kim[17], Minsoo Kim[17], Hyeong Mi Kim[17], Chao Kong[18], Shelly J. Krebs[19], Fei Lan[9,20], Guojun Lang[18], Sooyoung Lee[17], Cheuk Lun Leung[8], Junli Liu[14], Yanan Lu[9,21], Anna MacCamy[22], Andrew T. McGuire[22], Anne L. Palser[15], Terence H. Rabbitts[5,23], Zahra Rikhtegaran Tehrani[24], Mohammad M. Sajadi[24], Rogier W. Sanders[4], Aaron K. Sato[11], Liang Schweizer[25], Jimin Seo[17], Bingqing Shen[25], Jonne L. Snitselaar[4], Leonidas Stamatatos[22], Yongcong Tan[18], Milan T. Tomic[26], Marit J. van Gils[4], Sawsan Youssef[10], Jian Yu[12], Tom Z. Yuan[11], Qian Zhang[25], Bjoern Peters[1,27], Georgia D. Tomaras[3], Timothy Germann[2], Erica Ollmann Saphire[1,27]*

Antibody-based therapeutics and vaccines are essential to combat COVID-19 morbidity and mortality after severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection. Multiple mutations in SARS-CoV-2 that could impair antibody defenses propagated in human-to-human transmission and spillover or spillback events between humans and animals. To develop prevention and therapeutic strategies, we formed an international consortium to map the epitope landscape on the SARS-CoV-2 spike protein, defining and structurally illustrating seven receptor binding domain (RBD)–directed antibody communities with distinct footprints and competition profiles. Pseudovirion-based neutralization assays reveal spike mutations, individually and clustered together in variants, that affect antibody function among the communities. Key classes of RBD-targeted antibodies maintain neutralization activity against these emerging SARS-CoV-2 variants. These results provide a framework for selecting antibody treatment cocktails and understanding how viral variants might affect antibody therapeutic efficacy.

Cell entry of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is mediated by its surface glycoprotein, spike. The S1 subunit of spike contains the N-terminal domain (NTD) and the receptor binding domain (RBD), which mediates recognition of the host cell receptor angiotensin-converting enzyme 2 (ACE2). The S2 subunit drives fusion between virus and host cell membranes. Spike, particularly the S1 subunit, is the primary target of neutralizing antibodies against SARS-CoV-2 (*1*).

Since SARS-CoV-2 first emerged, recurrent mutations in spike arose during both human-to-human transmission (*2*–*4*) and spillover or spillback events between humans and animals (*5*–*8*). Distinct variants of concern (VOCs) or variants of interest (VOIs)—including those first identified in the UK (Alpha, B.1.1.7), South Africa (Beta, B.1.351), Brazil (Gamma, P.1), India (Delta, B.1.617.2), and California (Epsilon, B.1.429)—carry several mutations associated with enhancement of human-to-human transmission (*9*). In particular, the receptor binding motif (RBM) mutations K417, L452, E484, and N501 affect ACE2-spike interactions (*10*). Variations at positions N439 and S477 are frequently detected in patient samples (*3*, *11*, *12*), whereas others such as V367F, Y453F, and F486L are associated with cross-species transmission (*6*, *8*). The NTD is also highly mutable and is especially prone to deletions: ΔHV69-70 and ΔY144 are both seen in B.1.1.7 and ΔHV69-70 is in the mink-associated Cluster V (*6*). ΔLAL242-244 appears in B.1.351, and ΔFR157-158 is found in B.1.617.2 (*9*). The NTD point mutations S13I and W152C alter disulfide bonding and conformation of the B.1.429 NTD (*13*) (fig. S1).

SARS-CoV-2 will continue to evolve. By understanding antibody footprints and the distinct ways by which antibodies target spike, we may deduce optimal combinations of monoclonal antibodies (mAbs) to prevent and treat infection by emerging variants and to minimize the risk of viral escape. We can also gauge the susceptibility of mapped antibodies to new mutations and predict whether newly identified mAbs might also be susceptible to viral escape. Thus, we sought to define functionally important groups in an array of therapeutic candidates and to dissect how key mutations, both individually and combined as in VOCs, affect antibody-mediated neutralization in a pseudovirus neutralization assay.

The Coronavirus Immunotherapeutic Consortium (CoVIC) was formed to analyze candidate antibody therapeutics side by side in standardized assays (*14*) and now includes more than 350 mAbs directed against the SARS-CoV-2 spike protein from 56 different partners across four continents (*15*). The panel includes antibodies derived from COVID-19 survivors, phage display, naïve libraries, in silico methods, and other strategies—each elicited, evaluated, and selected using distinct criteria. The panel thus represents a broader and deeper array of antibodies from which both fundamental information and therapeutic cocktails can be derived. With the goals of FAIR (findable, accessible, interoperable, reusable) data analysis and management as well as inclusion of otherwise inaccessible clinical candidates, candidate antibody therapeutics were blinded and tested in multiple in vitro and in vivo assays with comparative data

[1]Center for Infectious Disease and Vaccine Research, La Jolla Institute for Immunology, 9420 Athena Circle, La Jolla, CA 92037, USA. [2]Carterra, 825 N. 300 W. Ste C309, Salt Lake City, UT 84103, USA. [3]Center for Human Systems Immunology, Departments of Surgery, Immunology, and Molecular Genetics and Microbiology and Duke Human Vaccine Institute, Duke University, Durham, NC 27701, USA. [4]Department of Medical Microbiology and Infection Prevention, Amsterdam University Medical Centers, Location AMC, University of Amsterdam, Amsterdam Infection & Immunity Institute, 1105 AZ Amsterdam, Netherlands. [5]Quadrucept Bio, Ltd., Cambridge CB23 6DW, UK. [6]Myrio Therapeutics Pty, Ltd., 1 Dalmore Drive, Scoresby, VIC 3179, Australia. [7]National Resilience, Inc., 13200 NW Nano Ct., Alachua, FL 32615, USA. [8]Generate Biomedicines, Inc., 26 Landsdowne Street, Cambridge, MA 02139, USA. [9]Activemotif, Inc., 1914 Palomar Oaks Way, Suite 150, Carlsbad, CA 92008, USA. [10]Centivax, Inc., 201 Gateway Blvd., Floor 1, South San Francisco, CA 94080, USA. [11]Twist Bioscience, 681 Gateway Blvd., South San Francisco, CA 94080, USA. [12]Aaron Diamond AIDS Research Center, Columbia University Vagelos College of Physicians and Surgeons, 701 West 168th St., HHSC 1102, New York, NY 10032, USA. [13]Division of Pediatric Infectious Diseases, Department of Pediatrics, Chang Gung Memorial Hospital and Research Center for Emerging Viral Infections, Chang Gung University, Taoyuan, Taiwan. [14]Shanghai Henlius Biotech, Inc., 9/F, Innov Tower, Zone A, no. 1801 Hongmei Road, Xuhui District, Shanghai, China. [15]Kymab, Ltd., The Bennet Building, Babraham Research Campus, Cambridge CB22 3AT, UK. [16]Department of Infectious Disease, Imperial College, London SW7 2AZ, UK. [17]Celltrion, Inc., Department of Research and Development, 23 Academy-ro Yeonsu-gu Incheon, Republic of Korea. [18]Sanyou Biopharmaceuticals Co., Ltd., no. 188 Xinjunhuan Road, Building 6B-C, 3rd Floor, Minhang District, Shanghai 201114, China. [19]Emerging Infectious Diseases Branch, Walter Reed Army Institute of Research, Silver Spring, MD 20910, USA. [20]Shanghai Key Laboratory of Medical Epigenetics, International Laboratory of Medical Epigenetics and Metabolism, Ministry of Science and Technology, Institutes of Biomedical Sciences, Fudan University, Shanghai, China. [21]AbCipher Biotechnology, 188 Xinjun Ring Road, Building 2, 4th Floor, Minhang District, Shanghai 201114, China. [22]Fred Hutchinson Cancer Research Center, Vaccines and Infectious Diseases Division, Seattle, WA, USA. [23]Institute of Cancer Research, Centre for Cancer Drug Discovery, London SM2 5NG, UK. [24]Division of Clinical Care and Research, Institute of Human Virology, University of Maryland, Baltimore, MD 21201, USA. [25]HiFiBiO, Inc., 237 Putnam Avenue, Cambridge, MA 02139, USA. [26]National Resilience, Inc., 2061 Challenger Dr., Alameda, CA 94501, USA. [27]Department of Medicine, University of California, San Diego, La Jolla, CA 92037, USA.

***Corresponding author. Email: erica@lji.org**

†These authors contributed equally to this work.

uploaded into a publicly accessible database (covic.lji.org).

We first measured the affinity of 269 CoVIC mAbs for D614-Hexapro spike ectodomain trimers and monomeric RBD and NTD and the ability of each of these mAbs to block ACE2-RBD binding (figs. S2 to S5, table S1, and covic.lji.org). The panel, formed by candidates for therapeutic use, includes NTD- or S2-directed antibodies but is dominated by those targeting the RBD. In contrast to previous studies that classified mAbs using germline or structural information (*10*, *16*), the 186 RBD-reactive mAbs of CoVIC analyzed in this study were instead distinguished by a competition profile created by high-throughput surface plasmon resonance (HT-SPR). RBD-directed antibodies can be sorted into seven core "communities" (Fig. 1, fig. S6A, and table S2) that are broadly defined by the competition profiles of each mAb relative to the others. Communities can be further divided into finer clusters and bins on the basis of their discrete competition with other clusters and/or their ability to compete with ACE2 (Fig. 1 and tables S1 and S2).

To understand the position of each community relative to the others, we next mapped the footprints by negative-stain electron microscopy (NS-EM) for 25 example RBD-reactive mAbs chosen to span the range of communities and key clusters (table S3). To have a relatively agnostic view of antibody interactions with spike, mAbs were not chosen on the basis of germline origin, CDR (complementarity-determining region) feature or length, neutralization potency, or particular antibody origin (e.g., human, mouse, in silico) or format (e.g., IgG, scFv-Fc, VHH-Fc, multivalent).

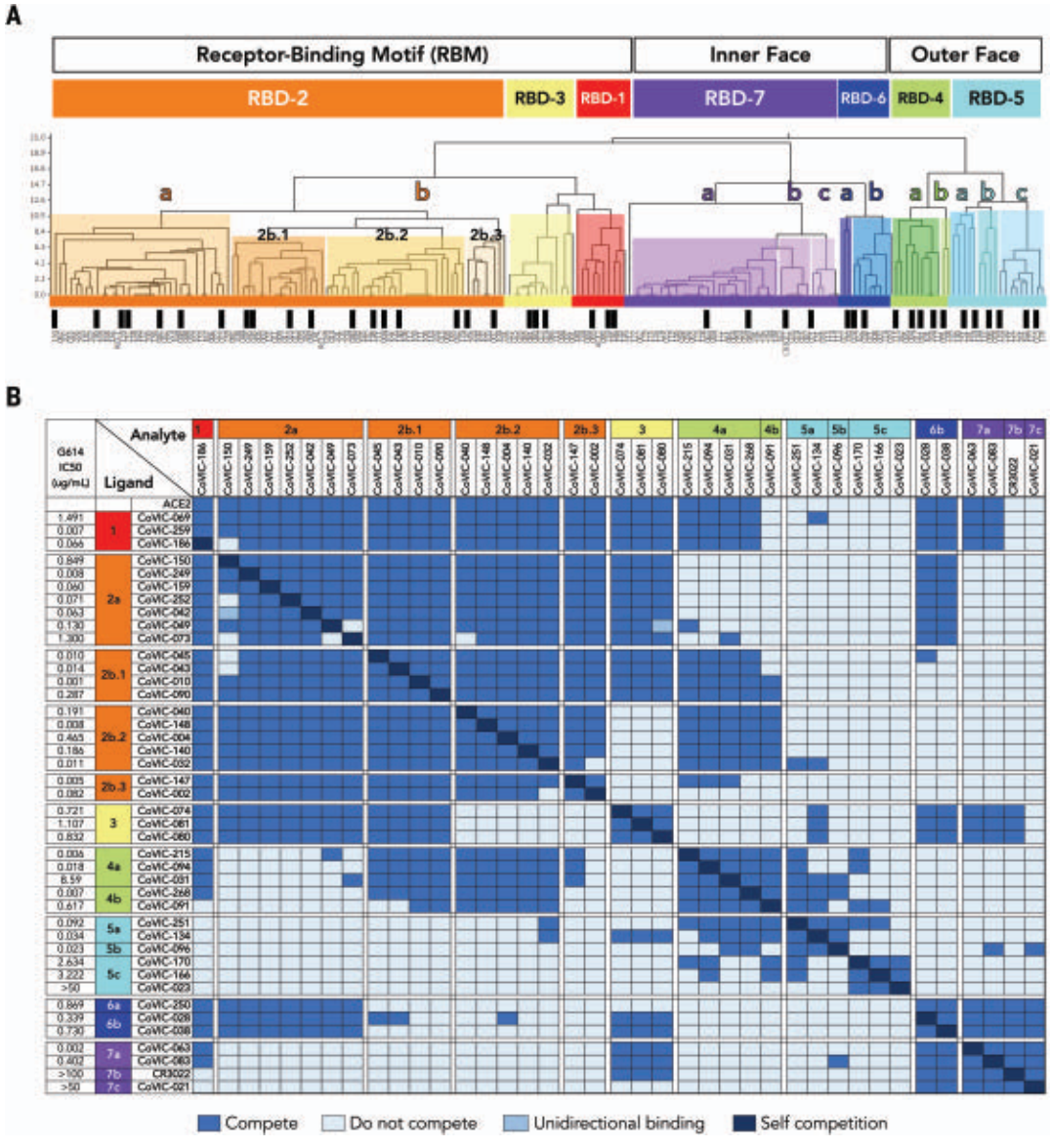


Fig. 1. The antigenic landscape of the SARS-CoV-2 RBD can be divided into seven binding communities. (**A**) HT-SPR was used to determine the competitive relationship between 186 RBD-directed mAbs. The dataset was analyzed by Carterra Epitope software to sort competition profiles of clones into related clusters, which are represented as regions of the dendrogram with shared color. The RBD epitope landscape can be broadly divided into seven communities containing mAbs that bind the RBM (RBD-1 through RBD-3), the outer face of the RBD (RBD-4 and RBD-5), or the inner face of the RBD (RBD-6 and RBD-7). Communities can be further divided into smaller clusters (e.g., RBD-2a and RBD-2b) and bins (e.g., RBD-2b.1, RBD-2b.2, and RBD-2b.3) on the basis of their discrete competition with other clusters and/or their ability to compete with ACE2 for spike binding. Black bars indicate single clones that were used in further analyses. Table S1 lists additional metrics (i.e., ACE2 blocking, kinetic analyses, and germline information) for the indicated mAbs; detailed information for the entire CoVIC panel can be found at covic.lji.org. (**B**) Binary heatmap matrix demonstrating the competition profile for the finer clusters and bins for the subset of single clones indicated by black bars in (A). The matrix here contains representative examples of each epitope community. RBD-2 can be divided into clusters a and b, which have varying ability to compete with mAbs in RBD-4 (e.g., RBD-2a mAbs do not compete, whereas most RBD-2b mAbs do). Cluster RBD-2b can be divided into three smaller bins that vary in their competition with both RBD-3 and RBD-4 mAbs: Those in 2b.1, but not 2b.2 or 2b.3, compete with RBD-3 mAbs, whereas mAbs in 2b.1 and 2b.2, but not 2b.3, compete with RBD-4 mAbs. RBD-4 contains mAbs that do (RBD-4a) and do not (RBD-4b) compete with ACE2. RBD-5 and RBD-7 have clusters of mAbs with lower neutralizing potency (i.e., RBD-5c, RBD-7b, and RBD-7c) relative to the other cluster in the same community (i.e., RBD-5a, RBD-5b, and RBD-7a). Rows and columns indicate the immobilized mAb and injected analyte mAb, respectively. Table S2 shows the complete matrix for competition among all 186 mAbs.

In parallel, we measured the neutralization activity of 41 RBD-directed mAbs (chosen to span the range of communities and key clusters) as well as a human ACE2-Fc fusion therapeutic candidate (CoVIC-069). Neutralization was measured against pseudoviruses that display the spike protein bearing (i) the globally dominant G614 variation, (ii) 15 single point mutations or deletions represented in circulating strains, (iii) constellations of mutations found in four VOCs [B.1.1.1 (Alpha), B.1.351 (Beta), P.1 (Gamma), and B.1.617.2 (Delta)] and one VOI [B.1.429 (Epsilon)], and (iv) two pseudovariants carrying four mutations (termed 4xM; containing G261D, Y453F, F486L, and N501T) or five mutations (termed 5xM; carrying the 4xM mutations plus V367F) identified in human–mink spillover events (fig. S1).

The mAbs in RBD-1 through RBD-3 target the RBM, compete with ACE2, and generally require the RBD to be in the "up" conformation for binding (footprints defined in Fig. 2B, table S3, and covic.lji.org). Community RBD-1 contains hACE2-derived molecules and IgGs (e.g., CoVIC-259, EMD-24335) that largely overlap with the RBM (Fig. 2B, fig. S6B, and table S3). The footprint for RBD-2 mAbs is shifted from the center of the ACE2 binding site toward the peak of the RBM (Fig. 2B, fig. S6B, and table S3). RBD-2 is the largest community and can be divided further into clusters and then bins on the basis of competition with other communities (Fig. 1). Cluster 2a antibodies (e.g., CoVIC-252, EMD-24339) bind toward the inner face of the RBD and its binding area overlaps highly with that of the therapeutic antibody REGN-10933 (*17*). Antibodies in 2b.1 [e.g., CoVIC-010, EMD-24343; similar to antibody COVA2-39 (*18*)] and 2b.2 [e.g., CoVIC-140, EMD-24383; similar to antibody C144 (*16*)] bind toward the outer face of the RBD, and mAbs in bin 2b.3 [e.g., CoVIC-002, EMD-24345; similar to antibody S2E12 (*19*)] bind to the peak of the RBD

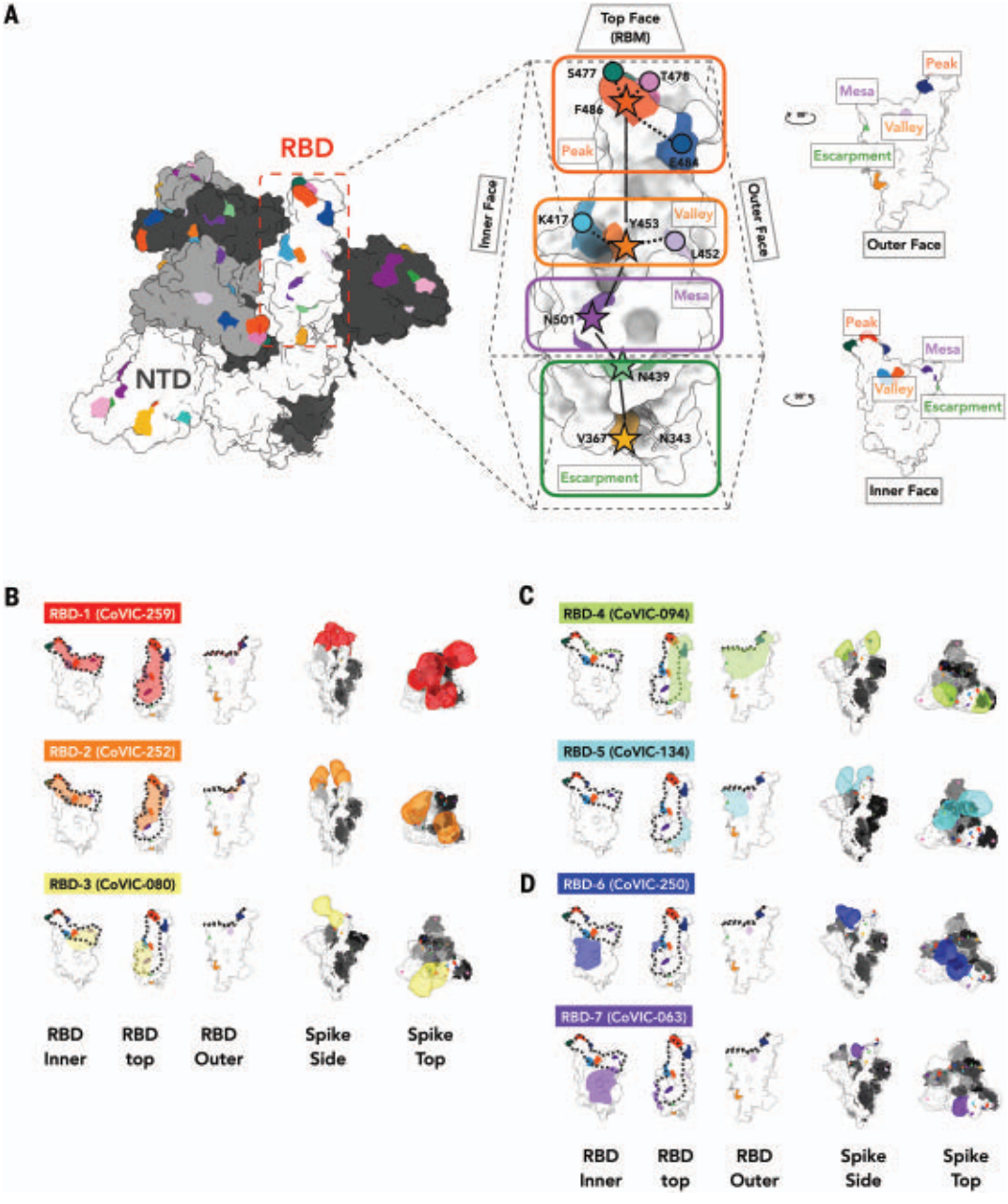


Fig. 2. NS-EM analysis of representatives from each RBD-directed community. (**A**) Location of important emerging mutations in a RBD. The spike trimer [adapted from PDB ID 7A94 (*39*)] viewed from the top with one "up" RBD is shown; individual spike monomers are colored white, gray, and black. The RBM can be topologically divided into three subsections: the "peak" that includes residues F486, S477, T478, and E484; the "valley" including residues Y453, K417, and L452; and the "mesa" with residue N501. Stars indicate residues on the central axis of the RBD. The "outer face" [exposed in the RBD down (closed) conformation] and "inner face" [buried inside the trimer in the RBD down (closed) conformation] define the lateral faces of the RBD and the "escarpment" (contains residues V367 and N439 and glycan 343). (**B** to **D**) NS-EM footprint of a representative antibody from each community mapped onto an RBD monomer. The colored shading corresponds to the community colors in Fig. 1. The ACE2 binding site is outlined with a dotted line. Side and top views of spike trimers show the Fab approach angle and binding stoichiometry for each representative. Table S3 shows NS-EM data for all of the 29 RBD-directed mAbs that we analyzed.

(Fig. 2B, fig. S7, and table S3). Lastly, RBD-3 mAbs bind down from the center of the ACE2 binding site toward the RBD "mesa" [Fig. 2B and table S3; e.g., CoVIC-080, EMD-24346; similar to antibody ADI-56046 (*20*)].

To simulate the authentic interactions between antibodies and spike, intact IgGs were used for NS-EM structural analysis whenever possible. RBD-1 IgGs tend to fully occupy all three RBDs on one spike and often cross-link two spike trimers, whereas most RBD-2 IgGs tend to bind bivalently to a single spike trimer (figs. S8, A and B, and S9 and table S3). RBD-3 IgGs can cross-link spikes, and bivalent binding was also observed in some cases (table S3 and fig. S9).

General epitope position, particularly that of RBM epitopes, is strongly associated with the propensity of particular spike mutations to escape antibody-mediated neutralization (Fig. 3, fig. S10, and table S4). Neutralization by RBD-2a antibodies is heavily affected by the K417N mutation but rarely by the E484K mutation; those in RBD-2b are affected by the E484K mutation but less so by K417N. Similarly, RBD-2a antibodies are resistant to the L452R mutation found in B.1.429 (Epsilon) and B.1.617.2 (Delta), whereas only some RBD-2b antibodies are sensitive to this mutation. Meanwhile, mAbs in RBD-3 are affected by both N501T/Y and E484K mutations (Fig. 3, figs. S10 and S11, and table S4). In contrast to RBD-2 and RBD-3, the susceptibility of neutralization activity of antibodies in RBD-1 to particular mutations is more variable (Figs. 2B and 3 and tables S3 and S4).

Regardless of the effect of particular single point mutations, nearly every RBD-1 or RBD-2 mAb analyzed showed additive decreases in potency against pseudovirus carrying constellations of multiple mutations in the RBM (Fig. 3, fig. S10, and table S4). For B.1.351 and P.1, almost all of the RBD-1 and RBD-2 antibodies that we analyzed suffer a complete loss of neutralization activity. For example, CoVIC-249 and CoVIC-010 show moderate or no change in IC_{50} (half-maximal inhibitory concentration) against the single point mutations K417N, E484K, and N501Y, but CoVIC-249 loses all neutralization activity and CoVIC-010 potency falls by 1000-fold against B.1.351 (Beta) and P.1 (Gamma), which contain all three mutations. Many RBD-2 antibodies also lose activity against the 4xM mink pseudovariant that carries Y453F, F486L, and N501T mutations (Fig. 3, figs. S1C and S10, and table S4).

By contrast, most RBD-1 and RBD-2 antibodies retain neutralization activity against B.1.1.7 (Alpha), B.1.429 (Epsilon), and B.1.617.2 (Delta) variants, which each contain only one or two RBM-located mutations (N501Y, L452R, or T478K/L452R respectively). Curiously, the V367F mutation identified in mink populations enhances neutralization by some RBD-2 mAbs, and in some cases, this mutation can offset decreases in potency resulting from other single point mutations. For example, CoVIC-040 has 14- and 8-fold decreases in potency against the F486L mutation and the F486L-containing 4xM mink pseudovariant, respectively, but only a 4-fold decrease against the 5xM mink pseudovariant, which contains V367F in addition to the four mutations present in 4xM (Fig. 3, fig. S10, and table S4). V367 is adjacent to an N-linked glycan at position 343, which was recently implicated in providing a gating mechanism for the RBD (*21*). Substitution of valine with phenylalanine could alter the local environment of the N343 glycan moieties and enable the RBD to adopt a conformation more amenable to antibody interaction.

Antibodies in communities RBD-4 and RBD-5 bind to the outer face of the RBD and, like the previously defined class 2 and class 3 mAbs (*16*), can do so in either the "up" or "down" RBD conformation without steric hindrance (Fig. 2C, figs. S6 and S12, and table S3). The footprints of these groups largely overlap, but RBD-4 mAbs bind toward the outer edge of the RBM and can block ACE2 [e.g., CoVIC-094, EMD-24350; similar to antibody C002 (*16*)], whereas RBD-5 mAbs bind away from the RBM, toward the S309 site, and block ACE2 weakly [e.g., CoVIC-134, EMD-24384; similar to antibody REGN-10987 (*17*)] (Figs. 1B and 2C, figs. S5 and S6B, and tables S1 to S3) (*10*). Some RBD-4 and RBD-5 IgGs can cross-link spike trimers in solution (fig. S8C and table S3).

Notably, in accordance with the five RBD-5 IgGs we imaged, only those IgGs that show spike-cross linking tendency have potent

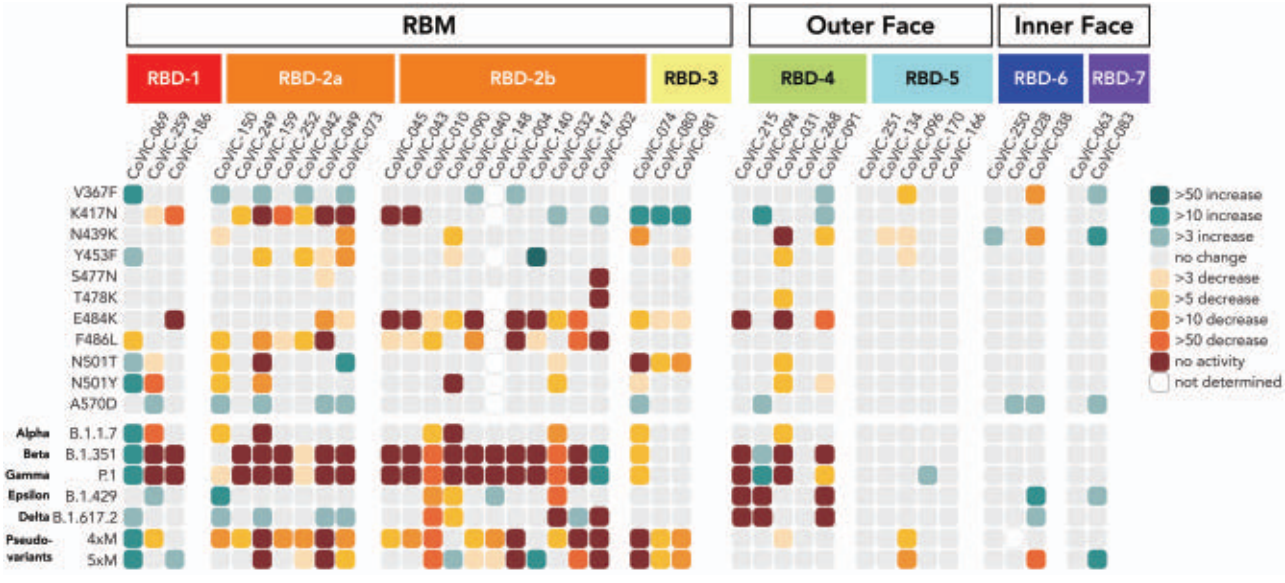


Fig. 3. RBD-5, -6, and -7 antibodies retain neutralization activity against pseudovirus bearing mutations singly or together in VOCs. Fold-change differences in potency for 38 RBD-directed antibodies and an ACE2-Fc fusion (CoVIC-069) are shown in a heatmap. In addition to VOCs, we also examined two pseudoviruses bearing clusters of mink-associated mutations: 4xM (G261D, Y453F, F486L, and N501T) and 5xM (G261D, Y453F, F486L, N501T, and V367F). Fig. S1 lists mutations represented in each variant. Fig. S10 shows neutralization curves for each virus–variant pair, and table S4 lists fold-change values corresponding to the heatmap. Single-letter abbreviations for amino acid residues: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

neutralizing activity (Fig. 1B, fig. S13, and table S3). A recent cryo–electron tomography study showed that native spike trimers on the SARS-CoV-2 virion surface tilt at variable degrees relative to the viral envelope (*22*). This finding provides a possibility for IgG-mediated spike cross-linking on virions and may contribute to the mechanism of neutralization of the RBD-5 mAbs in the absence of ACE2 blocking (fig. S8D).

Most RBD-4 mAbs are affected by E484K and/or L452R mutations (represented in the B.1.429 variant) (Fig. 3, fig. S10, and table S4), and some are affected by the N439K mutation, which is highly represented in sequences worldwide (*3*). RBD-5 mAbs, however, show broad resistance to nearly all mutations analyzed, with only two mAbs in this group showing moderate decreases in potency against V367F and N439K (Fig. 3, fig. S10, and table S4).

RBD-6 (e.g., CoVIC-250, EMD-24352) and RBD-7 (e.g., CoVIC-063, EMD-24353) antibodies bind to the inner face of the RBD and access a previously described cryptic epitope (*23*, *24*) (Fig. 2D, fig. S6B, and table S3). Similar to the binding of class 4 antibodies (*16*), binding of spike by RBD-6 and RBD-7 antibodies requires two RBDs to be in the up configuration (fig. S12). The representative IgGs in RBD-6 and RBD-7 each show stronger propensities to cross-link spike trimers than do RBM-directed antibodies (fig. S9 and table S3). RBD-6 and RBD-7 antibodies primarily vary in their competition with RBD-2a antibodies: The downward shift of the RBD-7 footprint on the inner face of the RBD relative to the RBD-6 footprint would allow simultaneous binding of RBD-2a antibodies with RBD-7 antibodies but not RBD-6 antibodies (Figs. 1B and 2D, fig. S6B, and table S2). This cryptic site targeted by RBD-6 and RBD-7 antibodies is also recognized by antibodies COVA1-16 (*23*) and CR3022 (*24*). Here, strategies of site recognition are further subdivided by competition subgroups—information that is useful for interpreting differences and antibody behavior and strategies for cocktail selection.

All RBD-6 and RBD-7a antibodies block ACE2, but antibodies in RBD-7b and RBD-7c do not (Fig. 1B and tables S1 to S3). The representatives from the RBD-7b and RBD-7c clusters (CR3022 and CoVIC-021, respectively) demonstrate poor neutralization of pseudoviruses in our assay. The characteristic difference in neutralization behavior between 7a and 7b or 7c suggests that at this cryptic epitope, competition with ACE2 is a determinant of neutralization (Fig. 1B and table S4) (*25*). Notably, owing to their location away from the RBM, RBD-6 and RBD-7 antibodies are resistant

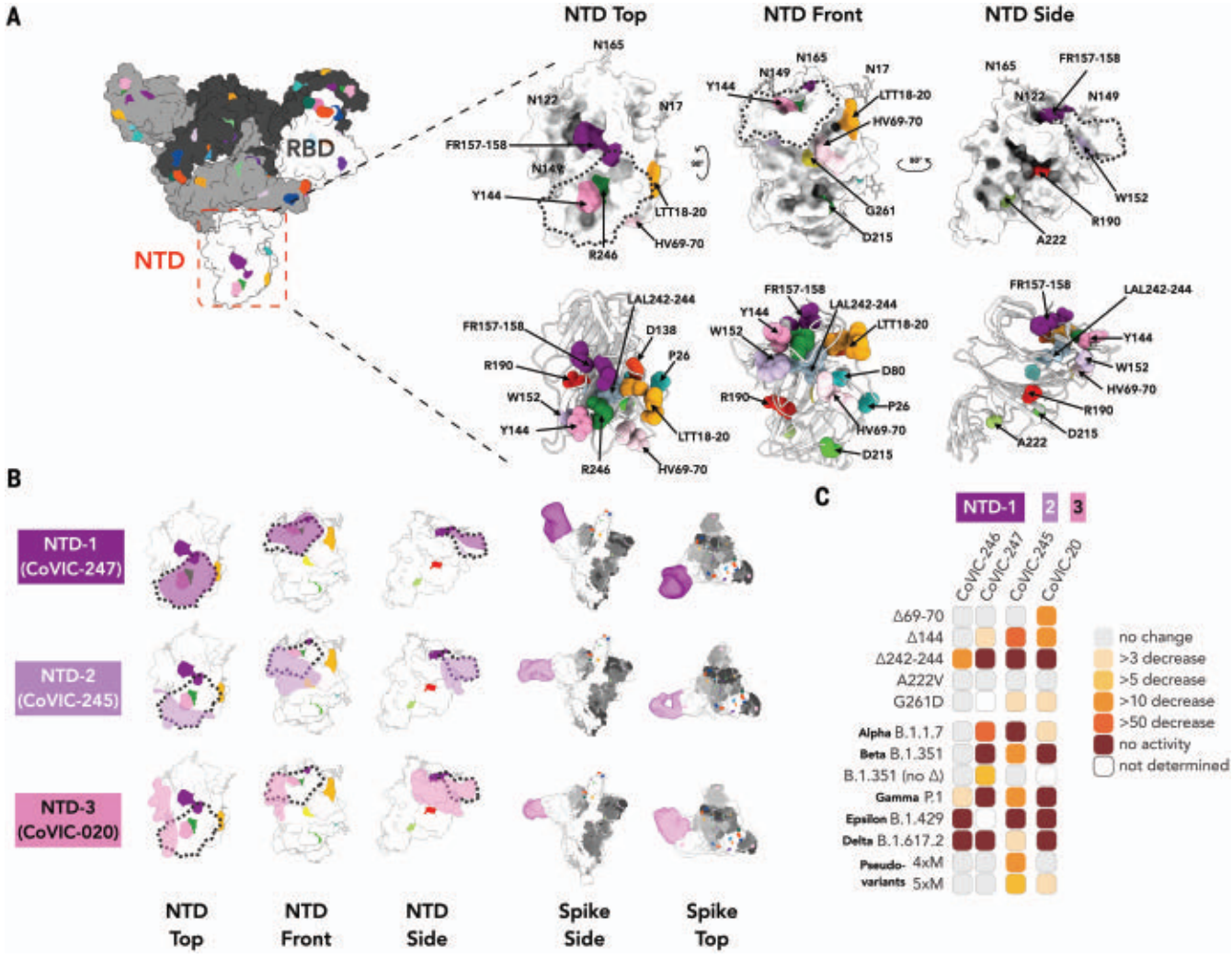


Fig. 4. NS-EM and neutralization analysis of mAbs targeting the NTD. (**A**) Surface and cartoon [adapted from PDB ID 7A94 (*39*)] representation of the spike NTD. The residue positions of mutations and deletions in circulating VOCs are indicated in three views of the NTD. Fig. S1 lists mutations represented in each variant. (**B**) Footprints for three NTD-targeted antibodies, with the NTD "supersite" (*26*) outlined with a dashed line. The NTD-directed antibodies shown here define the approximate boundaries of the neutralizing epitope landscape. Additional NS-EM data are in table S3. (**C**) Fold-change in potency of pseudovirus neutralization experiments for each antibody–variant pair.

to the mutations and variants analyzed (Fig. 3, fig. S10, and table S4).

Previous reports identified a "supersite" as the primary target for neutralizing antibodies directed against the NTD (*26*). In addition to RBD-directed antibodies, we also analyzed four CoVIC NTD-directed antibodies by NS-EM and in neutralization assays. Together, these four antibodies (grouped as NTD-1 through NTD-3), encompass the approximate boundaries of the supersite. The two NTD-1 antibodies bind from the top side of the NTD to cover the NTD N terminus and residue Y144 [Fig. 4; e.g., CoVIC-247, EMD-24355 (table S3)]. The NTD-1 epitope overlaps with that of mAb 4A8 (*27*) and other supersite binders (*28*, *29*). The NTD-2 antibody (CoVIC-245, EMD-24360) approaches from the front side of the NTD and contacts Y144 as well as residues H69, V70, W152, and G261, all of which are deleted or substituted in emerging variants (Fig. 4 and fig. S1). The NTD-2 footprint is similar to the footprint of antibodies in the antigenic site V group (*26*). The NTD-3 mAb (CoVIC-020, EMD-24356) binds to the left side of the NTD, proximal to the RBD of the adjacent monomer and in contact with residue W152 (Fig. 4B). The NTD-3 mAb represents a novel epitope and binding location of an anti-NTD antibody.

Unlike the RBD-directed antibodies, for which neutralization escape is strongly associated with antibody footprint, the NTD-directed antibodies are conformationally sensitive and affected by mutations outside of the discrete footprint. This finding is consistent with that for antibodies elicited by vaccines (*30*). Each of the four NTD mAbs analyzed exhibits a decreased or total loss of neutralization capacity for one or more of the NTD-located deletions (Δ69/70, ΔY144, Δ157-158, and Δ242-244) found in circulating VOCs, regardless of their binding location on the NTD (Fig. 4C, fig. S10, and table S4). All NTD mAbs were affected by P.1 (Gamma), which lacks deletions and instead has several point mutations in the NTD. For B.1.429 (Epsilon), altered disulfide bonding in the NTD arising from the S13I and W152C mutations (*13*) also abrogated mAb-mediated neutralization. Our results indicate that NTD mutations decrease not only neutralization potency but also the total fraction of virus neutralized (fig. S10).

Several therapeutic antibody cocktails that include pairs of different mAbs against spike are currently under investigation for postexposure treatment of COVID-19 (*16*, *17*, *31*, *32*). However, the potency of some antibodies in these cocktails is compromised by emerging SARS-CoV-2 variants (*33*, *34*). Meanwhile, exposure of virus to monoclonal or polyclonal antibodies can promote antibody-resistant mutations in spike (*34*–*37*). Notably, SARS-CoV-2 variants that share critical mutations with B.1.1.7 (Alpha) were isolated from an immunocompromised COVID-19 patient who received three rounds of convalescent plasma treatment, indicating that even a polyclonal therapeutic can drive evolution of resistant virus strains in unresolved infections (*38*).

Potency, variant resistance, and the ability to cobind are important considerations when selecting antibodies for therapeutic cocktails. The analysis of the 186 RBD-directed mAbs presented here—each donated by different groups around the world and selected in different ways—describes discrete antibody communities and functionally relevant subclusters and/or bins. This analysis provides a competition grid and a framework for cocktail selection. Notably, combining these data with neutralization potency and mutational analysis can guide selection of broadly protective therapeutic cocktails.

Overall, antibodies from community RBD-1 through RBD-4 and those directed against the NTD are generally more potent than antibodies of other communities. The high potency and nonoverlapping epitopes of RBD- and NTD-directed antibodies make them attractive as pairs for therapeutic cocktails. However, members of each of these groups are also highly susceptible to neutralization escape by mutations and deletions found in emerging VOCs. Indeed, a CoVIC bispecific antibody targeting the RBD-1 and NTD-1 sites could still neutralize single point mutations in the RBD (where the NTD arm could compensate) but was ineffective against B.1.351 (Beta) and P.1 (Gamma), which contain mutations that simultaneously escape both arms of the bispecific (fig. S14).

By contrast, RBD-5, RBD-6, and RBD-7 antibodies often have lower potency but are more resistant to escape. Notably, the epitopes targeted by RBD-5, RBD-6, and RBD-7 antibodies have high sequence conservation among the *Sarbecovirus* subgenus of *Betacoronavirus* (fig. S15). Enhanced potency for these communities might be achieved through engineering them as multivalent formats, making them key members of a variant-resistant cocktail that may also be suitable for treating other *Sarbecovirus* infections.

Taken together, the analysis presented here, made possible by broad participation of a few hundred therapeutic candidates in a global study, offers a detailed structural and competitive landscape of key antibody binding sites on spike. The results of this effort can be used to predict and interpret effects of VOCs and for strategic selection of durable therapeutics and cocktails against emerging variants.

REFERENCES AND NOTES

1. L. Piccoli *et al.*, *Cell* **183**, 1024–1042.e21 (2020).
2. B. Korber *et al.*, *Cell* **182**, 812–827.e19 (2020).
3. E. C. Thomson *et al.*, *Cell* **184**, 1171–1187.e20 (2021).
4. CDC, Science Brief: Emerging SARS-CoV-2 Variants (2021); www.cdc.gov/coronavirus/2019-ncov/more/science-and-research/scientific-brief-emerging-variants.html.
5. N. Oreshkova *et al.*, *Euro Surveill.* **25**, (2020).
6. WHO, COVID-19 - Denmark (2020); www.who.int/emergencies/disease-outbreak-news/item/2020-DON301.
7. B. B. Oude Munnink *et al.*, *Science* **371**, 172–177 (2021).
8. L. van Dorp *et al.*, bioRxiv 2020.11.16.384743 [Preprint] (2020); https://doi.org/10.1101/2020.11.16.384743.
9. CDC, SARS-CoV-2 Variant Classifications and Definitions (2021); www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html.
10. M. Yuan *et al.*, *Science* **373**, 818–823 (2021).
11. J. Chen, R. Wang, M. Wang, G.-W. Wei, *J. Mol. Biol.* **432**, 5212–5226 (2020).
12. E. B. Hodcroft *et al.*, *Nature* **595**, 707–712 (2020).
13. M. McCallum *et al.*, *Science* **373**, 648–654 (2021).
14. F. S. Collins *et al.*, "Therapeutic Neutralizing Monoclonal Antibodies: Report of a Summit sponsored by Operation Warp Speed and the National Institutes of Health" (2020); www.nih.gov/sites/default/files/research-training/initiatives/activ/20200909-mAb-summit-pub.pdf.
15. Coronavirus Immunotherapy Consortium (2020); https://covic.lji.org/.
16. C. O. Barnes *et al.*, *Nature* **588**, 682–687 (2020).
17. J. Hansen *et al.*, *Science* **369**, 1010–1014 (2020).
18. N. C. Wu *et al.*, *Cell Rep.* **33**, 108274 (2020).
19. M. A. Tortorici *et al.*, *Science* **370**, 950–957 (2020).
20. A. Z. Wec *et al.*, *Science* **369**, 731–736 (2020).
21. T. Sztain *et al.*, *Nat. Chem.* **13**, 963–968 (2021).
22. H. Yao *et al.*, *Cell* **183**, 730–738.e13 (2020).
23. H. Liu *et al.*, *Immunity* **53**, 1272–1280.e5 (2020).
24. M. Yuan *et al.*, *Science* **369**, 1119–1123 (2020).
25. J. Huo *et al.*, *Cell Host Microbe* **28**, 445–454.e6 (2020).
26. M. McCallum *et al.*, *Cell* **184**, 2332–2347.e16 (2021).
27. X. Chi *et al.*, *Science* **369**, 650–655 (2020).
28. G. Cerutti *et al.*, *Cell Host Microbe* **29**, 819–833.e7 (2021).
29. M. McCallum *et al.*, bioRxiv 2021.01.14.426475 [Preprint] (2021); https://doi.org/10.1101/2021.01.14.426475.
30. Y. Cao *et al.*, *Cell Res.* **31**, 732–741 (2021).
31. A. Baum *et al.*, *Science* **370**, 1110–1115 (2020).
32. NIH, Clinical trials of monoclonal antibodies to prevent COVID-19 now enrolling (2020); www.nih.gov/news-events/news-releases/clinical-trials-monoclonal-antibodies-prevent-covid-19-now-enrolling.
33. Z. Wang *et al.*, *Nature* **592**, 616–622 (2021).
34. P. Wang *et al.*, *Nature* **593**, 130–135 (2021).
35. Y. Weisblum *et al.*, *eLife* **9**, e61312 (2020).
36. Z. Liu *et al.*, *Cell Host Microbe* **29**, 477–488.e4 (2021).
37. A. J. Greaney *et al.*, *Cell Host Microbe* **29**, 44–57.e9 (2021).
38. S. A. Kemp *et al.*, *Nature* **592**, 277–282 (2021).
39. D. J. Benton *et al.*, *Nature* **588**, 327–330 (2020).

ACKNOWLEDGMENTS

We are grateful for the multiple generous contributions of antibodies to the CoVIC study, with special thanks to M. Beasley, S. Daraeikia, V. Dussupt, B. Kiefel, S. Liao, C. Liu, L. Mendez-Rivera, P. Rijal, L. Schimanski, P. Smith, T. Tan, A. Townsend, J. Wang, R. Yan, and L. Yang. We thank the electron microscope facility of La Jolla Institute for Immunology for the EM data collection.
Funding: We gratefully acknowledge philanthropic support of the Overton family for this urgent study (Coronavirus Immunotherapeutic Consortium), COVID-19 Therapeutics Accelerator INV-006133 (Coronavirus Immunotherapeutic Consortium), Bill and Melinda Gates Foundation OPP1210938 (Coronavirus Immunotherapeutic Consortium), GHR Foundation (Coronavirus Immunotherapeutic Consortium), and NIH/NIAID grant U19 AI142790-S1 (Coronavirus Immunotherapeutic Consortium). We also acknowledge philanthropic support of C. Lee, FastGrants from Emergent Ventures at the Mercatus Center, George Mason University for support of essential instrumentation, and T.S. for support of the Saphire laboratory efforts during the pandemic (E.O.S.); Early Postdoc Mobility Fellowship of the Swiss

National Science Foundation P2EZP3_195680 (D.Z.); and Translating Duke Health Immunology Initiative (G.T.). **Author contributions:** Conceptualization: K.H., H.L., S.S., and E.O.S. Methodology: K.H., S.S., H.L., D.B., S.D., K.L., G.T., and E.O.S. Software: J.I., G.G., and K.L. Validation: K.H., H.L., S.M.D., K.L., D.B., G.T., and E.O.S. Formal analysis: K.H., H.L., and D.B. Investigation: K.H., H.L., S.S., D.B., D.Z., V.R., M.Z., R.D.A., C.M., T.B., X.Y., S.H., K.S., C.H., J.Y., E.O., A.E., D.B., S.M.D., D.A., S.D., K.L., M.A., G.H., and L.F. Resources: H.A., R.F., J.J.F., J.G., R.G., G.G., J.I., A.L.H., W.J., C.K., M.K., H.M.K., C.K., F.L., G.L., S.L., A.L., J.L., Y.A., A.L.P., P.K., R.C., M.T., A.M., A.T.M., T.R., Z.R.-T., M.S., A.S., L.S., M.J.v.G., T.Y., J.S., B.S., L.S., Y.T., and Q.Z. Data curation: CoVIC-DB team, B.P., and S.S. Writing – original draft: K.H., H.L., and S.S. Writing – review & editing: EOS. Visualization: K.H., H.L., D.B., D.Z., K.L., and M.S. Supervision: B.P., G.T., T.G., and E.O.S. Project administration: S.S. and E.O.S. Funding acquisition: G.T., T.G., and E.O.S. The CoVIC-DB team includes Brendan Ha, Mari Kojima, Mahita Jarjapu, Randi Vita, Anaïs Gambiez, Jason A. Greenbaum, and James A. Overton under the direction of B.P. We thank members of the Antibody Dynamics platform of the Global Health-Vaccine Accelerator Platforms: S. Mudrak, V. Bekker, and K. Makar for program management; D. Beaumont and M. Sampson for data management; and N. Eisel and L. Williams for technical expertise. We are grateful to all of the discovery teams who were involved in isolating and developing the antibodies contributed to the CoVIC panel. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** EM maps have been uploaded to the EMDataResource. The Electron Microscopy Data Bank (EMDB) accession numbers are available in the main text and table S3. Information concerning particular antibodies can be requested through the Coronavirus Immunotherapeutics Consortium at https://covic.lji.org. 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.abh2315
Materials and Methods
Figs. S1 to S15
Tables S1 to S4
References (*40*–*56*)
MDAR Reproducibility Checklist

24 February 2021; accepted 21 September 2021
Published online 23 September 2021
10.1126/science.abh2315

METALLURGY

In situ design of advanced titanium alloy with concentration modulations by additive manufacturing

Tianlong Zhang[1,2], Zhenghua Huang[3], Tao Yang[1]*, Haojie Kong[1], Junhua Luan[1], Anding Wang[1], Dong Wang[2], Way Kuo[1], Yunzhi Wang[4]*, Chain-Tsuan Liu[1]*

Additive manufacturing is a revolutionary technology that offers a different pathway for material processing and design. However, innovations in either new materials or new processing technologies can seldom be successful without a synergistic combination. We demonstrate an in situ design approach to make alloys spatially modulated in concentration by using laser-powder bed fusion. We show that the partial homogenization of two dissimilar alloy melts—Ti-6Al-4V and a small amount of 316L stainless steel—allows us to produce micrometer-scale concentration modulations of the elements that are contained in 316L in the Ti-6Al-4V matrix. The corresponding phase stability modulation creates a fine scale–modulated β + α′ dual-phase microstructure that exhibits a progressive transformation-induced plasticity effect, which leads to a high tensile strength of ~1.3 gigapascals with a uniform elongation of ~9% and an excellent work-hardening capacity of >300 megapascals. This approach creates a pathway for concentration-modulated heterogeneous alloy design for structural and functional applications.

Additive manufacturing (AM), also referred to as three-dimensional (3D) printing, integrates multiple metallurgical processes into one in which the making, shaping, and treating of alloys take place all at once in a single process (*1*–*8*). However, AM has largely been regarded as a forming technology that produces material components near net shape, without taking full advantage of the capabilities offered by AM for simultaneous and synergistic advancements of both alloy and process together (*2*, *9*–*12*). By building a component flexibly point by point and layer by layer, AM provides the opportunity to create heterogeneous alloys with location-specific compositions and microstructures (*8*, *13*–*15*). During the laser-powder bed fusion (L-PBF), because of the short lifetime and fast cooling of the melt pool, even finer-scale (such as micrometer-scale) concentration modulations (microCMs hereafter) can be achieved through partial homogenization within each melt pool by using a premixed mixture of different alloy or elemental powders (*13*, *14*). Such microCMs have been shown to have the potential to offer various alloy-specific benefits, such as a gradient precipitate microstructure in β-Ti (*16*); martensite-austenite heterostructures in steels (*17*, *18*); and controlled strain release, linear superelasticity, and ultralow elastic modulus in ferroelastic materials (*19*, *20*).

We have demonstrated how to design such a microCM Ti alloy through L-PBF by using a mixture of two commercial alloy powders: Ti-6Al-4V (Ti64) and 316L (67.5Fe-18Cr-12Ni-2.5Mo, both in wt %). The choice of these two alloys is based on the following considerations. Ti64 by means of AM has long suffered from the problem of both large detrimental columnar grains and poor work-hardening capacity (*21*–*24*). The elements in 316L (Fe, Cr, Ni, and Mo; we refer to these as 316L elements) are effective grain refiners (*25*, *26*) and also effective β stabilizers in Ti alloys. In addition, high-quality powders of both alloys are commercially available for L-PBF (*27*). With an appropriate choice of the amount of 316L to be added and the L-PBF processing parameters, microCMs of these elements in the Ti64 matrix are created, and the corresponding phase stability modulation that is associated with the concentration modulation results in a fine-scale modulated β + α′ dual-phase microstructure in the as-printed alloy. This type of microCM Ti alloy exhibits a high yield strength and a progressive transformation-induced plasticity (TRIP) effect in a broad range of external loads during deformation (*19*, *20*, *28*), which leads to a prolonged uniform elongation and enhanced work-hardening effect.

The typical microCM architecture in the as-printed Ti64-(4.5%)316L (wt %) alloy is shown in Fig. 1, A to C. For example, our scanning electron microscopy–energy-dispersive spectroscopy (SEM-EDS) images (Fig. 1A) show obvious depletion of Ti, Al, and V (Ti64 elements) and enrichment of Fe, Cr, and Ni (316L elements) in one of the swirls within the melt pool. Higher-magnification transmission electron microscopy–EDS (TEM-EDS) line scan results (Fig. 1, B and C) illustrate clearer intragranular concentration gradients alongside the α′-β interphase interface region. The α′ martensitic phase regions have a much lower content of the 316L elements (the β stabilizers) as compared with that of the β phase regions. At the center of one of the α′ and one of the β phase regions, the precise compositions we measured with 3D atom probe tomography (3D-APT) (table S1) are Ti-6.0Al-4.1V-0.9Fe-0.3Cr-0.1Ni-0.01Mo and Ti-5.8Al-3.9V-6.4Fe-1.7Cr-1.6Ni-0.3Mo, respectively [which can also be simplified as Ti64-(1.3%)316L and Ti64-(9.9%)316L]. Therefore, the microCM alloy

[1]Department of Materials Science and Engineering, Hong Kong Institute for Advanced Study, College of Science and Engineering, City University of Hong Kong, Hong Kong, China. [2]Center of Microstructure Science, Frontier Institute of Science and Technology, State Key Laboratory for Mechanical Behavior of Materials, Xi'an Jiaotong University, Xi'an 710049, China. [3]Guangdong–Hong Kong Joint Research and Development Center on Advanced Manufacturing Technology for Light Alloys, Institute of New Materials, Guangdong Academy of Sciences, Guangzhou 510650, China. [4]Department of Materials Science and Engineering, Ohio State University, Columbus, OH 43210, USA.
***Corresponding author. Email: taoyang6-c@my.cityu.edu.hk (T.Y.); wang.363@osu.edu (Y.W.); chainliu@cityu.edu.hk (C.-T.L.)**

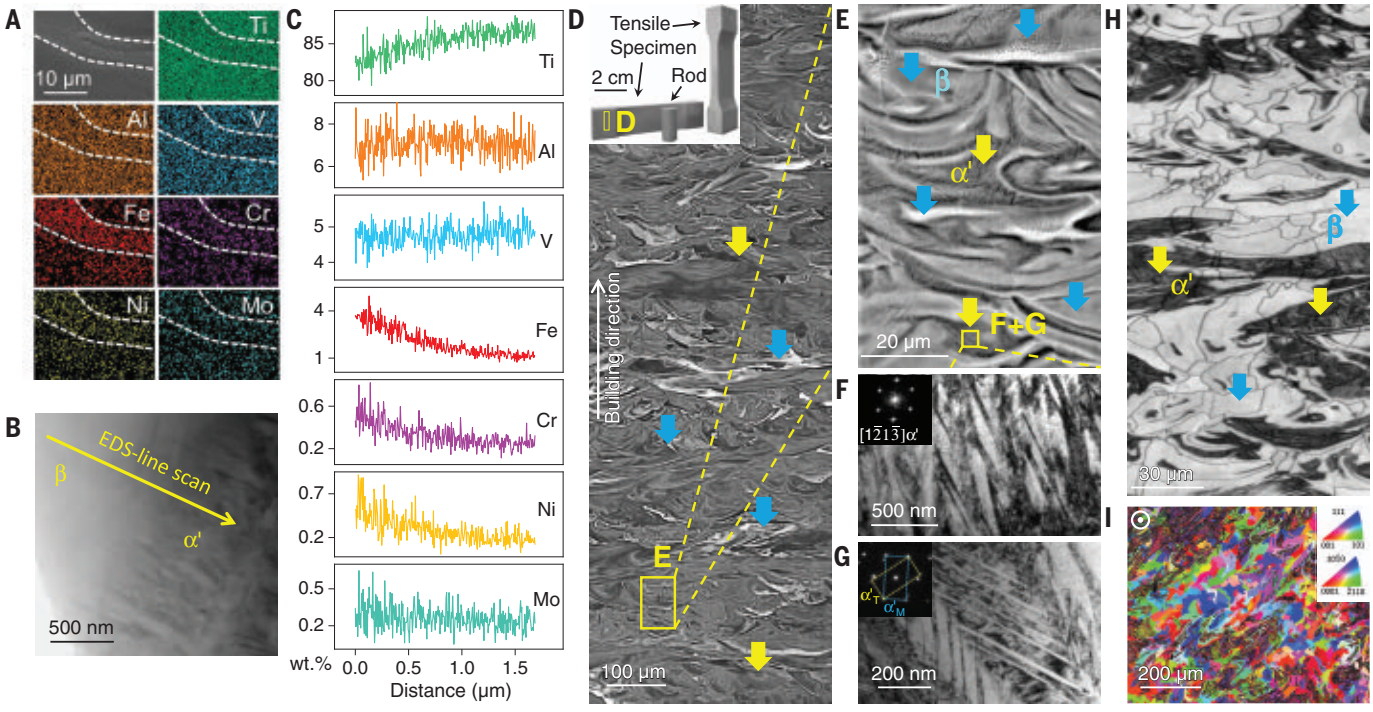


Fig. 1. Micrometer-scale concentration modulation and microstructures of the as-printed Ti64-(4.5%)316L alloy. (**A**) SEM-EDS maps showing the micrometer-scale depletion of Ti, Al, and V and enrichment of Fe, Cr, and Ni elements within one of the swirls of the melt pool. (**B**) Scanning TEM image of the β-α′ interphase region. The yellow arrow denotes the EDS line scan direction in (C). (**C**) Composition profiles across the β-α′ interphase boundary in (B) showing the enrichment of 316L elements (Fe, Cr, Ni, and Mo) in the β phase and depletion of 316L elements in the α′ phase. (**D**) Side-view cross-sectional BSE image showing lava-like microstructure with the clear swirl pattern from Marangoni convection during mechanical mixing in the melt pool. Brighter regions are enriched in heavy elements (Fe, Cr, Ni, and Mo in 316L, blue arrows), and darker regions are enriched in light elements (Ti, Al, and V in Ti64, yellow arrows). The inset shows a photograph of the as-printed Ti64-(4.5%)316L tensile specimen and rod. (**E**) Enlarged BSE image showing the lava-like heterogeneous microstructure with the coexistence of acicular α′ martensite (yellow arrows) and ultrafine β grains with solidification cellular structure (blue arrows). (**F**) Fine acicular α′ martensite as observed by TEM bright-field image. (**G**) Ultrafine twin structure as observed by TEM bright-field image. (**H** and **I**) EBSD images of as-printed Ti64-(4.5%)316L. (H) Band contrast image of side-view cross section showing the ultrafine grain structure without columnar grains. The bright regions (blue arrows) consist of bulk β grains, and the dark regions (yellow arrows) consist of fine acicular α′ martensite. (I) Inverse pole figure (IPF) map of the top-view cross section showing grain orientations of β and α′ phases.

has location-specific compositions throughout the build.

The microCM is produced through partial homogenization of the two alloy melt swirls in the melt pool during L-PBF (*13*, *14*), in which both physical mixing by means of fluid flow with Marangoni motion and chemical mixing by means of atomic diffusion occur between the two alloy melts within the melt pool (*29*, *30*). Although the physical mixing tends to homogenize the melt at the melt-pool scale (creating swirls of the two alloy melts), a complete homogenization (for example, complete dissolution of the 316L elements into Ti64) requires sufficient chemical mixing—sufficient interdiffusion between the adjacent 316L element–rich and Ti64 element–rich swirls in the melt pool. Because of the small melt-pool size and rapid solidification associated with L-PBF (*31*, *32*), a partially homogenized state is preserved after solidification. The measured 316L concentration variation in the microCM Ti64-(4.5%)316L alloy by means of APT, TEM-EDS, and SEM-EDS ranges from ~1.3 to ~9.9 wt % ($\Delta c \approx$ 8.6 wt %) (table S1). This sort of concentration nonuniformity has been both reported and considered undesirable for AM (*13*, *14*, *33*, *34*). However, we actively used this characteristic of L-PBF to achieve microCMs in Ti alloys that lead to a modulated dual-phase (α′ + metastable β) microstructure with excellent mechanical properties.

One direct impact of the microCM is on the phase stability and microstructure modulation in the as-printed alloy. From a side-view cross section of the as-printed Ti64-(4.5%)316L, by using backscattered electron (BSE) imaging, we observed a lava-like microstructure that showed clear swirl patterns of the fluid flow within the melt pools (Fig. 1, D and E). Within each individual melt-pool region, we found two distinct microstructural features within the dark (316L element–poor) and bright (316L element–rich) swirls. The dark swirls consist of fine acicular α′ martensite (yellow arrows) with a typical lath thickness of ~100 nm, along with some ultrafine twin structures (~20 nm of twin spacing) inside (Fig. 1, F and G). The bright swirls, however, consist of a single β phase with a cellular microstructure at the center (blue arrows). We confirmed the highly heterogeneous and fine-scale modulated β + α′ microstructure by using electron backscatter diffraction (EBSD) results (Fig. 1, H and I). Instead of the conventional large columnar grains along the build direction observed in AM Ti64 (*35*–*37*), there are ultrafine grains throughout the build. In addition, the as-printed alloy has a high density (>99.9%) (fig. S1).

The second impact of the microCM is on the metastability-induced mechanical property enhancement. Upon solidification, there exist 316L element–rich and 316L element–poor regions that are due to partial homogenization. Upon further cooling, the 316L element–poor

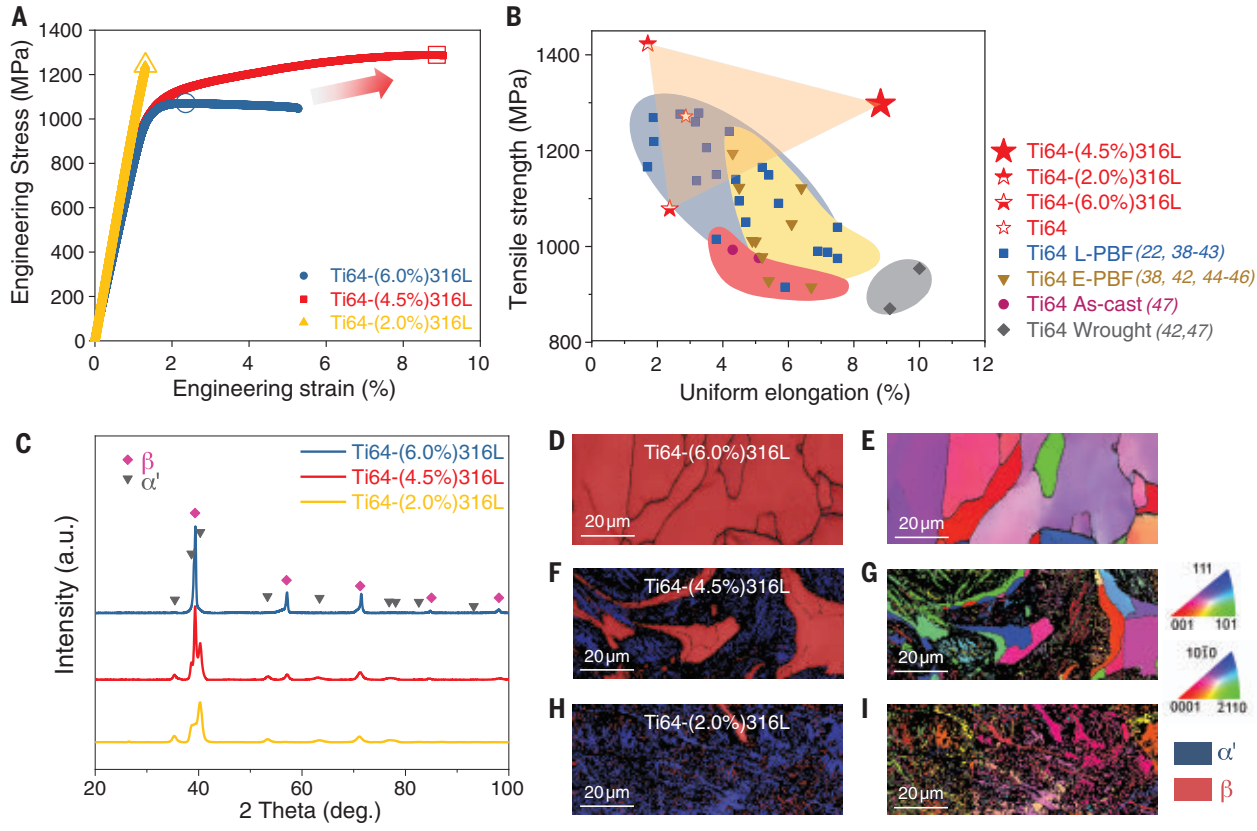


Fig. 2. Microstructure and mechanical properties of the as-printed microCM Ti64-x316L alloy. (**A**) Engineering stress-strain curves of as-printed Ti64-(6.0%)316L, Ti64-(4.5%)316L, and Ti64-(2.0%)316L alloys, respectively. The UTS is marked by big open symbols in the curves. (**B**) Engineering tensile strength versus uniform elongation of the as-printed Ti64-x316L alloy as compared with Ti64 alloy produced by different AM methods and conventional technologies (*22*, *38*–*47*). (**C** to **I**) Influence of microCMs on microstructure of Ti64-x316L alloys. (C) XRD profiles of phase constitution of three alloys. a.u., arbitrary units. [(D) to (I)] EBSD phase images and IPF maps of three different as-printed alloys showing phase constituents and grain orientations.

regions undergo the β-to-α′ martensitic transformation, and the 316L element–rich regions are retained as metastable β phase, which creates a highly dispersed β + α′ dual-phase microstructure. Furthermore, the composition variation within the β phase regions leads to a phase stability variation (*28*), which results in a progressive TRIP effect upon loading and, thus, an excellent work-hardening capacity and prolonged uniform elongation. We performed tensile tests on the as-printed Ti64-(4.5%)316L microCM alloy (Fig. 2A) and obtained a yield strength (σ_y) of 984 ± 14 MPa, an ultimate tensile strength (σ_{UTS}) of 1297 ± 10 MPa, and a uniform elongation (ε_u) of 8.8 ± 0.2% (Fig. 2A) with a ductile fracture mode (fig. S6). Compared with all classes of Ti alloys (α, α + β, and β) that are manufactured by different AM technologies (such as L-PBF, electron beam–PBF (E-PBF), and laser-directed energy deposition (L-DED)] (Fig. 2B and fig. S7) (*22*, *38*–*47*), the microCM Ti64-(4.5%)316L alloy exhibits simultaneously a higher σ_{UTS} and larger ε_u, which breaks the tradeoff between strength and ductility. In addition, the excellent work-hardening ability ($\sigma_{UTS} - \sigma_y$ = 313 ± 11 MPa) is advantageous in structural applications to guarantee a large safety margin before fracture (*48*).

Most additively manufactured Ti alloys lack substantial work hardening (*1*, *47*, *49*). Advanced AM-fabricated Ti64 can achieve a UTS of ~1.2 GPa (*43*, *49*). However, an early onset of plastic instability is usually observed, which leads to a premature failure before ε reaches 4% (Fig. 2B) (*43*, *49*). In general, the β phase volume fraction in as-printed or heat-treated Ti64 ranges from ~0 to 7% (*43*, *49*), whereas in our microCM Ti64-(4.5%)316L alloy, the β phase volume fraction reaches as high as 60% (Fig. 2C). This high-volume fraction is not surprising because all the four elements (Fe, Cr, Ni, and Mo) in 316L powders are strong β stabilizers in Ti alloys. Thus, in the microCM Ti64-(4.5%)316L alloy, the regions enriched in these elements remain as β even under the extremely high cooling rates during L-PBF. However, alloys with a large volume fraction of β phase may not necessarily have an enhanced work-hardening ability. For example, the microCM Ti64-(6.0%)316L alloy consists of mainly the β phase (Fig. 2, C to E) after printing but exhibits little work hardening during deformation (Fig. 2A). What is different about the retained β phase in Ti64-(4.5%)316L microCM alloy is that (i) it is largely metastable because of the amounts of Fe, Cr, Ni, and Mo elements present and (ii) it has a continuous variation in its metastability because of the continuous variations in Fe, Cr, Ni, and Mo concentrations (*28*). Those retained β phase regions with relatively high concentrations of these elements show higher mechanical stability under load and have higher triggering stresses for the stress-induced β-to-α′ martensitic transformation (SIMT) (*28*), whereas those β phase regions with relatively low concentrations of these elements have lower triggering stresses for SIMT (*28*). Thus, upon loading, the SIMT first starts within the less-stable β phase regions and then gradually propagates to those more-stable β phase regions as

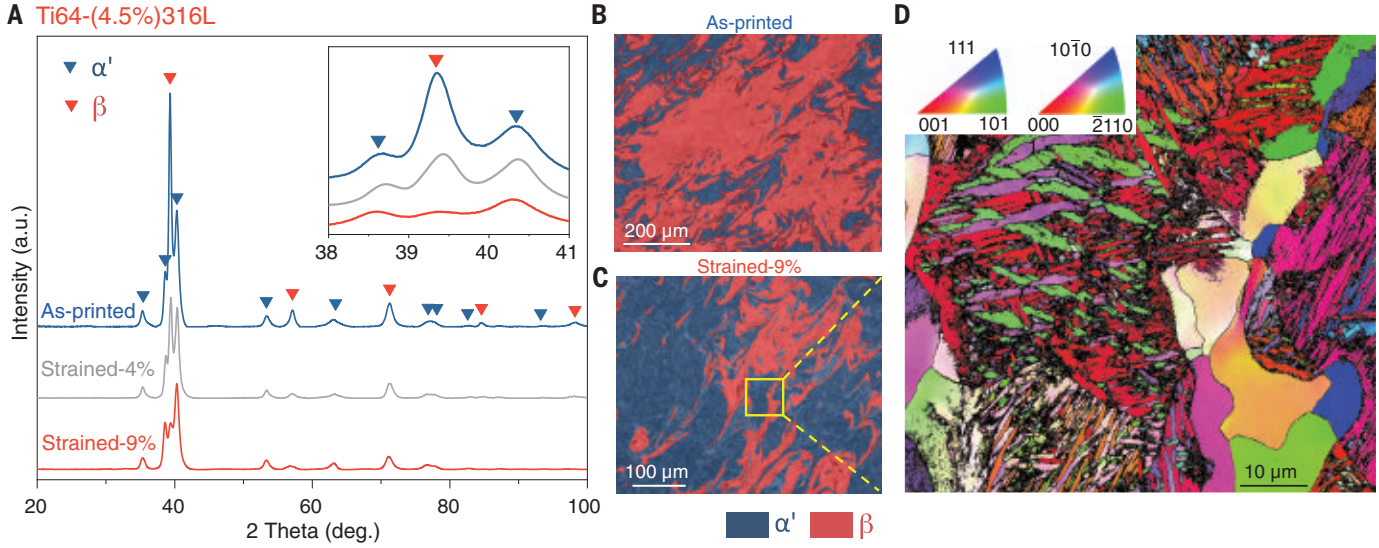
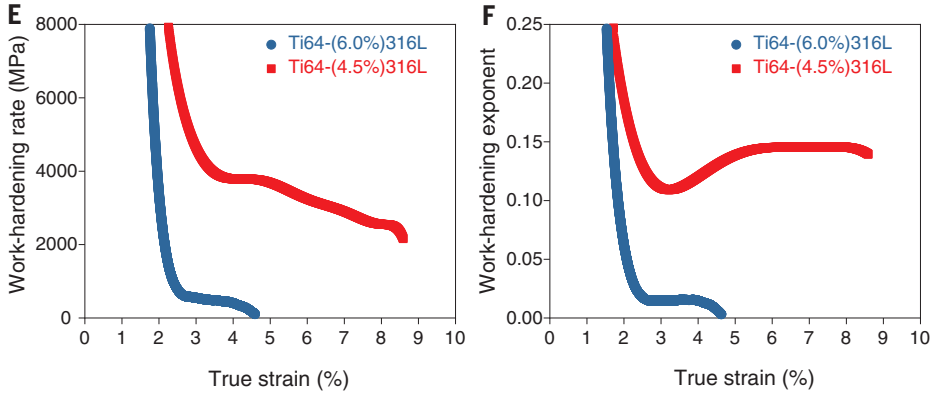


Fig. 3. Work-hardening behavior and microstructure evolution of the microCM alloys. (**A**) XRD profiles of the as-printed Ti64-(4.5%)316L alloy showing the continuous β → α′ phase transformation during tensile test. (**B** and **C**) EBSD phase images showing decrease of β phase volume fraction after being strained to failure. (**D**) Enlarged EBSD IPF map at the yellow box in (C) that shows the morphology of acicular α′ martensite and retained β grains after the tensile test. (**E**) Work-hardening rate curves of the microCM alloys. (**F**) Dynamic change of the work-hardening exponent versus true strain within the uniform deformation process of the microCM alloys.



the applied load increases. This substantially enhances the work-hardening capacity of the microCM Ti64-(4.5%)316L alloy. The x-ray diffraction (XRD) patterns (Fig. 3A) show that, during deformation, the peak intensities of $(110)_\beta$ at 39.4° and $(200)_\beta$ at 57.5° decrease continuously. The calculated β phase fraction decreases from 58 to 25% after deformation, reflecting a continuous deformation-induced phase transformation from the metastable β phase to the α′ martensitic phase, which is also confirmed by the EBSD phase maps (Fig. 3, B and C). The stress-induced martensite is confined within the prior-β grains (Fig. 3D and fig. S8). By contrast, even though microCMs of the β stabilizers also exist in Ti64-(6.0%)316L, most of the β phase in this alloy is too stable (because of the high Fe, Cr, Ni, and Mo contents) to transform to the martensite during deformation (fig. S9), and thus, we did not observe a noticeable TRIP effect. A comparison of the incremental work-hardening behavior of the two alloys as a function of strain is shown in Fig. 3, E and F. After being strained at >2%, the Ti64-(6.0%)316L alloy exhibits a quick

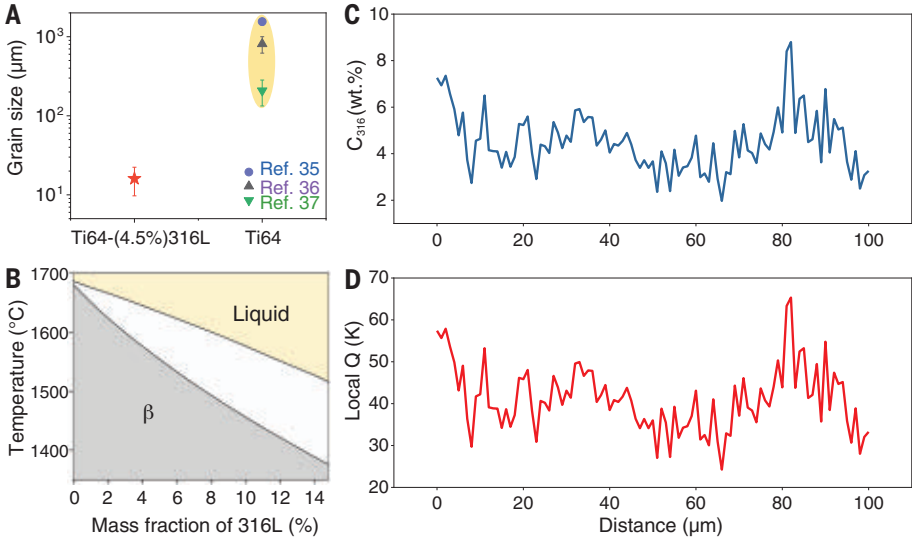


Fig. 4. Effects of microCM on grain refinement. (**A**) Comparison of average grain size between the as-printed Ti64 and microCM Ti64-(4.5%)316L alloys (*35*–*37*). (**B**) Pseudobinary Ti64-316L phase diagram calculated by Thermo-Calc software. (**C**) SEM-EDS long-time line scan curve shows the composition variation through several melt pools. (**D**) Calculated growth-restriction factor Q as a function of local composition on the basis of (C).

drop of the work-hardening rate to <600 MPa, with a small work-hardening exponent. On the contrary, the Ti64-(4.5%)316L alloy exhibits a much higher and gradually decreased work-hardening rate and a much higher and steady work-hardening exponent over a much larger strain range, which reflects a steady and continuous SIMT during the deformation (Fig. 3A).

The high strength of the as-printed microCM Ti64-(4.5%)316L alloy also originates from the fine and highly dispersed mixture of the α′ and metastable β dual-phase microstructure. As they accompany the modulated concentration, the super-refined α′ martensite regions or swirls are dispersed in between the β phase regions or swirls and show an average swirl spacing of ~5 μm between adjacent swirls, which provides a high yield strength to the alloy. When the amount of the 316L powders added is low, the as-printed Ti64-(2.0%)316L alloy comprises mainly super-refined α′ martensite (Fig. 2, C, H, and I) because of the low contents of the β stabilizers on average. The high fraction of the super-refined α′ martensite microstructure makes the alloy strong but brittle (Fig. 2A), having a yield strength of >1400 MPa but fracturing even before yielding for most samples. In view of this, the control of the mean of the microCM by controlling the amount of the 316L powders added is critically important for achieving the above-mentioned dual-phase microstructure for a balanced strength and ductility.

The third important impact of the microCM is on the grain refinement. As has been verified by the EBSD images (for example, Fig. 1H), we measured the average grain size of the as-printed Ti64-(4.5%)316L alloy as ~16 ± 6 μm, which is one of the finest grain structures achieved in an AM Ti alloy (Fig. 4A) (*35*–*37*). The grain size of the as-printed alloy is influenced to a large degree by its increased constitutional supercooling capacity, which is usually measured by the growth restriction factor Q-value (*25*). A high Q-value indicates a high constitutional supercooling, which activates more nuclei and restricts the growth of existing grains (*25*, *26*, *35*), which leads to a finer grain structure (*7*, *50*). On the basis of our Thermo-Calc calculations (Fig. 4B and fig. S10) and combined with the measured local composition (Fig. 4C), the microCM Ti64-(4.5%)316L alloy is readily seen to have an average growth-restriction factor Q-value of ~40 K (Fig. 4D), which is about five times as much as that of Ti64 (Q ≈ 8 K). Because of the local variations of the 316L elements in the microCM alloy (Fig. 4C), a corresponding spatial variation of the Q-value (Fig. 4D) exists from ~24 K in the 316L element–poor regions to ~65 K in the 316L element–rich regions. The 316L element–rich β phase regions (higher Q-values) tend to have smaller grains as compared with those of the 316L element–poor α′ phase regions (smaller Q-values) (Fig. 1H).

We have demonstrated a method of microCM alloy design that is enabled by L-PBF with fine-scale concentration modulations for achieving adaptive microstructures with excellent mechanical properties. The microCM Ti64-(4.5%)316L alloy that is made through partial homogenization has a fine and highly dispersed mixture of α′ + metastable β dual-phase microstructure with fine grains (~16 μm) in its as-printed state. With the super-refined α′ phase regions offering the alloy a high yield strength and the continuous variations in Fe, Cr, Ni, and Mo concentrations in the retained metastable β phase regions offering the alloy a progressive TRIP effect in a broad stress range, the microCM Ti64-(4.5%)316L alloy exhibits a high tensile strength of ~1.3 GPa, a uniform elongation of ~9%, and an excellent work-hardening capacity of >300 MPa in its as-printed state. These excellent properties may enable many advanced applications of the microCM Ti alloys in the aerospace, automotive, chemical, and medical industries. We expect the microCMs to produce different types of spatially modulated phase stabilities and microstructures in different alloy systems (*16*, *17*, *19*, *20*), which will offer some distinctive properties that are superior to those of their homogeneous counterparts.

REFERENCES AND NOTES

1. D. Herzog, V. Seyda, E. Wycisk, C. Emmelmann, *Acta Mater.* **117**, 371–392 (2016).
2. W. E. Frazier, *J. Mater. Eng. Perform.* **23**, 1917–1928 (2014).
3. W. J. Sames, F. A. List, S. Pannala, R. R. Dehoff, S. S. Babu, *Int. Mater. Rev.* **61**, 315–360 (2016).
4. C. Panwisawas, Y. T. Tang, R. C. Reed, *Nat. Commun.* **11**, 2327 (2020).
5. Y. M. Wang *et al*., *Nat. Mater.* **17**, 63–71 (2018).
6. J. H. Martin *et al*., *Nature* **549**, 365–369 (2017).
7. D. Zhang *et al*., *Nature* **576**, 91–95 (2019).
8. P. Kürnsteiner *et al*., *Nature* **582**, 515–519 (2020).
9. S. A. M. Tofail *et al*., *Mater. Today* **21**, 22–37 (2018).
10. T. M. Pollock, *Nat. Mater.* **15**, 809–815 (2016).
11. M. Zhang *et al*., *Sci. Adv.* **6**, eaba5581 (2020).
12. D. Raabe, C. C. Tasan, E. A. Olivetti, *Nature* **575**, 64–74 (2019).
13. B. Vrancken, L. Thijs, J.-P. Kruth, J. van Humbeeck, *Acta Mater.* **68**, 150–158 (2014).
14. I. Yadroitsev, P. Krakhmalev, I. Yadroitsava, *JOM* **69**, 2725–2730 (2017).
15. D. Gu *et al*., *Science* **372**, eabg1487 (2021).
16. T. Zhang, D. Wang, Y. Wang, *Acta Mater.* **196**, 409–417 (2020).
17. R. Ding *et al*., *Sci. Adv.* **6**, eaay1430 (2020).
18. W. W. Sun, Y. X. Wu, S. C. Yang, C. R. Hutchinson, *Scr. Mater.* **146**, 60–63 (2018).
19. J. Zhu, D. Wang, Y. Gao, T.-Y. Zhang, Y. Wang, *Mater. Today* **33**, 17–23 (2020).
20. J. Zhu, Y. Gao, D. Wang, T.-Y. Zhang, Y. Wang, *Acta Mater.* **130**, 196–207 (2017).
21. B. E. Carroll, T. A. Palmer, A. M. Beese, *Acta Mater.* **87**, 309–320 (2015).
22. C. J. Todaro *et al*., *Nat. Commun.* **11**, 142 (2020).
23. X. Wu *et al*., *Mater. Des.* **25**, 137–144 (2004).
24. L. Thijs, F. Verhaeghe, T. Craeghs, J. van Humbeeck, J.-P. Kruth, *Acta Mater.* **58**, 3303–3312 (2010).
25. M. J. Bermingham, S. D. McDonald, M. S. Dargusch, D. H. StJohn, *J. Mater. Res.* **23**, 97–104 (2008).
26. D. Zhang *et al*., *Metall. Mater. Trans., A Phys. Metall. Mater. Sci.* **51**, 4341–4359 (2020).
27. Materials and methods are available as supplementary materials.
28. C. Li, J. H. Chen, X. Wu, W. Wang, S. van der Zwaag, *J. Mater. Sci.* **47**, 4093–4100 (2012).
29. S. J. Clark *et al*., *IOP Conf. Ser.: Mater. Sci. Eng.* **861**, 12010 (2020).
30. Q. Guo *et al*., *Addit. Manuf.* **31**, 100939 (2020).
31. R. Cunningham *et al*., *Science* **363**, 849–852 (2019).
32. W. Huang, H. Wang, T. Rinker, W. Tan, *Mater. Des.* **195**, 109056 (2020).
33. A. M. Vilardell *et al*., *Addit. Manuf.* **36**, 101436 (2020).
34. M. Simonelli *et al*., *Metall. Mater. Trans., A Phys. Metall. Mater. Sci.* **51**, 2444–2459 (2020).
35. M. J. Bermingham, D. H. StJohn, J. Krynen, S. Tedman-Jones, M. S. Dargusch, *Acta Mater.* **168**, 261–274 (2019).
36. Y. M. Ren *et al*., *Acta Mater.* **132**, 82–95 (2017).
37. P. W. Liu *et al*., *J. Mater. Process. Technol.* **257**, 191–202 (2018).
38. H. K. Rafi, N. V. Karthik, H. Gong, T. L. Starr, B. E. Stucker, *J. Mater. Eng. Perform.* **22**, 3872–3883 (2013).
39. T. Vilaro, C. Colin, J. D. Bartout, *Metall. Mater. Trans., A Phys. Metall. Mater. Sci.* **42**, 3190–3199 (2011).
40. W. Xu *et al*., *Acta Mater.* **85**, 74–84 (2015).
41. G. Kasperovich, J. Hausmann, *J. Mater. Process. Technol.* **220**, 202–214 (2015).
42. L. Facchini *et al*., *Rapid Prototyping J.* **16**, 450–459 (2010).
43. W. Xu, E. W. Lui, A. Pateras, M. Qian, M. Brandt, *Acta Mater.* **125**, 390–400 (2017).
44. C. de Formanoir *et al*., *Acta Mater.* **162**, 149–162 (2019).
45. C. de Formanoir, S. Michotte, O. Rigo, L. Germain, S. Godet, *Mater. Sci. Eng. A* **652**, 105–119 (2016).
46. H. Gong *et al*., *Mater. Des.* **86**, 545–554 (2015).
47. M. Niinomi, *Mater. Sci. Eng. A* **243**, 231–236 (1998).
48. T. Yang *et al*., *Science* **362**, 933–937 (2018).
49. T. Voisin *et al*., *Mater. Des.* **158**, 113–126 (2018).
50. M. A. Easton, D. H. StJohn, *Acta Mater.* **49**, 1867–1878 (2001).

ACKNOWLEDGMENTS

Funding: The authors from City University of Hong Kong (CityU) are grateful for the internal funding from City University of Hong Kong under the programs 9042635, 9360161, and 9380060. T.Y. and C.-T.L. are also grateful for the financial support from the Hong Kong Institute for Advanced Study (grant no. 9360157). D.W. and T.Z. acknowledge the support of the National Key Research and Development Program of China (grant no. 2016YFB0701302) and the National Natural Science Foundation of China (grants nos. 51671156 and 51671158). Z.H. acknowledges the GDAS's Project of Science and Technology Development (2019GDASYL-0203002). Y.W. acknowledges the support by the US National Science Foundation (grant DMR – 1923929). **Author contributions:** C.-T.L., Y.W., and T.Z. conceived the idea and designed the research. T.Z. designed the processing parameters and prepared the specimens. T.Z. characterized the specimens by means of SEM, EDS, EBSD, and XRD and conducted Thermo-Calc calculations. T.Z. and Z.H. performed the micro-CT analysis. Z.H. and A.W. operated the 3D printer and manufactured the samples. T.Y. carried out the TEM. J.L. conducted the 3D-APT. H.K. and T.Z. performed the mechanical testing. T.Z., C.-T.L., Y.W., T.Y., and D.W. analyzed the data and discussed the results. T.Z., Y.W., C.-T.L., T.Y., and W.K. wrote and revised the manuscript. All authors reviewed and contributed to the final manuscript. **Competing interests:** C.-T.L. and T.Z. are inventors on a US patent (application no. 17/012,507) that has been filed related to the alloy design described in this work. The rest of the authors declare no conflict of 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.abj3770
Materials and Methods
Figs. S1 to S11
Tables S1 and S2
References (*51*–*98*)

9 May 2021; accepted 8 September 2021
10.1126/science.abj3770

POACHING IMPACTS

Ivory poaching and the rapid evolution of tusklessness in African elephants

Shane C. Campbell-Staton[1,2,3]*†, Brian J. Arnold[4,5]†, Dominique Gonçalves[6,7], Petter Granli[8], Joyce Poole[8], Ryan A. Long[9], Robert M. Pringle[1]

Understanding the evolutionary consequences of wildlife exploitation is increasingly important as harvesting becomes more efficient. We examined the impacts of ivory poaching during the Mozambican Civil War (1977 to 1992) on the evolution of African savanna elephants (*Loxodonta africana*) in Gorongosa National Park. Poaching resulted in strong selection that favored tusklessness amid a rapid population decline. Survey data revealed tusk-inheritance patterns consistent with an X chromosome–linked dominant, male-lethal trait. Whole-genome scans implicated two candidate genes with known roles in mammalian tooth development (*AMELX* and *MEP1a*), including the formation of enamel, dentin, cementum, and the periodontium. One of these loci (*AMELX*) is associated with an X-linked dominant, male-lethal syndrome in humans that diminishes the growth of maxillary lateral incisors (homologous to elephant tusks). This study provides evidence for rapid, poaching-mediated selection for the loss of a prominent anatomical trait in a keystone species.

The selective killing of species that bear anatomical features such as tusks and horns is the basis of a multibillion-dollar illicit wildlife trade (*1*) that poses an immediate threat to the survival of ecologically important megafauna worldwide (*2*, *3*). Megaherbivores are especially vulnerable to overharvesting because of their large habitat requirements, small population sizes, and long generation times (*4*, *5*). As ecosystem engineers, these species also behaviorally regulate ecological processes (*5*–*8*); anthropogenic selection on phenotypes that influence these behaviors may, therefore, have cascading effects on ecosystem functioning. However, most work that details human-driven selection has focused on smaller species in which evolutionary change is more readily studied (*9*, *10*). It remains unclear to what extent, at what rates, and through what mechanisms harvest-induced phenotypic change occurs in the world's largest land animals.

Warfare is associated with intensified exploitation and population declines of wildlife throughout Africa (*11*), and organized violence has long been intertwined with the ivory trade (*12*–*14*). In Gorongosa National Park, the Mozambican Civil War (1977 to 1992) reduced large-herbivore populations by >90% (*15*), and armies on both sides of the conflict targeted elephants for ivory (*15*, *16*). Intensive poaching in Africa has been associated with an increase in the frequency of tuskless elephants, exclusively (or nearly so) among females (table S3). No record of tuskless male elephants within Gorongosa National Park exists (table S2). Analyses of historical video footage and contemporary sighting data (supplementary materials) show that the precipitous decline of the Gorongosa elephant population was accompanied by a nearly threefold increase in the frequency of tuskless females, from 18.5% (n = 52) to 50.9% (n = 108) (two-sample equality of proportions test with continuity correction, $P < 0.001$) (Fig. 1A).

To test whether the increased frequency of female tusklessness was a chance event associated with the severe population bottleneck (*17*), we simulated the observed population decline in Gorongosa from 1972 (n = 2542 individuals) to 2000 (n = 242) (*15*) under a scenario of equal survival probabilities for tusked and tuskless females (see methods). On the basis of these simulations, the observed increase in tusklessness is extremely unlikely to have occurred in the absence of selection (hypergeometric distribution, $P = 1.8 \times 10^{-15}$) (Fig. 1B). The relative survival of tuskless females across this 28-year period was estimated to be more than five times that of tusked individuals (maximum-likelihood estimate = 5.13, 95% confidence interval 3.98 to 6.60) (Fig. 1C). Thus, we conclude that the population bottleneck in Gorongosa was accompanied by strong selection favoring the tuskless phenotype.

If there were strong selection against tusked elephants, we might also observe divergent genomic signatures of population-size change between the two tusk morphs. We sequenced whole genomes from blood samples of 18 female elephants (n = 7 tusked, 11 tuskless). We mapped sequence reads to the annotated African savanna elephant genome (Loxafr3.0) and generated alignments with ~30× coverage for 13 samples and 14× coverage for 5 samples (supplementary materials). Using the 30× coverage samples (n = 6 tusked, 7 tuskless), we calculated Tajima's D (*18*) genome-wide in nonoverlapping 10-kb windows. Both groups displayed a slight excess of rare variants, indicated with negative D values (tuskless: −0.27, tusked: −0.2). However, tusked samples had significantly fewer rare variants than tuskless samples (Welch's two-sample t test: $P < 0.0001$) (Fig. 1D and supplementary materials), which is consistent with a more severe population contraction of tusked individuals.

To evaluate the evolutionary response to selection, we quantified the frequency of tusk phenotypes among adult females born after the war (estimated birth years 1995 to 2004). We found that tusklessness among female offspring of survivors (33%, n = 91) remained significantly elevated over the pre-conflict proportion (18.5%, two-sample equality of proportions test with continuity correction, $P = 0.046$) (Fig. 1A) and was greater than expected in the absence of selection (hypergeometric distribution, $P = 4.3 \times 10^{-8}$) (Fig. 1B). These results indicate a heritable genetic basis for tusklessness and an evolutionary response to poaching-induced selection in Gorongosa.

Given the evidence for heritability and female-specificity of tusklessness in Gorongosa, we hypothesized that the phenotype is genetically inherited through a sex-linked locus (*17*, *19*–*21*). We therefore searched for a pattern of inheritance that could explain the observed variation in tusk morphology. Phenotypes displaying extreme female bias are commonly attributed to X chromosome–linked dominant inheritance with male lethality (*22*). Accordingly, we used mother-offspring phenotype surveys in Gorongosa to test the a priori hypothesis that tusklessness is an X-linked dominant, male-lethal trait governed by a single locus. Under this hypothesis, we expect two-tusked females (X_+X_+) and males (X_+Y) to carry only the unaffected allele (X_+). As such, X_+X_+ mothers should exclusively produce X_+X_+ daughters. Furthermore, tuskless females should appear only in the heterozygous state (X_+X_-) owing to male lethality (females would always inherit the unaffected allele from X_+Y fathers); thus, X_+X_- mothers should produce daughters with a 1:1 ratio of X_+X_- and X_+X_+ phenotypes, and only 50% of male offspring (X_+Y) conceived by X_+X_- mothers should be viable. As a result, two-thirds of offspring born to X_+X_- mothers should be female, assuming that all

[1]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08540, USA. [2]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA. [3]Institute for Society and Genetics, University of California, Los Angeles, CA 90095, USA. [4]Department of Computer Science, Princeton University, Princeton, NJ 08540, USA. [5]Center for Statistics and Machine Learning, Princeton University, Princeton, NJ 08540, USA. [6]Gorongosa National Park, Sofala 00000, Mozambique. [7]Durrell Institute of Conservation and Ecology, University of Kent, Canterbury CT2 7NR, UK. [8]ElephantVoices, San Francisco, CA 94111, USA. [9]Department of Fish and Wildlife Sciences, University of Idaho, Moscow, ID 83844, USA.
*Corresponding author. Email: scampbellstaton@princeton.edu
†These authors contributed equally to this work.

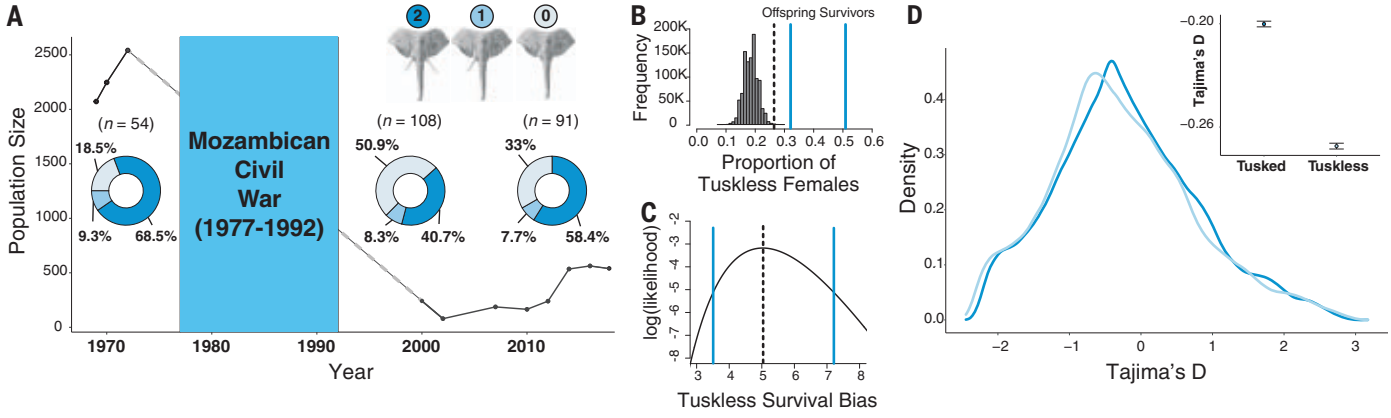


Fig. 1. Demographic shifts during the Mozambican Civil War and evidence of selection for tusklessness in Gorongosa National Park. (**A**) Change in population size and tusk morphology. The line depicts the minimum elephant population size in Gorongosa National Park by year from aerial censuses (*30*); the dashed segment indicates a period for which no robust census data are available. The pie charts show the proportion of two-tusked (dark blue), single-tusked (blue), and tuskless (light blue) females observed prewar (n = 54), in survivors of the war (n = 108), and in the first generation born postwar (n = 91) on the basis of contemporary surveys and historical photos. Individuals for which tusk morphology could not be accurately determined are not represented. (**B**) Simulated samples from a Wallenius distribution of the proportion of female tusklessness expected after the population bottleneck, assuming no selection for tusklessness. Solid vertical lines indicate the observed proportions of tuskless females among survivors of the war and the first postwar generation (born between 1995 and 2004). The dashed line shows the upper 1% quantile of the simulated distribution. (**C**) Log(likelihood) of the increased survivorship of tuskless females observed in Gorongosa, given a range of survival odds of tuskless relative to tusked females. The x axis represents the odds ratio of survival for tuskless individuals compared with that of their tusked counterparts. The dotted line shows the maximum-likelihood estimate; solid blue lines indicate the 95% confidence interval. (**D**) Density plot of genome-wide estimates of Tajima's D for two-tusked (dark blue) and tuskless (light blue) morphs, calculated in 10-kb windows. (Inset) Mean ± 1 SE of genome-wide D for each tusk morph.

$X_$ carriers display the phenotype (i.e., complete penetrance).

We found that 91.3% of daughters born to two-tusked mothers carried two tusks (n = 21 two-tusked, 1 tuskless, 1 one-tusked) (Fig. 2A). By contrast, the daughters of tuskless mothers displayed approximately equal proportions of tusked and tuskless phenotypes [n = 19 two-tusked (40.9%), 21 tuskless (44.7%), two-sample equality of proportions test with continuity correction, P = 0.42] (Fig. 2A). The mothers of both tusk morphs were observed with daughters displaying an intermediate one-tusked phenotype [two-tusked mothers: n = 1 (4.3%), tuskless mothers: n = 7 (14.9%), equality of proportions test P = 0.37]. Tuskless mothers also displayed a biased offspring sex ratio (≤5 years old, n = 67); 65.7% were female, which differs significantly from the null hypothesis of equal sex ratios (exact binomial test, P = 0.027) and is statistically indistinguishable from the 66.7% female bias expected under complete male lethality (P = 0.90) (Fig. 2B). We found no evidence for sex bias among tusked females (54.2% female offspring, n = 48, P = 0.67) (Fig. 2B), and previous research has shown no general sex-biased birth in African elephants (*23*), which suggests that the observed skew in offspring sex ratio is correlated with expression of the tuskless phenotype.

Altogether, 87.1% of mother-offspring phenotypic associations were consistent with a single-locus X-linked dominant model of inheritance, and the sex bias associated with tusklessness was within 1% of that expected under complete male lethality. The unexplained variability in the trait, including the presence of unilateral tusklessness (if genetic), suggests that epistatic interactions between at least two loci may influence the expression of tusk morphology. Genotype-phenotype relationships associated with variation in dental morphogenesis are known to be highly variable, and distinctive mutations have been identified as population or even family specific (*24*). Furthermore, epigenetic patterning and mosaic X-chromosome inactivation can result in the variable phenotypic expression of identical mutations associated with dental agenesis, even between monozygotic twins (*25*). Nevertheless, the Gorongosa data support the hypothesis that the tuskless phenotype is controlled by at least one X-linked dominant, male-lethal locus of large effect, with possible additional modifier loci affecting phenotypic expression of the trait.

We used whole-genome data to identify the putative major-effect locus underpinning the hypothesized X-linked dominant inheritance of the tuskless phenotype and searched for signatures of selection associated with recent intensive poaching. We analyzed genomes to identify the strongest candidate loci for selection on a sex-biased, male-lethal trait. We first conducted a genome-wide scan for loci that exhibit evidence of strong recent selection, specific to the tuskless phenotype, and focused on genomes of tuskless individuals (n = 11). Although tusklessness has not swept to fixation in the Gorongosa population (Fig. 1), we would expect strong signals of recent selection when analyzing genomes from the tuskless subgroup.

Using a sliding window analysis of 10-kb windows with a step size of 2 kb, we scanned for two signatures that we expected under the hypothesis of selection on an X-linked dominant, male-lethal trait. Using 30× coverage samples (6 tusked, 7 tuskless), we first searched for genomic regions that displayed excess heterozygosity within alleles private to the tuskless morph and quantified heterozygosity as deviations from Hardy-Weinberg equilibrium. We next searched for loci displaying highly correlated mutations among tuskless samples [measured as linkage disequilibrium (LD)]. We quantified LD by using single-nucleotide polymorphism (SNP) vectors (μLD), as implemented in the RAiSD (*26*) software.

Subsequently, we compared all tusked (n = 7) and tuskless (n = 11) genomes to search for

patterns of genetic divergence between tusk morphs. Under the hypothesized model of inheritance, we expected the genomes of tusked individuals to lack the specific mutation(s) causing tusklessness. Tusked individuals should also be less likely to harbor SNPs that are linked to the causal mutation(s). We quantified genetic differentiation by using both normalized differences in allele frequency, F_{ST} (*27*), and the average number of pairwise differences, D_{XY} (*28*). Genomic windows that were significant outliers (above the 95% quantile of the genome-wide distribution) for all four of these summary statistics (heterozygosity, LD, F_{ST}, and D_{XY}) were considered candidate loci for selection on tusklessness. The overlap of outliers across these four statistics revealed 305 candidate windows (Fig. 2C).

If tusklessness is an X-linked dominant trait, then the major-effect locus should reside on the X chromosome. We therefore filtered the 305 candidate windows for those located on the X chromosome and found 8 windows that fell within two contiguous genomic intervals. Five of these windows overlap an ~100-kb region that contains the X-linked isoform of amelogenin (*AMELX*) (Fig. 2D), which encodes an extracellular matrix protein involved in biomineralization of enamel and putatively regulates periodontium formation and cementum-associated genes (*29*, *30*). Several mutations within this locus are associated with enamel hypomineralization and tooth brittleness in humans (*30*).

A genomic deletion in the syntenic region of the human X chromosome (Xp22.2), which encompasses *AMELX* and several adjacent genes, results in amelogenesis imperfecta accompanied by an X-linked dominant, male-lethal syndrome (*31*). In such cases, women display several craniofacial abnormalities, including microdonty and/or agenesis of the maxillary lateral incisors (*31*), which are homologous to elephant tusks. Notably, skewed X-chromosome inactivation in amelogenesis imperfecta contributes to pronounced phenotypic variation in heterozygotes (*32*). Previous studies have shown that male survival can be rescued in mice engineered with a deletion in this region by forced expression of the human holocytochrome c-type synthetase (*HCCS*) gene, which lies directly adjacent to *AMELX* (*33*). The high degree of LD that we observed across this region of the tuskless elephant X chromosome (Fig. 2D and fig. S1) suggests that physical proximity between *AMELX* and neighboring male-lethal loci may underlie the inferred association between tusklessness and male lethality in the Gorongosa population. The remaining three candidate windows of the X chromosome encompass the unprocessed pseudogene *FAM115B* and an adjacent intergenic region. We have found no known link between *FAM115B* and odontogenesis or craniofacial development.

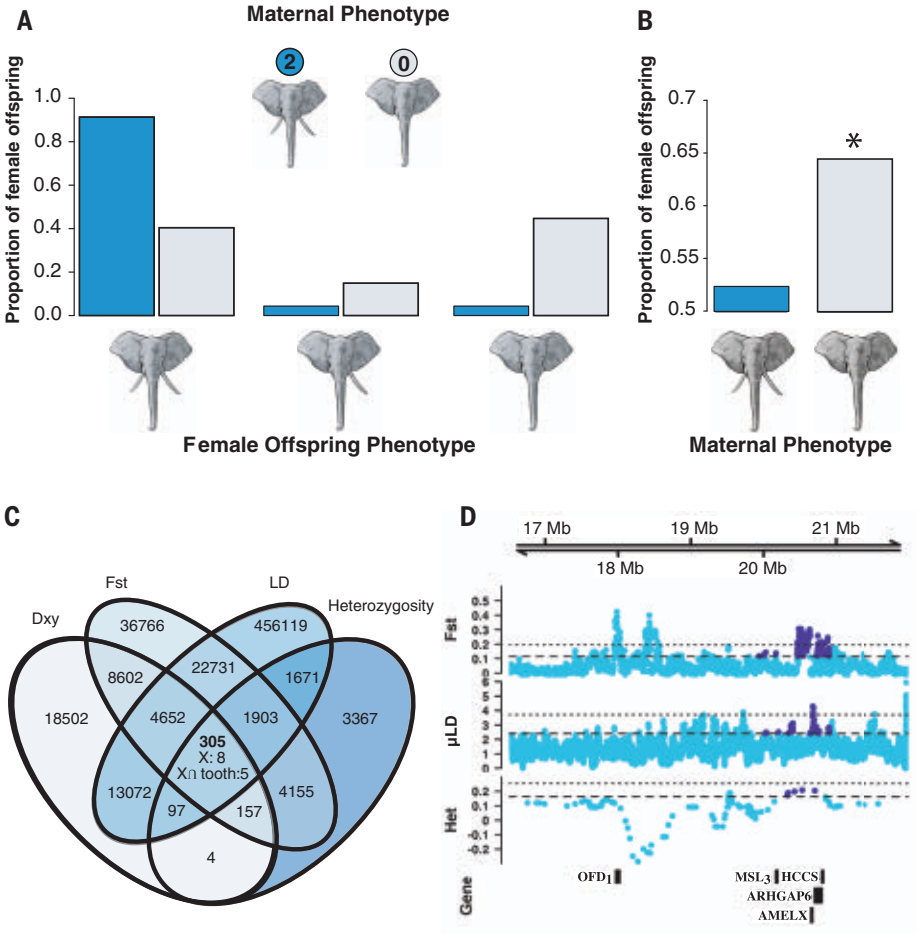


Fig. 2. Evidence for X-linked dominant inheritance with male lethality and an underpinning candidate locus. (**A** and **B**) Inference of inheritance patterns from population offspring survey. (A) Observed phenotypic distribution of female offspring with two-tusked (dark blue) and tuskless (light blue) mothers. (B) Observed sex ratio of offspring for tusked and tuskless females. Asterisk indicates significant deviation from a 1:1 sex ratio among tuskless mothers (binomial test, $P = 0.027$) but not tusked mothers ($P = 0.67$). (**C** and **D**) Evidence for selection on sex-linked candidate locus. (C) Venn diagram of four summary statistics computed in sliding windows across the genome, showing the numbers of overlapping windows in the 5% tails of each statistical distribution. Summary statistics include genetic differentiation between tusked and tuskless samples (F_{ST} and D_{XY}), along with LD and deviation in heterozygosity from Hardy-Weinberg equilibrium ("Heterozygosity") within tuskless samples. Included in the four-way intersection are the number of windows on the X chromosome and the subset of these that overlap known tooth genes (X ∩ tooth). (D) Magnified Manhattan plots of F_{ST}, LD, and heterozygosity (Het) show the genomic location of five X ∩ tooth windows that are contiguous and overlap *AMELX* and flanking regions. Dashed and dotted lines indicate upper 5% and 1% quantiles, respectively. Dark blue dots represent outlier windows ($P < 0.05$) within the candidate region. Light blue dots represent nonoutlier windows. The candidate region contains 23 genes, but for clarity, only genes known to be involved in tooth development and/or male lethality are labeled.

Given the potential influence of epistasis on expression of tusk morphology in Gorongosa, we conducted additional genome-wide scans for regions displaying extreme genetic divergence between tusk morphs (F_{ST} or D_{XY}, $P < 0.001$) (Fig. 3A) and low genetic diversity specific to the tuskless morph (relative diversity = $\pi_{tuskless} - \pi_{tusked}$, $P < 0.001$. Three contiguous genomic windows met these two criteria (F_{ST} $P < 0.001$, relative diversity $P < 0.001$) (Fig. 3B). This region overlaps with a single autosomal gene on chromosome 1: *MEP1a*. This gene encodes

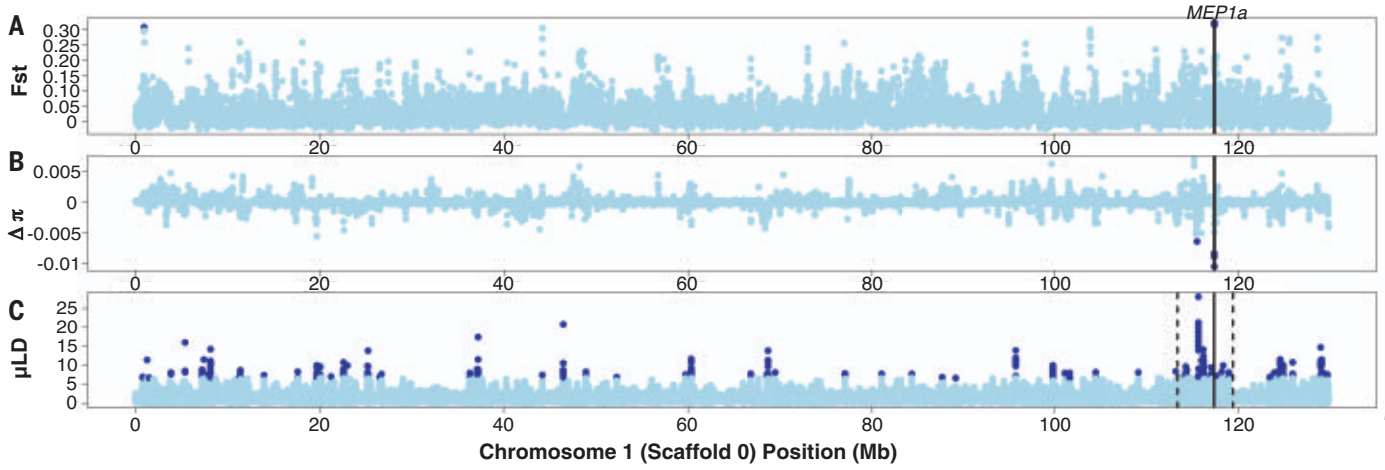


Fig. 3. Autosomal candidate locus for tusklessness in Gorongosa elephants. (**A** to **C**) Scaffold 0 within African elephant chromosome 1, which highlights a candidate locus for tusklessness, *MEP1a*. (A) Genetic divergence (F_{ST}) between two-tusked and tuskless morphs. (B) Relative diversity ($\Delta\pi = \pi_{tuskless} - \pi_{two\text{-}tusked}$) among tuskless individuals. Low values represent regions with low relative genetic diversity in tuskless morphs. (C) Linkage disequilibrium quantified by using SNP vectors (μLD) among tuskless samples. Extreme outliers in each distribution ($P < 0.001$) are indicated with dark blue dots. Light blue dots indicate nonextreme values. Solid vertical lines represent the position of *MEP1a* along chromosome 1 of the elephant genome. Dashed lines represent the boundaries of the surrounding region that displays a high density of elevated LD.

meprin subunit alpha, a matrix metalloprotein that plays an important role in dentin mineralization by processing a precursor, dentin sialophosphoprotein (DSPP). Abnormalities in DSPP are associated with several odontogenic disorders, including dentin dysplasia, which results in malformation of the tooth root and premature tooth loss (*34*). *MEP1a*$^{-/-}$ mice display significant alterations in dentin bone mineral density (*35*). These three windows are nested within an extended genomic interval (~6 Mb) that displays elevated LD ($P < 0.001$) (Fig. 3C and fig. S2), which suggests recent positive selection across this region. Together, *AMELX* and *MEP1a* have functional associations with the development of several distinct regions of the mammalian tooth, including enamel, dentin, cementum, and periodontium (Fig. 4). However, analyses of divergent polymorphisms and structural variants (including deletions, duplications, and copy number variants) between tusk morphs did not reveal obvious causal genetic variants for either locus (supplementary materials).

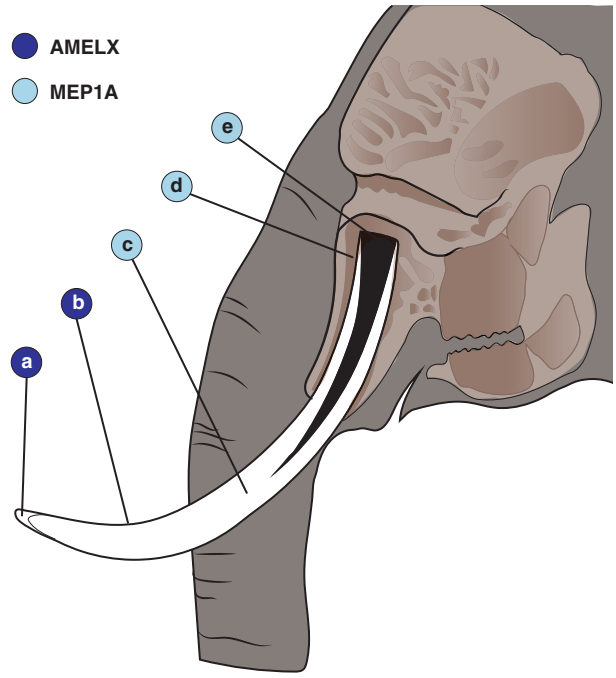


Fig. 4. Putative functional effects of candidate loci on tusk morphology. A cross section of an African elephant tusk shows the anatomical position of (a) enamel, (b) cementum, (c) dentin (ivory), (d) periodontium, and (e) root of the tusk. Dark blue circles indicate regions known or proposed to be affected by candidate gene *AMELX*. Light blue circles are proposed to be affected by candidate gene *MEP1a*. Neither gene is known to affect the formation of the dental pulp (black interior of cross section).

In summary, human-mediated selection for tusklessness during the Mozambican Civil War appears to be driven by recent selection on at least one X-linked locus (*AMELX*) and one autosomal locus (*MEP1a*). Physical linkage between *AMELX* and proximate male-lethal loci on the X chromosome, such as *HCCS* (*31*, *33*), may underpin the proposed X-linked dominant, male-lethal inheritance of tusklessness in the Gorongosa population. If our interpretation is correct, this study represents a rare example of human-mediated selection favoring a female-specific trait despite its previously unknown deleterious effect in males (sexually antagonistic selection). Given the timeframe of selection, speed of evolutionary response, and known presence of the selected phenotype before the selective event, the selection of standing genetic variation at these loci is the most plausible explanation for the rapid rise of tusklessness during this 15-year period of conflict. However, the exact genetic and developmental mechanisms leading to tusklessness and/or male nonviability remain unresolved. Although tuskless males do not occur in Gorongosa or in surveys of large sample sizes from Africa's most intensively studied elephant populations (*17*, *21*, *36*, *37*), there are anecdotal reports of tuskless males in several locations (*20*, *38*, *39*). We are unaware

of any study that has firmly established a frequency of tuskless males beyond what could plausibly be explained by rare injuries or observer error (supplementary text and table S3), but we cannot rule out the possibility of alternative genetic mechanisms and/or genotype–environment interactions. Furthermore, intermediate single-tusked phenotypes commonly co-occur in family groups that also include bilaterally tuskless females (*17*, *20*, *37*). Although the evidence from Gorongosa is consistent with an X-linked dominant, male-lethal trait, continent-wide patterns of tusk expression and heritability may be the result of geographic variation in LD between *AMELX* and adjacent male-lethal loci, additional loci elsewhere in the species' genome, individual variation in patterns of X-chromosome inactivation, or some entirely different genetic mechanism. Further study is needed to establish the exact number and identity of causal variants that encode tusklessness, and comparative studies across multiple populations will be necessary to reveal the geographic structure of genetic variation and inheritance underlying the trait.

Social conflict and commercial harvest can intertwine to devastate animal populations (*11*, *40*). However, most known instances of harvest-induced evolutionary change occur gradually over longer time periods, and the selective effects of harvest can be difficult to disentangle from other factors (*9*, *41*–*44*). Our study shows how a sudden pulse of civil unrest can cause abrupt and persistent evolutionary shifts in long-lived animals even amid extreme population decline. In Gorongosa, recovery of both elephant abundance and ancestral tusk morphology may be crucial for ecosystem restoration. Elephant tusks are multipurpose tools that are used for excavating subterranean food and minerals (*45*, *46*) and gouging and peeling bark, which can kill trees (*47*, *48*). These behaviors can catalyze forest-to-grassland transitions at large scales (*45*) and create habitat for other species at local scales (*49*, *50*). Accordingly, a population-wide increase in tusklessness may have downstream impacts such as reduced bioturbation, shifts in plant species composition, reduced spatial heterogeneity, and increased tree cover—any of which could affect myriad other ecosystem properties. Elsewhere, evolution in species that perform key ecological functions has exerted potent effects on food-web structure, community composition, and nutrient transport (*51*, *52*). Restoration of these functions may require disproportionately longer time scales than the initial selection event (*44*) and may thus constrain the pace of rewilding efforts. Understanding the dynamics of rapid evolution in the Anthropocene is therefore essential, not only for revealing the biological impacts of contemporary human activities but also for designing strategies to mitigate them.

REFERENCES AND NOTES

1. S. K. Wasser *et al*., *Sci. Adv.* **4**, eaat0625 (2018).
2. S. Wasser *et al*., *Science* **327**, 1331–1332 (2010).
3. D. Biggs, F. Courchamp, R. Martin, H. P. Possingham, *Science* **339**, 1038–1039 (2013).
4. M. Cardillo *et al*., *Science* **309**, 1239–1241 (2005).
5. W. J. Ripple *et al*., *Sci. Adv.* **1**, e1400103 (2015).
6. J. J. Midgley, D. Balfour, G. I. Kerley, *S. Afr. J. Sci.* **101**, 213–215 (2005).
7. G. P. Asner *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **106**, 4947–4952 (2009).
8. M. S. Waldram, W. J. Bond, W. D. Stock, *Ecosystems* **11**, 101–112 (2008).
9. C. T. Darimont *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **106**, 952–954 (2009).
10. V. Fugère, A. P. Hendry, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 10070–10075 (2018).
11. J. H. Daskin, R. M. Pringle, *Nature* **553**, 328–332 (2018).
12. I. Douglas-Hamilton, *Oryx* **21**, 11–24 (1987).
13. R. L. Beyers *et al*., *PLOS ONE* **6**, e27129 (2011).
14. V. Varun, T. Ewing, "Ivory's curse: The militarization and professionalization of poaching in Africa" (Born Free USA, 2014); www.bornfreeusa.org/campaigns/wildlife-trade/ivorys-curse-the-militarization-and-professionalization-of-poaching-in-africa-2014/.
15. M. E. Stalmans, T. J. Massad, M. J. S. Peel, C. E. Tarnita, R. M. Pringle, *PLOS ONE* **14**, e0212864 (2019).
16. J. O. J. Hatton, M. Couto, "Biodiversity and war: A case study of Mozambique" (Biodiversity Support Program, 2001).
17. A. M. Whitehouse, *S. Afr. J. Zool.* **257**, 249–254 (2002).
18. F. Tajima, *Genetics* **123**, 585–595 (1989).
19. J. H. Poole, in *The Ivory Trade and the Future of the African Elephant. Volume 2: Technical Reports* (Ivory Trade Review Group, 1989).
20. H. Jachmann, P. S. M. Berry, H. Imae, *Afr. J. Ecol.* **33**, 230–235 (1995).
21. G. Steenkamp, S. M. Ferreira, M. N. Bester, *J. S. Afr. Vet. Assoc.* **78**, 75–80 (2007).
22. D. Germain, A. Mehta, M. Beck, G. Sunder-Plassmann, in *Fabry Disease: Perspectives from 5 Years of FOS*, M. Atul, M. Beck, G. Sunder-Plassmann, Eds. (Oxford PharmaGenesis, 2006).
23. P. C. Lee, V. Fishlock, C. E. Webber, C. J. Moss, *Behav. Ecol. Sociobiol.* **70**, 337–345 (2016).
24. N. Chhabra, M. Goswami, A. Chhabra, *Med. Oral Patol. Oral Cir. Bucal* **19**, e112–e119 (2014).
25. M. Varela, M. J. Trujillo-Tiebas, P. García-Camba, *Eur. Arch. Paediatr. Dent.* **12**, 318–322 (2011).
26. N. Alachiotis, P. Pavlidis, *Commun. Biol.* **1**, 79 (2018).
27. K. E. Holsinger, B. S. Weir, *Nat. Rev. Genet.* **10**, 639–650 (2009).
28. M. Nei, J. C. Miller, *Genetics* **125**, 873–879 (1990).
29. H. L. Viswanathan *et al*., *J. Periodontol.* **74**, 1423–1431 (2003).
30. C. W. Gibson *et al*., *J. Biol. Chem.* **276**, 31871–31875 (2001).
31. G. M. Hobson *et al*., *Am. J. Med. Genet. A.* **149A**, 1698–1705 (2009).
32. X. Duan *et al*., *J. Dent. Res.* **98**, 870–878 (2019).
33. S. K. Prakash *et al*., *Hum. Mol. Genet.* **11**, 3237–3248 (2002).
34. S. Malik, S. Gupta, V. Wadhwan, G. P. Suhasini, *J. Oral Maxillofac. Pathol.* **19**, 110 (2015).
35. P. Arnold *et al*., *Cell Tissue Res.* **367**, 351–358 (2017).
36. S. R., J. Sarukhan, A. Whyte, R. Hassan, "Millennium Ecosystem Assessment: Ecosystems and human well-being" (Island Press, 2006).
37. C. J. Moss, *Elephant Memories: Thirteen Years in the Life of an Elephant Family* (William Morrow, 1988).
38. I. J. Whyte, A. Hall-Martin, *Pachyderm* **55**, 31–40 (2018).
39. E. Abe, *Pachyderm* **22**, 46–47 (1996).
40. J. S. Brashares *et al*., *Science* **345**, 376–378 (2014).
41. J. B. S. Haldane, *J. Genet.* **44**, 296–304 (1942).
42. K. H. Andersen, K. Brander, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 11657–11660 (2009).
43. M. Heino, B. Díaz Pauli, U. Dieckmann, *Annu. Rev. Ecol. Evol. Syst.* **46**, 461–480 (2015).
44. F. W. Allendorf, J. J. Hard, *Proc. Natl. Acad. Sci. U.S.A.* **106** (suppl. 1), 9987–9994 (2009).
45. J. S. Weir, *J. Zool.* **158**, 293–310 (1969).
46. W. H. Elder, *Zool. Africana* **158**, 293–310 (1969).
47. A. H. K. Beuchner, H. C. Dawkins, *Ecology* **42**, 752–766 (1961).
48. L. D. Wing, I. O. Buss, *Wildl. Monogr.* **19**, 3–92 (1970).
49. R. M. Pringle, *Ecology* **89**, 26–33 (2008).
50. T. C. Coverdale *et al*., *Ecology* **97**, 3219–3230 (2016).
51. E. P. Palkovacs, M. T. Kinnison, C. Correa, C. M. Dalton, A. P. Hendry, *Evol. Appl.* **5**, 183–191 (2012).
52. K. B. Oke *et al*., *Nat. Commun.* **11**, 4155 (2020).

ACKNOWLEDGMENTS

We thank A. Gadani, R. Branco, M. Stalmans, P. Tonecas, L. Van Wyk, M. Pingo, G. Carr, the staff of Gorongosa National Park, and the Republic of Mozambique. **Funding:** This research was supported by funding support to S.C.C.-S. from the University of California, Los Angeles. D.G. was supported by the National Geographic Society (GORONG1) and the Gorongosa Project. ElephantVoices (P.G. and J.P.) was supported by the National Geographic Society, the Crystal Springs Foundation, the Elephant Crisis Fund, the Elephant Sanctuary in Tennessee, the WildiZe Foundation, the Oakland Zoo, and the Gorongosa Project. R.A.L. was supported by the US National Science Foundation (IOS-1656642) and the National Geographic Society (WW-268C-17). R.M.P. was supported by the US National Science Foundation (IOS-1656527), the Cameron Schrier Foundation, the Greg Carr Foundation, and the High Meadows Environmental Institute of Princeton University. **Author contributions:** S.C.C.-S. conceived the project. S.C.C.-S., B.J.A., D.G., R.A.L., and R.M.P. participated in animal capture and genetic sample collection. J.P. collected all tusk configuration and offspring survey data. J.P. and P.G. created, maintained, and populated the elephant identification and sightings database. S.C.C.-S. and B.J.A. analyzed all survey, simulation, and genomic data. All authors participated in data interpretation and writing. **Competing interests:** Several authors have current or previous relationships with the Gorongosa Project (GP), a US 501(c)(3) nonprofit organization that supports conservation and community outreach in Gorongosa National Park, where the research was conducted. D.G. is employed by GP (2017 to present); P.G. and J.P. are directors of ElephantVoices, a US 501(c)(3) nonprofit organization that received a consultancy fee for research conducted in Gorongosa National Park; R.M.P. was an unpaid member of an advisory committee to GP from 2010 to 2020. **Data and materials availability:** Sequence data are available through the Short Read Archive within BioProject PRJNA 761769. Data and code associated with this study are stored and publicly available at Github (https://github.com/brian-arnold/AfricanElephantGenomics). The Gorongosa Elephant Who's Who & Whereabouts Database is archived at www.elephantvoices.org/gorongosaelephants/. This database is password protected as a precaution against poaching, but the data used in this study will be made available by P.G. and J.P. upon reasonable request in line with confidentiality agreements. All data from this database that are necessary for the analyses conducted in this study are openly available in the provided GitHub link.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abe7389
Materials and Methods
Supplementary Text
Figs. S1 and S2
Tables S1 to S8
References (*53*–*72*)
MDAR Reproducibility Checklist

11 September 2020; accepted 3 September 2021
10.1126/science.abe7389

MICROBIOLOGY

Rapid evolutionary turnover of mobile genetic elements drives bacterial resistance to phages

Fatima Aysha Hussain[1]†, Javier Dubert[1,2]†, Joseph Elsherbini[1], Mikayla Murphy[1], David VanInsberghe[1], Philip Arevalo[1], Kathryn Kauffman[1], Bruno Kotska Rodino-Janeiro[1,3], Hannah Gavin[1], Annika Gomez[1], Anna Lopatina[3], Frédérique Le Roux[4,5]*, Martin F. Polz[1,3]*

Although it is generally accepted that phages drive bacterial evolution, how these dynamics play out in the wild remains poorly understood. We found that susceptibility to viral killing in marine *Vibrio* is mediated by large and highly diverse mobile genetic elements. These phage defense elements display exceedingly fast evolutionary turnover, resulting in differential phage susceptibility among clonal bacterial strains while phage receptors remain invariant. Protection is cumulative, and a single bacterial genome can harbor 6 to 12 defense elements, accounting for more than 90% of the flexible genome among close relatives. The rapid turnover of these elements decouples phage resistance from other genomic features. Thus, resistance to phages in the wild follows evolutionary trajectories alternative to those predicted from laboratory-based evolutionary experiments.

Bacterial viruses (phages) are ubiquitous across Earth's biosphere and control microbial populations through predatory interactions (*1*–*3*). Because successful killing depends on molecular interaction with the host, phages display higher specificity than most other microbial predators. This is one reason for the renewed interest in the clinical use of phages as alternatives to antibiotics (*4*). Like antibiotics, however, phage killing exerts strong selection for resistance in bacterial hosts (*5*). Consequently, how bacteria naturally acquire resistance has important implications for understanding microbial community dynamics as well as the long-term success of phage therapy. Laboratory coevolution studies have consistently identified phage-receptor mutations as key drivers of resistance (*6*, *7*), with genetic analyses indicating secondary contributions by restriction-modification (RM) (*8*–*10*) and abortive infection (Abi) systems (*11*–*13*).

However, because phages target important surface structures, such as the lipopolysaccharide (LPS) and membrane proteins as receptors, it is questionable whether mutations in receptors represent primary adaptive strategies in complex microbial communities because such modifications frequently incur fitness costs (*6*). Indeed, many additional defense mechanisms have recently been discovered, including CRISPR systems (*14*) and several other innate immunity mechanisms, many of which remain to be mechanistically characterized (*15*–*19*). Genes encoding both receptors and defense systems have been shown to occur frequently in variable genomic islands (*6*, *20*–*22*) or to be associated with mobile genetic elements (*23*–*26*). For CRISPR-based adaptive immunity, which is present in more than 40% of bacteria (*27*), evolutionary processes have been documented on ecological time scales (*28*, *29*). By contrast, much less is known for innate immunity mechanisms, which defend most bacteria in the wild (*30*). Neither the predominant mechanisms of resistance to phages nor the dynamics of resistance gain and loss are well understood (*31*, *32*), limiting our knowledge of how phage predation structures diversity and drives the evolution of wild microbial populations.

Here, we combined population genomic and molecular genetic approaches to determine how phage resistance arises in natural microbial populations lacking adaptive immunity. We recently created a large, genomically resolved phage-host cross-infection network, using sympatric environmental isolates (*33*–*35*). The Nahant Collection was established in the context of a 93-consecutive-day coastal ocean time series (*36*) and comprises more than 1300 strains of marine *Vibrio* ranging in relatedness from near-clonal to species-level divergence (*34*, *37*). *Vibrio* hosts isolated on three different days (ordinal dates 222, 261, and 286) were used as "bait" to isolate 248 co-occurring lytic viruses by quantitative plaque assays, and each of 245 plaque-positive hosts was subsequently challenged with all phage isolates to establish an all-by-all cross-infection matrix (fig. S1A). Phages tend to be highly specific, as indicated by the general sparsity of the matrix. Additionally, even the most closely related hosts, differentiated by a few single-nucleotide polymorphisms (SNPs) across the core genome, are preyed upon by different phages (fig. S1B). Furthermore, this extends to broader host-range phages that, although capable of infecting multiple hosts, are typically limited to a single strain within a host clade (fig. S1C).

Because the observation that nearly clonal bacterial strains are subject to differential predation suggests extremely rapid evolution of phage resistance, we sought to identify the mechanisms in an exemplary set of 22 nearly clonal isolates of *Vibrio lentus* (Fig. 1A, left) that are differentially infected by two highly divergent groups of lytic siphovirus phages consisting of four and 18 isolates, respectively ("orange" and "purple" in Fig. 1A, top; see also fig. S2, A to D, and data S1). Testing all pairwise interactions of these phages with all potential hosts at varying multiplicity of infection (MOI) revealed two phenotypes: (i) host lysis and phage production and (ii) host lysis only at high phage concentrations without phage production (fig. S3). The latter effect, termed "lysis from without" (*38*), along with observed adsorption of representative phages to differing hosts (fig. S4), suggests that the phages can adsorb to a broader set of hosts than they can kill.

Supporting our hypothesis, we found that although the two phage types use different receptors, all bacterial genomes encode both sets of receptors. Molecular genetic analysis suggests that the "orange" phage receptor is the type II secretion system (T2SS) pseudopilus (GspH), whereas the "purple" phages have two receptors: LPS, an extracellular receptor, and the sodium transporter NqrC, a membrane receptor (*39*) (fig. S5 and tables S1 and S2). Comparative genomics showed that every identified gene encoding a phage receptor is identical across all 23 "orange," "purple," and outgroup strains. The noteworthy exception is the single outgroup strain (10N26154A12) that differs by one SNP in *gspN*: N201K, which prevents a single "orange" phage 1.127.O from attaching (Fig. 1A). Although this suggests that receptor mutations play a secondary role, the data overall support the idea that receptors are not primarily responsible for the observed phage specificity, leading us to explore potential intracellular mechanisms of host resistance by a combination of comparative genomics and molecular genetics.

Clustering the pangenome of the 22 near-clonal *Vibrio lentus* isolates using high-quality genomes, together with a slightly more divergent outgroup for additional genomic context (Fig. 1B, right), revealed the presence-absence patterns of large flexible genomic regions containing putative phage defense genes. Using a *k*-mer–based approach to conduct all-by-all

[1]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [2]Department of Microbiology and Parasitology, University of Santiago de Compostela, Santiago de Compostela, Spain. [3]Division of Microbial Ecology, Department of Microbiology and Ecosystem Science, Center for Microbiology and Environmental Systems Science, University of Vienna, Vienna, Austria. [4]Ifremer, Unité Physiologie Fonctionnelle des Organismes Marins, CS 10070, F-29280 Plouzané, France. [5]Sorbonne Universités, UPMC Paris 06, CNRS, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, F-29688 Roscoff Cedex, France.
*Corresponding author. Email: martin.f.polz@univie.ac.at (M.F.P.); fleroux@sb-roscoff.fr (F.L.R.)
†These authors contributed equally to this work.

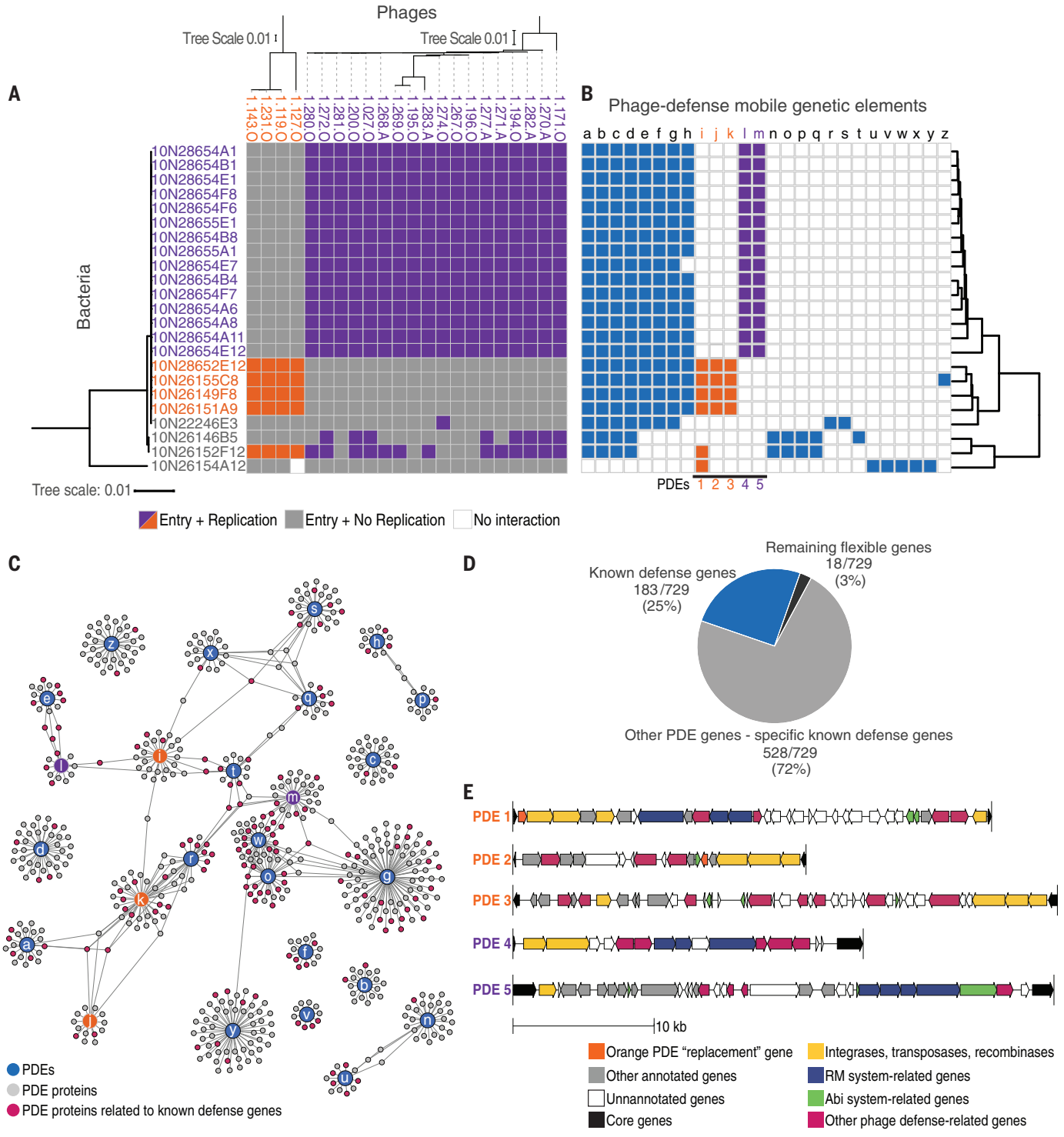


Fig. 1. Near-clonal strains of *Vibrio lentus* differ in sensitivity to phage predation and in their carriage of mobile genetic elements encoding for phage defense genes. (**A**) Phage host range matrix. Rows represent bacterial strains; columns represent phages. Left: Core genome phylogenetic tree of hosts. Top: Whole-genome tree of viruses. (**B**) Distribution of 26 PDEs in 23 hosts. PDEs responsible for defense of the "orange" and "purple" host groups are indicated in the same colors. Right: Hierarchical clustering of presence or absence of genomic regions in bacterial hosts corroborating "orange" and "purple" phage infection patterns. (**C**) Shared genes (>50% amino acid sequence identity) among the 26 PDEs, displayed as a network. PDEs are indicated in blue, and the orange and purple PDEs are color-coded accordingly. Gray nodes are genes; defense genes are indicated in magenta. (**D**) Fraction of the bacterial flexible genome attributed to phage defense. Among the 23 clones, an all-by-all genomic comparison shows that 97% of flexible regions greater than 5 kbp are putative PDEs. Only 25% of the total flexible genes match known defense genes; the majority are other PDE-associated genes (72%), many of which are unannotated. The remaining genes, not associated with PDEs, account for 3% of the flexible genome. (**E**) Gene diagrams of mobile genetic elements specific to the "orange" and "purple" host groups.

pairwise genome comparisons (data S2), we identified a total of 28 flexible genome regions (>5 kb in length), 26 of which show clear hallmarks of being mobile genetic elements (MGEs) and harboring at least one gene related to a known phage defense gene (data S3). These elements range in size from 8500 to 89,600 bp and contain genes annotated as recombinases/integrases and transposases (data S3), but the elements appear overall unrelated because only a few genes are shared (Fig. 1C and data S3). Except for two episomal elements, one of which is a prophage, all others are chromosomally integrated and their presence-absence patterns in different clonal genomes indicate that these regions are transferred as independent units (fig. S6). Of the chromosomally integrated elements, three are also putative prophages. We refer to these 26 MGEs as phage defense elements (PDEs) because no other function obviously relevant to the host cell can be identified. Their number ranges between six and 12 in each strain, and collectively the PDEs account for more than 95% of the flexible genomic regions (Fig. 1D and data S3). Even if only known defense genes, excluding the entire PDEs, are considered, phage defense genes account for >20% of the flexible gene content (Fig. 1D). We observed a similar range of 12 to 17% in diverse *Vibrio* species in our collection (fig. S7) even though these comprise higher genomic diversity than the clonal strains and have known flexible genome elements other than those involved in phage defense (*40*). If we consider that most of the remaining PDE genes may be involved in maintenance and spread, and that we did not detect any other functional annotations on the PDEs, we predict that a large, if not the largest, portion of the flexible genome might be involved in phage defense.

To investigate the impact of PDEs on phage defense, we looked more closely at a near-clonal subset of these hosts that have differential presence of PDEs matched by phage predation profiles. These hosts consist of a group of four strains and another group of 15 strains that are differentiated by only 14 SNPs across the core genome, yet are differentially infected by the two phage groups (Fig. 1A, fig. S8, and table S3). The flexible genome of these "orange" and "purple" hosts is differentiated by three (PDEs 1 to 3) and two (PDEs 4 and 5) large PDEs, respectively (Fig. 1E). PDE1, PDE4, and PDE5 have type 1 RM systems, and PDE2 and PDE3 have several putative defense genes. These include many genes related to abortive infection (Abi) systems, as well as genes with domains homologous to retrons, reverse transcriptases, and guanosine triphosphatases (GTPases) (*17*, *22*). A striking feature of the PDEs is that their insertion does not appear to disrupt host functions. For example, PDE1 inserts into "orange" hosts' 5′-deoxynucleotidase nucleosidase (*yfbr*) gene, thereby truncating it, but encodes its own divergent copy of the same gene. Similarly, although PDE2 disrupts a gene encoding a thiol peroxidase upon insertion, it encodes a copy in the middle of the element (Fig. 1E). These five PDEs are also found in more divergent *Vibrio* genomes, which indicates that they transfer via horizontal gene transfer (fig. S9), sometimes inserting into different genomic neighborhoods, although at conserved insertion motifs with the exception of a few SNPs (fig. S10).

We investigated the extent to which each of the PDEs contributes to the differential resistance observed between the "purple" and "orange" hosts. All three PDE-encoded RM systems (PDEs 1, 4, and 5; Fig. 1E) appear active according to methylome sequencing analysis (table S4) (*39*). To further characterize the contribution to resistance of other potential defense genes, we focused on the three PDEs specific to "orange" host strains as an exemplar. First, we found that all three PDEs are transcribed during an infection with a "purple" phage (fig. S11). Second, we systematically knocked out large regions of the three PDEs containing putative defense as well as unidentified candidate genes hypothesized to drive resistance of a representative "orange" host strain (10N26155C8) to "purple" phages (Fig. 2A) and then challenged the knockouts with all "purple" and "orange" phages (Fig. 2B and figs. S12 and S13). This genetic analysis shows that in addition to the RM-encoding PDE, both PDEs encoding alternative defense systems are also active and together contribute to phage resistance because deletion of all three PDEs is required for full sensitivity.

The complexity of the interaction among all PDEs leading to full resistance is illustrated by deletion experiments covering all possible combinations. Knocking out RM-containing PDE1 alone increased the "lysis from without" phenotype by a factor of 10 (from 10^{-1} to 10^{-2} phage dilution) but did not yield viable phage progeny when assayed using a restreaking test (Fig. 2C). This phenotype led us to hypothesize that there may be a multilevel defense structure involving the remaining two PDEs. Knocking out PDE2 and PDE3 independently resulted in no change from the wild-type phenotype. However, knocking out PDE2 and PDE3 together allowed for both killing and propagation of the phage at a wider range of high MOIs ($\leq 10^{-5}$ phage dilution). This phenotype is consistent with RM escape mutants (*39*) (fig. S14). Knocking out PDE1 and PDE2 together resulted in the same factor-of-10 increase in killing observed when deleting PDE1 alone, whereas knocking out PDE1 and PDE3 together yielded cell death at much lower MOIs (10^{-5} phage dilution); however, in neither case were viable phages propagated (Fig. 2C). These data suggest that both PDEs afford protection but that PDE3 provides stronger resistance than PDE2. Finally, knocking out all three PDEs simultaneously resulted in the "orange" strain becoming just as susceptible to the "purple" phages as are "purple" wild-type host strains. Therefore, we conclude that all three elements are needed for full, wild-type–level defense.

Further genetic characterization of defense genes within the PDEs underscores the diversity of undiscovered or poorly characterized genes with anti-phage function. We took advantage of our double PDE mutants to scrutinize the remaining active PDE for phage defense. Although we were only able to find a single gene (the restriction enzyme) necessary for defense on PDE1 (Fig. 2A and fig. S15), three unannotated genes and two putative GTPases within a module on the 5′ end of PDE2 appear involved in defense (Fig. 2A and fig. S16). For PDE3, we inactivated 11 different genes as well as two regions associated with the two retrons present; however, none showed a loss-of-resistance phenotype. We therefore analyzed the host transcriptome during infection with a "purple" phage and identified 14 highly expressed and/or cotranscribed regions (fig. S17A). Using the triple-knockout mutant ΔΔΔPDE123, we complemented eight of these regions with their native promoter and observed restoration of phage defense by region R3 containing a gene with DUF2971 (Fig. 2A and fig. S17, B and C). Hence, two of the three PDEs carry uncharacterized phage defense genes.

The invariance of receptor genes across the two near-clonal host groups challenges the notion that receptor variation is primarily responsible for resistance (*6*, *20*). Thus, we asked to what extent these receptors are variable across more diverse populations. Populations are defined here as gene flow clusters that also represent ecological units (*37*, *41*). The recognition of population boundaries is key for interpretation of gene or allele frequency in the context of differential selective forces. Surprisingly, looking across 107 *Vibrio* isolates spanning 10 populations, all putative receptors are highly monomorphic at the population level, indicating that they are under purifying selection. The two genes identified as receptors in *V. lentus* are identical, or nearly so, at the amino acid sequence level across populations and thus are well below the average diversity of core genes (fig. S18A). This is corroborated by phylogenetic trees based on nucleotide sequences showing that all members of each population carry the same or highly similar gene variants, a pattern consistent with the occurrence of recent gene-specific selective sweeps. A notable exception is *gspH*, which is more diverse in two of the populations (fig. S18B). Finally, the LPS,

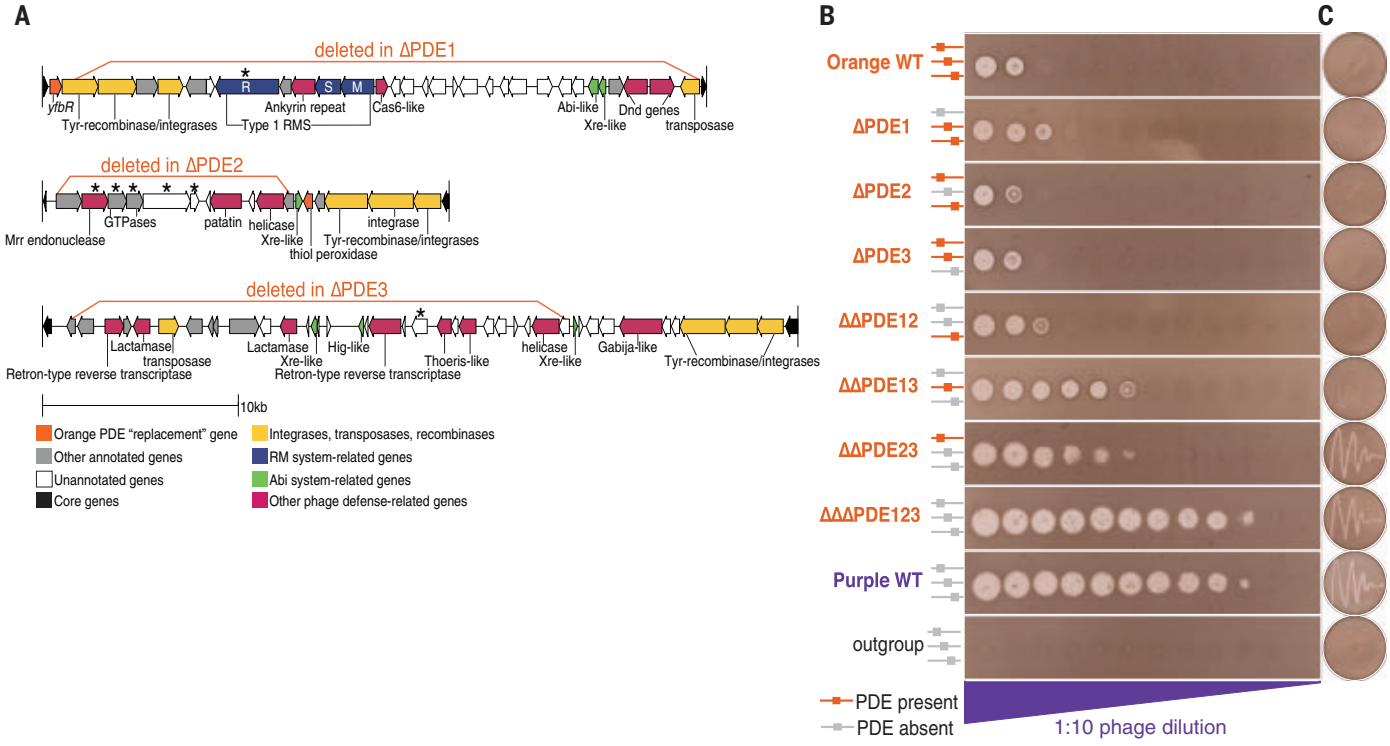


Fig. 2. Changes in susceptibility to phage killing observed for phage defense element (PDE) deletions. (**A**) Detailed gene diagrams indicate markerless deletions of defense gene–containing portions of the different "orange" PDEs as well as genes determined to be responsible for defense phenotypes (marked with an asterisk; see figs. S15 to S17). (**B**) Lawns of bacterial hosts with drop spots of a 1:10 dilution series of "purple" phage (image shows representative "purple" phage 1.281.O; plating completed in triplicate for each phage, fig. S12). Cartoons on the left depict presence or absence of different PDEs in each strain. From top to bottom: "orange" wild-type host (10N26155C8), ΔPDE1, ΔPDE2, ΔPDE3, ΔΔPDE12, ΔΔPDE13, ΔΔPDE23, ΔΔΔPDE123, "purple" wild-type host (10N28654F7, positive control), outgroup (10N26149C11, negative control). (**C**) Restreak test for propagation of phage progeny from drop spot clearings. Only infections of ΔΔPDE23, ΔΔΔPDE123, and "purple" wild-type hosts produce viable phages, indicated by clearing on the restreak plates.

which frequently serves as the primary receptor for many phages, also appears similar at the population level because the genes responsible for its synthesis display population-specific presence/absence patterns suggesting that the synthesis pathway is conserved (fig. S18C).

Thus, although putative receptors can reside in variable regions when more divergent genomes are compared (*20*), their evolution appears to be constrained when population structure is considered. This constraint may arise because in wild populations, these surface structures are optimized for ecological interactions, and its existence indicates a key difference from the lab, where, in rich media, receptor mutations frequently arise in phage-host cocultures (*42*). This observed invariance suggests that other selective forces that compete with phage resistance play an important role in receptor evolution in the wild. This interpretation is consistent with predictions that intracellular defenses should be important under such conditions (*43*, *44*), as well as with phage-host evolution in nature differing from the lab (*45*). It is possible that receptor-mediated defenses may be advantageous only under extreme predation regimes, or under regimes of low effective diversity, such as a clonal infection or laboratory culture.

If defense is largely determined by differential PDE carriage at the clonal level, this suggests that the rate of turnover is surprisingly fast and confirms theoretical considerations that resistance genes should be mobile because the cost of resistance limits their utility under changing predation pressures (*25*). Genomic comparisons of the 23 strains illustrates this high turnover because genomes with zero SNPs can differ by at least one putative PDE, and the "purple" and "orange" host groups carry a minimum difference of five PDEs while being separated by only 14 SNPs (Fig. 1 and fig. S8). Further, across the time series of a mere 90 days, we observed a differential shift in relative PDE abundance. We designed droplet digital polymerase chain reaction (ddPCR) assays to span the insertion sites of three of the PDEs to constrain the detection of the resistance elements to closely related genetic backgrounds (fig. S18). PDE3 and PDE4 (representing "orange" and "purple" genetic backgrounds, respectively) were near perfectly correlated across the entire time series, whereas PDE2 gained in importance only toward the end of the time series. These dynamics suggest rapid evolution at the population level, either from acquisition of PDE2 or a shift in abundance of host clones with differential PDE carriage. Both scenarios reflect a rapidly shifting selective landscape of host-phage interactions.

PDEs structuring host defense among near-clonal isolates is likely general beyond the detailed description provided here. Within other *Vibrio* species in our collection, clonal isolates also differ in their phage predation profiles (fig. S1), and, even among bacterial pathogens that—in contrast to *Vibrio* species—follow a primarily nonrecombinogenic mode of evolution (*37*, *46*–*48*), we have observed similarly rapid turnover of putative phage defense elements, several of which are prophages themselves. When analyzing publicly available genomes with identical ribosomal protein gene sequences from *Listeria* (fig. S20 and data S4), *Salmonella* (fig. S21 and data S5), and *Clostridium* (fig. S22 and data S6), we observe a surprisingly similar dominance (~90%) of the flexible genome associated with phage defense regions across these diverse species.

We have shown that innate immunity mechanisms can be rapidly turned over on the population level, with three major implications. First, because coexisting diversity in many natural bacterial populations is high (*49*), resistance at the clonal level effectively lowers prey concentration and thus the frequency of phage-host encounter. Second, encoding resistance on highly mobile elements can make phage resistance a fast-changing phenotypic trait independent of physiological or metabolic traits encoded by the core genome. This allows the core genome to be maintained over the long term even in the face of phage predation. In other words, although phages may exert negative frequency-dependent selection on their hosts, this selection is largely acting at the level of MGEs. Finally, rapid acquisition of resistance may also have to be considered for repeated or longer-term use of phages in therapy because resistance may be easily acquired and quickly spread through bacterial populations, just as the connection to MGEs (i.e., primarily plasmids) has led to an unanticipated rise in antibiotic resistance. It is therefore necessary to determine the mechanisms of transfer and the life history strategies of PDEs. Together, our findings suggest that phage resistance is an important, if not the most important, selective force determining clonal bacterial diversity, with phage defense elements potentially explaining a very large portion of the previously enigmatic bacterial flexible genome.

REFERENCES AND NOTES

1. K. E. Wommack, R. R. Colwell, *Microbiol. Mol. Biol. Rev.* **64**, 69–114 (2000).
2. M. Breitbart, F. Rohwer, *Trends Microbiol.* **13**, 278–284 (2005).
3. H. G. Hampton, B. N. J. Watson, P. C. Fineran, *Nature* **577**, 327–336 (2020).
4. K. E. Kortright, B. K. Chan, J. L. Koff, P. E. Turner, *Cell Host Microbe* **25**, 219–232 (2019).
5. S. J. Labrie, J. E. Samson, S. Moineau, *Nat. Rev. Microbiol.* **8**, 317–327 (2010).
6. S. Avrani, O. Wurtzel, I. Sharon, R. Sorek, D. Lindell, *Nature* **474**, 604–608 (2011).
7. J. R. Meyer *et al*., *Science* **335**, 428–432 (2012).
8. W. Arber, *Annu. Rev. Microbiol.* **19**, 365–378 (1965).
9. G. G. Wilson, N. E. Murray, *Annu. Rev. Genet.* **25**, 585–627 (1991).
10. G. Bertani, J. J. Weigle, *J. Bacteriol.* **65**, 113–121 (1953).
11. I. J. Molineux, *New Biol.* **3**, 230–236 (1991).
12. L. Snyder, *Mol. Microbiol.* **15**, 415–420 (1995).
13. A. Lopatina, N. Tal, R. Sorek, *Annu. Rev. Virol.* **7**, 371–384 (2020).
14. R. Barrangou *et al*., *Science* **315**, 1709–1712 (2007).
15. S. Doron *et al*., *Science* **359**, eaar4120 (2018).
16. A. Millman *et al*., *Cell* **183**, 1551–1561.e12 (2020).
17. L. Gao *et al*., *Science* **369**, 1077–1084 (2020).
18. R. P. Novick, G. E. Christie, J. R. Penadés, *Nat. Rev. Microbiol.* **8**, 541–551 (2010).
19. A. A. Govande, B. Duncan-Lowey, J. B. Eaglesham, A. T. Whiteley, P. J. Kranzusch, *Cell Rep.* **35**, 109206 (2021).
20. F. Rodriguez-Valera *et al*., *Nat. Rev. Microbiol.* **7**, 828–836 (2009).
21. K. S. Makarova, Y. I. Wolf, S. Snir, E. V. Koonin, *J. Bacteriol.* **193**, 6039–6056 (2011).
22. A. Bernheim, R. Sorek, *Nat. Rev. Microbiol.* **18**, 113–119 (2020).
23. N. D. McDonald, A. Regmi, D. P. Morreale, J. D. Borowski, E. F. Boyd, *BMC Genomics* **20**, 105 (2019).
24. A. C. McKitterick, K. D. Seed, *Nat. Commun.* **9**, 2348 (2018).
25. E. V. Koonin, K. S. Makarova, Y. I. Wolf, M. Krupovic, *Nat. Rev. Genet.* **21**, 119–131 (2020).
26. K. N. LeGault *et al*., *Science* **373**, eabg2166 (2021).
27. K. S. Makarova *et al*., *Nat. Rev. Microbiol.* **9**, 467–477 (2011).
28. A. F. Andersson, J. F. Banfield, *Science* **320**, 1047–1050 (2008).
29. E. Laanto, V. Hoikkala, J. Ravantti, L.-R. Sundberg, *Nat. Commun.* **8**, 111 (2017).
30. T. Dimitriu, M. D. Szczelkun, E. R. Westra, *Curr. Biol.* **30**, R1189–R1202 (2020).
31. A. Chevallereau, B. J. Pons, S. van Houte, E. R. Westra, *Nat. Rev. Microbiol.* 10.1038/s41579-021-00602-y (2021).
32. S. van Houte, A. Buckling, E. R. Westra, *Microbiol. Mol. Biol. Rev.* **80**, 745–763 (2016).
33. K. M. Kauffman *et al*., *Sci. Data* **5**, 180114 (2018).
34. K. M. Kauffman *et al*., *Nature* **554**, 118–122 (2018).
35. K. M. Kauffman *et al*., BioRxiv 449121 [preprint]. 27 June 2021.
36. A. M. Martin-Platero *et al*., *Nat. Commun.* **9**, 266 (2018).
37. P. Arevalo, D. VanInsberghe, J. Elsherbini, J. Gore, M. F. Polz, *Cell* **178**, 820–834.e14 (2019).
38. M. Delbrück, *J. Gen. Physiol.* **23**, 643–660 (1940).
39. See supplementary materials.
40. O. X. Cordero, M. F. Polz, *Nat. Rev. Microbiol.* **12**, 263–273 (2014).
41. B. J. Shapiro *et al*., *Science* **336**, 48–51 (2012).
42. E. R. Westra *et al*., *Curr. Biol.* **25**, 1043–1049 (2015).
43. S. Zborowsky, D. Lindell, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 16899–16908 (2019).
44. E. O. Alseth *et al*., *Nature* **574**, 549–552 (2019).
45. C. A. Hernandez, B. Koskella, *Evolution* **73**, 2461–2475 (2019).
46. K. E. Dingle *et al*., *PLOS ONE* **6**, e19993 (2011).
47. B. J. Shapiro, *Curr. Opin. Microbiol.* **31**, 116–123 (2016).
48. Y. Yin *et al*., *Microbiol. Res.* **175**, 84–92 (2015).
49. J. R. Thompson *et al*., *Science* **307**, 1311–1313 (2005).
50. Code is provided in data S2 and online (DOI: 10.5281/zenodo.5228182).
51. Supporting data for phylogenetic trees are provided in data S9 and online (DOI: 10.5061/dryad.sj3tx965k).

ACKNOWLEDGMENTS

We thank M. Cutler for experimental support; D. Newman, K. Costa, and H. Wildschutte for advice on mutagenesis experiments; O. X. Cordero, S. Kearney, and A. F. Takemura for valuable suggestions throughout; S. W. Chisholm, L. Kelly, and S. Pollak for thoughtful comments on the manuscript; and D. Kwon, X. Yu, O. Abudayyeh, and J. Gootenberg for use of ddPCR equipment and/or lab space during the SARS-CoV-2 pandemic. **Funding:** Simons Foundation (Life Sciences Project Award-572792) (M.F.P.); National Science Foundation Division of Ocean Sciences (OCE-1435868) (M.F.P.); MIT J-WAFS seed grant (M.F.P.); NSF GRFP (F.A.H.); MIT Martin Society of Fellows for Sustainability (F.A.H.); Xunta de Galicia Postdoctoral Fellowship (ED481B 2016/032) (J.D.); Agence Nationale de la Recherche (ANR-16-CE32-0008-01) (F.L.R.); European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 884988, Advanced ERC Dynamic) (F.L.R.). **Author contributions:** Conceptualization: F.A.H., M.F.P. Data curation: F.A.H., with contributions from J.E., J.D., M.M., P.A., D.V., B.K.R.-J., K.K. Methodology: F.A.H., J.D., M.M., J.E., F.L.R. Investigation: F.A.H., J.D., F.L.R., with contributions from J.E., M.M., P.A., D.V., K.K., B.K.R.-J., H.G. Validation: F.A.H., J.D., F.L.R., with contributions from J.E., H.G., A.L., A.G. Visualization: F.A.H., J.E., with contributions from D.V., P.A., K.K. Software: F.A.H., J.E., M.M., with contributions from D.V., P.A. Funding acquisition: M.F.P., F.L.R., F.A.H., J.D. Project administration and supervision: M.F.P., F.L.R. Writing–original draft: F.A.H., M.F.P. Writing–review and editing: F.A.H., M.F.P., with contributions from J.D., F.L.R. **Competing interests:** Authors declare that they have no competing interests. **Data and materials availability:** New genomes used in this work have been deposited under the NCBI BioProject with accession number PRJNA328102. Main code is provided in data S2 and online (*50*). Supporting data are provided in data S9 and online (*51*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abb1083
Materials and Methods
Figs. S1 to S22
Tables S1 to S4
References (*52*–*97*)
Data S1 to S9

1 February 2020; resubmitted 10 May 2021
Accepted 1 September 2021
10.1126/science.abb1083

AAAS.ORG/COMMUNITY

AAAS' Member Community is a one-stop destination for scientists and STEM enthusiasts alike. It's "Where Science Gets Social": a community where facts matter, ideas are big and there's always a reason to come hang out, share, discuss and explore.

Member COMMUNITY AAAS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

Who's the top employer for 2020?

Science Careers' annual survey reveals the top companies in biotech & pharma voted on by *Science* readers.

Read the article and employer profiles at sciencecareers.org/topemployers

Science 2020
TOP EMPLOYER

EMPIRE INNOVATION PROGRAM FACULTY POSITIONS IN VISION RESEARCH

The Department of Ophthalmology and Visual Sciences (OVS) and the interdepartmental Center for Vision Research (CVR) at SUNY Upstate Medical University are seeking candidates to fill **faculty positions** supported by the prestigious State University of New York Empire Innovation Program (EIP). The EIP supports the recruitment of faculty with strong track records of research accomplishments and that are recognized leaders in their fields with international reputation for outstanding research. The Center for Vision Research has been recognized for its strength in eye research and was competitively awarded two EIP positions to be filled immediately.

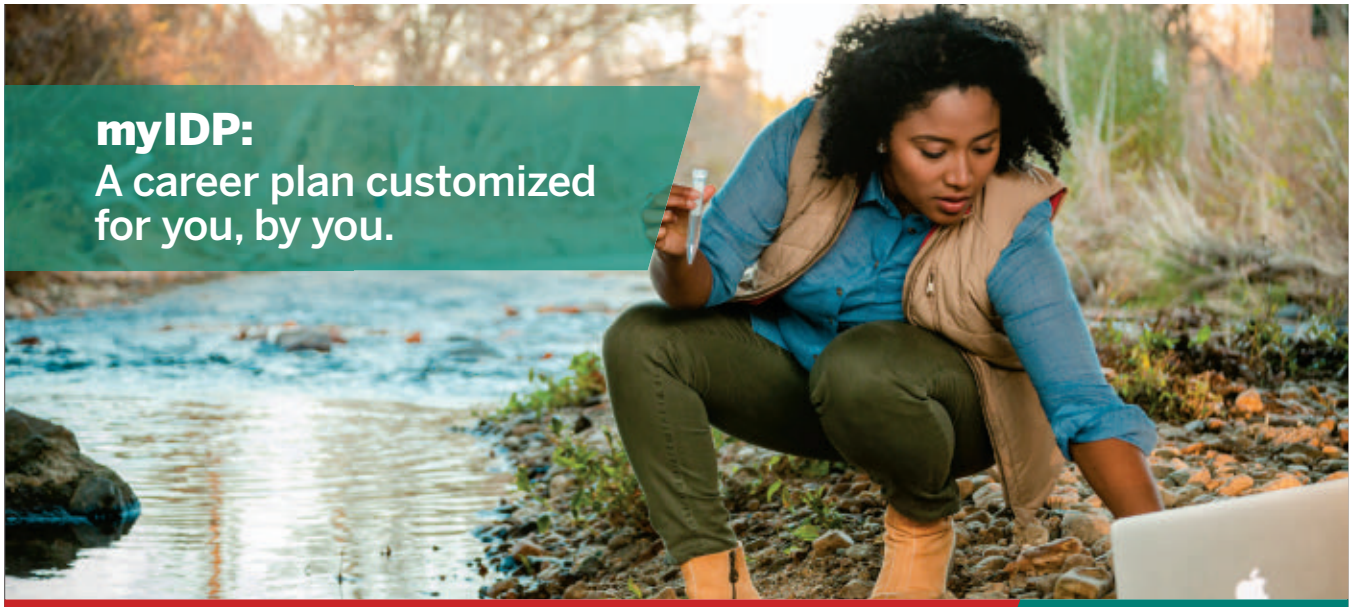
We are seeking researchers with expertise in imaging of the living eye and in developing artificial intelligence approaches, including machine learning or deep learning, applied to eye research and/or diagnostics. Candidates for one position should have an outstanding research and funding track record and demonstrated international reputation. Candidates for the second position may be earlier in their career path with demonstrated potential for a vigorous independent career. Candidates should have MD, PhD or MD/PhD degrees or equivalents.

Upstate Medical University strives for an environment that encourages varied perspectives from faculty with diverse life experiences. A respect for diversity is one of our core values. We actively seek applications from women and members of underrepresented groups to contribute to the diversity in support of our teaching, research and clinical missions. The research faculty in OVS is 40% women, 20% LatinX, 10% LGBTQ; we are committed to enhancing that diversity.

In 2021, Upstate Medical University ranked #35 on the Forbes list of America's best large employers. By earning this award, Upstate joins the ranks only two-dozen universities in the nation. Upstate Medical University is located in Central New York (CNY), which is well known for its natural beauty. CNY boasts world-class outdoor activities year around. Syracuse offers outstanding quality of life whether urban or in surrounding suburbs, with excellent schools and low cost of living. The city has a walkable downtown entertainment district and is close to New York City, Toronto, Montreal, or Boston making urban weekend getaways easy.

Please send a letter of interest, CV, statement of research interests, and a list of references as a single PDF document to:
Peter D. Calvert, PhD
Professor of Ophthalmology and Visual Sciences
Chair of EIP Search Committee
email: **OVS_EIPSearch@upstate.edu**
https://www.upstate.edu/cvr

ScienceCareers online @sciencecareers.org



For your career in science, there's only one Science

Features in myIDP include:

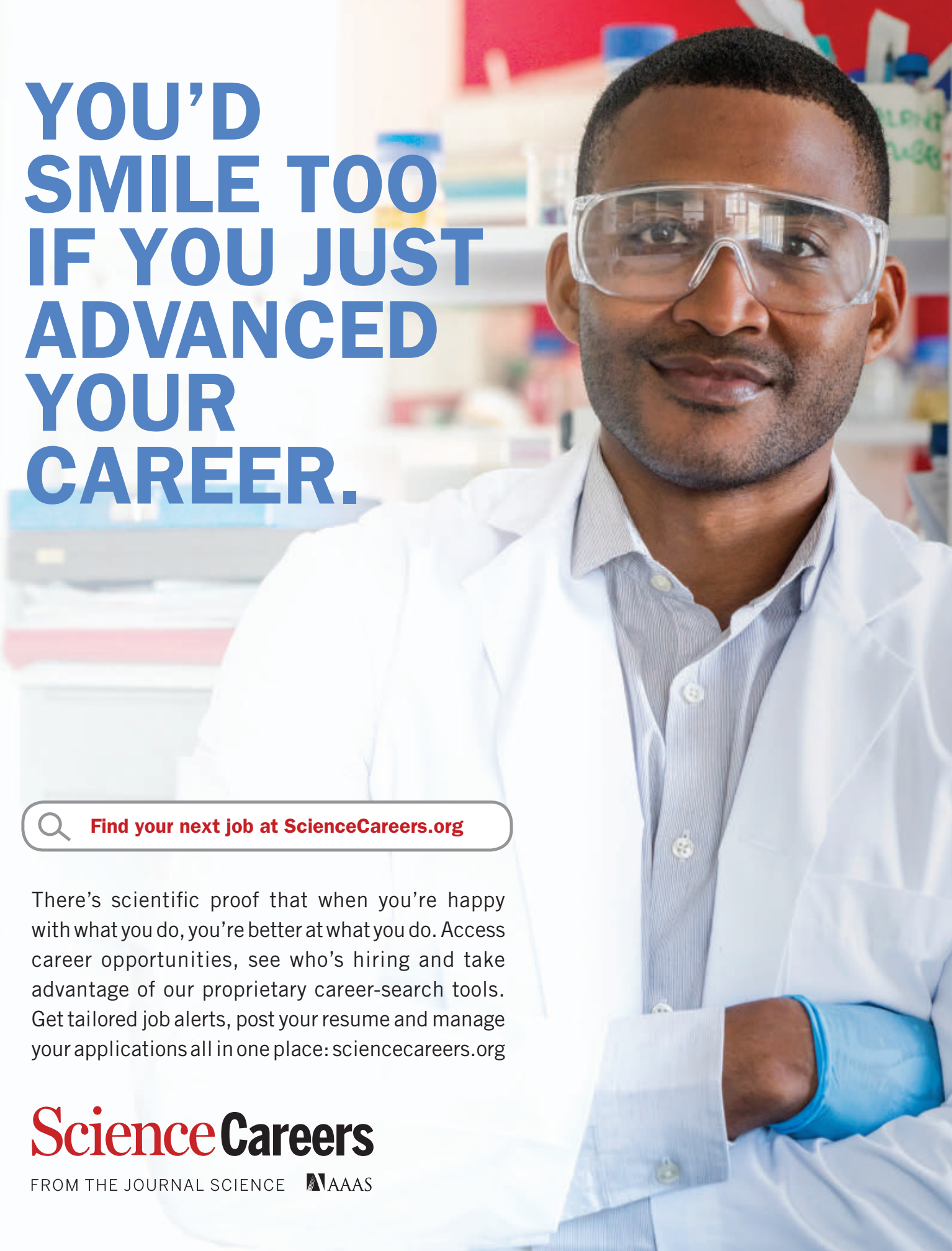
- Exercises to help you examine your skills, interests, and values.
- A list of 20 scientific career paths with a prediction of which ones best fit your skills and interests.
- A tool for setting strategic goals for the coming year, with optional reminders to keep you on track.
- Articles and resources 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 AAAS In partnership with:

FASEB Federation of American Societies for Experimental Biology
University of Massachusetts Medical School
UCSF University of California San Francisco
UIC
BURROUGHS WELLCOME FUND



YOU'D SMILE TOO IF YOU JUST ADVANCED YOUR CAREER.

Find your next job at ScienceCareers.org

There's scientific proof that when you're happy with what you do, you're better at what you do. Access career opportunities, see who's hiring and take advantage of our proprietary career-search tools. Get tailored job alerts, post your resume and manage your applications all in one place: sciencecareers.org

ScienceCareers

FROM THE JOURNAL SCIENCE AAAS

By **Mingde Zheng**

When to say 'no'

I was sitting at my desk when another grad student in my lab approached me. "Can you help?" he pleaded. His experiment wasn't working and he desperately needed help troubleshooting. I was a fifth year Ph.D. student at the time, and I took pride in being the senior member of the lab, whom everyone looked up to. But that also meant I was the one everyone turned to for help—which ate away hours, days, and sometimes weeks that I could have spent on my own research. As I had many times before, I caved and said yes. It was a habit that took years to break.

There were many reasons I had a hard time saying no to such pleas. I was new to the United States, having moved from China for graduate school. I found it difficult to make new friends and discover afterwork activities I enjoyed. So I spent a huge chunk of my time in the lab, and my lab-mates served as my primary source of social connection. I feared that if I brushed them off, I'd lose their favor.

It was even harder to say no to my Ph.D. adviser, who had the power to withhold funding and recommendation letters. When he asked me to train new researchers in the lab, I gladly took on the responsibility and led them through a full day of lab orientation and training. When a machine broke down and he requested that I be the technician, I did my best to fix the problem. And when he asked me to set up and conduct elaborate demonstrations for a class he was teaching, I put aside my experiments and immediately got to work on his request.

"It was wonderfully addictive to be needed ... but the extra responsibilities came at a cost."

It was wonderfully addictive to be needed. My own experiments were technically challenging and never went the way I planned. I got a much-needed confidence boost when I could solve a problem or do a service for someone else.

But the extra responsibilities came at a cost. I had to work extra hours to catch up with my own work, and I often made sacrifices to my personal life. One holiday break, I spent hours in the lab fixing a piece of equipment that I rarely used, when I could have been spending time with my family, who had traveled from China to see me.

After I graduated and started a job in industry, I continued to carry the same mentality with me, and it continued to cost me. When a colleague asked for help with his short-staffed project, I dropped everything to help, hoping he'd become a friend at work and that he'd put in a good word about my contributions to my manager. But he moved to another division of the company a few weeks later, and I never saw him again. I missed my own project deadline because of the extra work, and I gained nothing in return.

It wasn't until my wife gave birth to our first child that I realized how thin I had stretched myself—and how misguided my priorities were. While she lay in a hospital bed in the early stages of labor, I sat nearby hunched over my laptop finishing up a work report. Hours later, after hearing my daughter's first cry and watching her tiny fingers grab tightly onto mine, it dawned on me: I should have been fully present during my daughter's birth. I was clearly spending too much time working if my job intruded into one of life's most precious moments.

From then on, I resolved to carve out more time for my family by paring down my work tasks and carefully considering each request for help. I still enjoyed collaborating with others, but I prioritized mutually beneficial tasks or those my manager asked me to take on, rather than accepting everything that came my way.

I noticed many benefits, and few downsides. My daily work agenda became less crammed, I no longer worked overtime—affording me more family time—and I was able to improve my work performance by focusing more on core assignments. I was also pleased to discover that "Sorry, I'd love to help but I have a deadline coming up" is an acceptable response to a request for assistance. Most colleagues seem to understand.

It's hard to say "no" to those you work with. But I've learned that sometimes that's the best course of action to avoid an excessive workload and lead a freer and happier life. ■

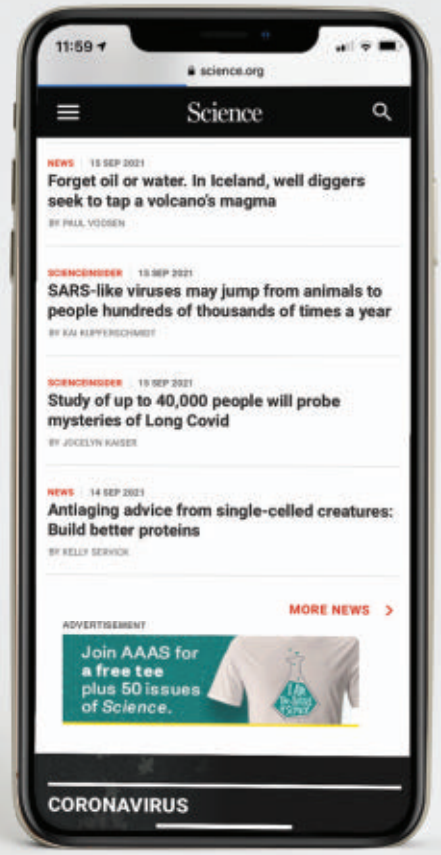
Mingde Zheng is a research scientist at Nokia Bell Labs. Do you have an interesting career story to share? Send it to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

NEWS FROM Science

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

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



Science.org/news

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