New funds, new scrutiny for Africa's Great Green Wall p. 666

Do microplastics affect human health? p. 672

Imaging Earth's crust with whale calls p. 731

Science

$15
12 FEBRUARY 2021
sciencemag.org

AAAS

DIVERSITY AND POLICING

Big data reveal threats to minorities policed by white and male officers pp. 677 & 696

Fundación BBVA

14th Edition

BBVA Foundation Frontiers of Knowledge Awards

With the collaboration of

CSIC
Consejo Superior de Investigaciones Científicas

The **BBVA Foundation Frontiers of Knowledge Awards** recognize and reward world-class research and artistic creation, prizing contributions of singular impact for their originality and significance. The name of the scheme is intended to denote not only research work that substantially enlarges the scope of our current knowledge – pushing forward the frontiers of the known world – but also the meeting and overlap of different disciplinary areas and the emergence of new fields.

The Frontiers of Knowledge Awards honor fundamental disciplinary or interdisciplinary advances across a broad expanse of the knowledge map of the 21st century.

The BBVA Foundation is assisted in the award process by the **Spanish National Research Council (CSIC).**

Categories

1/ **Basic Sciences (Physics, Chemistry, Mathematics)**
2/ **Biology and Biomedicine**
3/ **Information and Communication Technologies**
4/ **Ecology and Conservation Biology**
5/ **Climate Change**
6/ **Economics, Finance and Management**
7/ **Humanities and Social Sciences**
8/ **Music and Opera**

In Humanities and Social Sciences, the award will alternate annually between these two disciplinary domains, with this fourteenth edition dedicated to the Social Sciences.

Nomination

Nominations are invited from scientific or artistic societies and organizations, public or private R&D centers, university and hospital departments, schools of music, orchestras, and organizations working on or around the issue of climate change, as well as other institutions specified in the call conditions.

Entry submission

The nomination period concludes at **23:00 GMT on June 30, 2021.**

www.frontiersofknowledgeawards-fbbva.es

Fundación BBVA
Plaza de San Nicolás, 4 · 48005 Bilbao
Paseo de Recoletos, 10 · 28001 Madrid
awards-info@fbbva.es

CONTENTS

12 FEBRUARY 2021 • VOLUME 371 • ISSUE 6530

663

NEWS

IN BRIEF

654 News at a glance

IN DEPTH

657 United States rushes to fill void in viral sequencing
Some experts say Biden administration plan is too modest to track fast-evolving variants as they arise and spread *By M. Wadman*

658 Genes for life on land evolved earlier in fish
Hidden genetic pathways behind land-water transition are found in living fish *By E. Pennisi*

660 Pandemic hit academic mothers hard, data show
New evidence highlights need for policy changes to aid female scientists *By K. Langin*

661 Study: Police diversity matters
Landmark analysis of 7000 Chicago police shows nonwhite and female officers make fewer stops *By D. Starr*
PERSPECTIVE p. 677; RESEARCH ARTICLE p. 696

662 Illicit centipede spurs taxonomy journal debate
Editors differ about whether they are obliged to ensure specimens are legally acquired
By Y. H. Law

FEATURES

663 A revealing flaw
A rare disease that cripples a key cellular organelle holds clues to treating more common conditions *By M. Leslie*

666 Taking root
A new influx of funds could help realize Africa's Great Green Wall. But will the massive tree-planting effort learn from past failures? *By R. Cernansky*
PODCAST

INSIGHTS

POLICY FORUM

668 Strengthen scientific integrity under the Biden administration
Accountability must extend to highest levels of leadership
By J. M. Carter et al.

PERSPECTIVES

672 Microplastics and human health
Knowledge gaps should be addressed to ascertain the health risks of microplastics
By A. D. Vethaak and J. Legler

674 Speedy galaxy evolution
Mature features are detected in an early galaxy *By J. Wardlow*
REPORT p. 713

676 Different origins for similar brain circuits
Gene expression profiles indicate separate evolution of cortical circuits in birds and mammals *By M. A. Tosches*
RESEARCH ARTICLE p. 695

677 Asking the right questions about race and policing
When is policing the right tool, and when is it the problem? *By P. A. Goff*
NEWS STORY p. 661; RESEARCH ARTICLE p. 696

678 Tidying up the mess
Cadmium doping improves the thermoelectric properties of $AgSbTe_2$ through ordering effects *By Y. Liu. and M. Ibáñez*
REPORT p. 722

680 Tracking the UK SARS-CoV-2 outbreak
The UK epidemic was seeded by more than 1000 imported viruses from Europe *By M. I. Nelson*
RESEARCH ARTICLE p. 708

681 Single-domain antibodies make a difference
A double hit with one antibody construct may avoid viral escape
By X. Saelens and B. Schepens
RESEARCH ARTICLE p. 691

BOOKS *ET AL.*

683 Dream interpretation meets modern science
Nighttime narratives could be critical to consolidating memories, argue two sleep researchers *By M. Frazer and G. Poe*

684 Bill Gates's guide to avoiding climate catastrophe
The Microsoft mogul offers a pragmatic approach to achieving greenhouse gas targets *By M. R. Aczel*

LETTERS

685 Reversing extinction in China's Père David's deer
By Y. Zhang et al.

685 A marine biodiversity plan for China and beyond
By W. Zhou et al.

ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*

686 Worst climate outcomes are still possible
By P. B. Duffy and M. L. Duffy

686 Technical Comment abstracts

RESEARCH

IN BRIEF

687 From *Science* and other journals

REVIEW

690 Physiology
Communicating clocks shape circadian homeostasis *K. B. Koronowski and P. Sassone-Corsi*
REVIEW SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABD0951

RESEARCH ARTICLES

691 Coronavirus
Structure-guided multivalent nanobodies block SARS-CoV-2 infection and suppress mutational escape *P.-A. Koenig* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABE6230
PERSPECTIVE p. 681

692 Molecular motor
Germanium nanospheres for ultraresolution picotensiometry of kinesin motors *S. Sudhakar* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABD9944

693 Heart valve disease
Network-based screen in iPSC-derived cells reveals therapeutic candidate for heart valve disease *C. V. Theodoris* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABD0724

694 Human evolution
Reintroduction of the archaic variant of *NOVA1* in cortical organoids alters neurodevelopment *C. A. Trujillo* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.AAX2537

695 Neuroscience
Cellular transcriptomics reveals evolutionary identities of songbird vocal circuits *B. M. Colquitt* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABD9704
PERSPECTIVE p. 676

696 Criminal justice
The role of officer race and gender in police-civilian interactions in Chicago *B. A. Ba* et al.
NEWS STORY p. 661; PERSPECTIVE p. 677

702 Nucleotide synthesis
Catalytic asymmetric and stereodivergent oligonucleotide synthesis *A. L. Featherston* et al.

708 Coronavirus
Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK *L. du Plessis* et al.
PERSPECTIVE p. 680

REPORTS

713 Galaxies
A massive stellar bulge in a regularly rotating galaxy 1.2 billion years after the Big Bang *F. Lelli* et al.
PERSPECTIVE p. 674

716 Superconductivity
Two-dimensional superconductivity and anisotropic transport at $KTaO_3$ (111) interfaces *C. Liu* et al.

722 Thermoelectrics
Enhanced atomic ordering leads to high thermoelectric performance in $AgSbTe_2$ *S. Roychowdhury* et al.
PERSPECTIVE P. 678

728 Geology
Orogenic quiescence in Earth's middle age *M. Tang* et al.

731 Oceanography
Seismic crustal imaging using fin whale songs
V. M. Kuna and John L. Nábělek
PODCAST

735 Coronavirus
Mosaic nanoparticles elicit cross-reactive immune responses to zoonotic coronaviruses in mice *A. A. Cohen* et al.

741 Coronavirus
Immunological characteristics govern the transition of COVID-19 to endemicity *J. S. Lavine* et al.

DEPARTMENTS

653 Editorial
Monitoring change
By H. Holden Thorp

750 Working Life
Balancing act
By Ryoichi Fujiwara

ON THE COVER

Diversification is a widely proposed policing reform, but data constraints make it difficult to test whether officers of different identities behave differently when facing comparable circumstances. Using microlevel records from Chicago, a new study shows that officers from marginalized groups use force less often and make fewer stops and arrests for minor infractions than white and male officers. These disparities are particularly pronounced during encounters with Black civilians. See pages 661, 677, and 696. *Illustration: Reyna Noriega*

New Products 746
Science Careers 747

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

PHOTO: ESO

EDITORIAL

Monitoring change

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

The *Science* family of journals is launching a number of new initiatives related to our editorial policies. We now have a web page listing several important metrics for scientific journals (www.sciencemag.org/journal-metrics). This information will be updated at the end of each year with new data.

One set of metrics is the numbers of research paper submissions and the rates at which they are peer reviewed and ultimately accepted for publication. The *Science* journals receive many submissions each year and select only a subset (11 to 32% across the journals) for external review. Decisions about peer review are made by professional editors at *Science*, *Science Immunology*, *Science Signaling*, and *Science Translational Medicine*; by academic editors for *Science Advances*; and by a combination of academic and professional editors for *Science Robotics*. Submissions to *Science Advances* continue to increase, showing the enthusiasm for the journal and its practices. There are now more than 300 members of its editorial board under the leadership of Ali Shilatifard of Northwestern University.

Another set of metrics shows the time that elapses at various stages of the review process. All of us—editors, authors, and reviewers—would love to move even faster, but we hope that by posting these numbers, the *Science* journals can work with reviewers and authors to get papers evaluated, revised, and published more efficiently than ever. As a reminder, the *Science* family of journals supports the posting of preprints when the authors choose to make information available while the formal peer review process proceeds. Our journal metrics web page also shows the numbers of citations for published papers at each journal over 2 years. These data are used to calculate the journal impact factor (JIF). The *Science* family is a signatory of the Declaration on Research Assessment (DORA), which seeks to broaden the ways in which the impact of journals and papers is assessed beyond the traditional JIF calculation. As the Declaration notes, citation distribution within a given journal can be highly skewed. Thus, a small number of highly cited papers can inflate the mean number of papers that have a high citation value well beyond the modal citation count. In the interest of transparency, we are displaying the full distribution of citations in a histogram format for each journal.

> **"...the *Science* journals remain committed...to produce a durable and accurate scientific record."**

An additional category of metrics relates to the impact of the journals taken from Web of Science and Scopus, two resources of comprehensive data on scholarly peer-reviewed publications. The criticisms of JIF are that it was never intended to be a measure of the quality of a paper published in a particular journal and that it is subject to manipulation by journals as well as by institutions that use the calculation as a key driver of decisions about promotions, tenure, and research funding. Thus, our policy is to report not only the JIF for the *Science* family of journals, but also a variety of other related metrics.

In 2017, *Science* joined a working group of publishers and academics that aimed to align standards for the transparent reporting of life science research. This led to the creation of the Materials Design Analysis Reporting (MDAR) framework, a means for scientific journals to apply the Transparency and Openness Promotion (TOP) Guidelines to articles in the life sciences. The framework provides both minimum and best practice recommendations. In 2019, *Science* participated in a pilot study with 13 journals from other publishers that tested implementation of the proposed minimum recommendations through the MDAR checklist. After the pilot, we adopted the draft checklist with the goal of improving reproducibility of our life science papers. We have received positive responses from authors and plan to continue using the checklist, which will be updated when final recommendations of the MDAR working group become available.

Also, we were pleased to join 14 other publishers in a working group to promulgate new guidelines for software citation, a critical element of the scientific record.

After a year like no other, the *Science* journals remain committed to partnering with scientists around the world to produce a durable and accurate scientific record. We are deeply grateful to the authors, reviewers, and others who work with us to push science forward every day.

–H. Holden Thorp

PHOTO: CAMERON DAVIDSON

10.1126/science.abg9540

NEWS

A patient with COVID-19 breathes oxygen at a hospital in Cape Town, South Africa, the African country hit hardest by the pandemic.

PHOTO: RODGER BOSCH/AFP/GETTY IMAGE

IN BRIEF

Edited by **Jeffrey Brainard**

DISPATCHES FROM THE PANDEMIC

South Africa halts vaccine drive

COVID-19 | South Africa this week halted plans to widely roll out a COVID-19 vaccine produced by AstraZeneca and the University of Oxford after a clinical trial in the country showed it had an efficacy of less than 25% against mild and moderate disease. Researchers blamed a virus variant that dodges key antibodies and has become widespread in that country. Scientists are still hopeful that the two-dose vaccine may do better against serious disease. The recent trial in 2000 people was too small, and its participants too young, to determine whether it does, but South Africa may further test it in 100,000 people. It's "a very plausible scenario" that the AstraZeneca-Oxford vaccine offers protection against serious illness, Katherine O'Brien of the World Health Organization said at an 8 February press conference. Other scientists also cautioned against dismissing the value of the shots; AstraZeneca has priced its vaccine for wide use in developing countries, at a cost less than that of other COVID-19 vaccines. The company projects it can produce 3 billion doses this year for about $3 each. A similar COVID-19 vaccine developed by Johnson & Johnson also protected less well against mild and moderate disease in South Africa than elsewhere, but still convincingly warded off severe illness and death, even in those infected with the new variant, named B.1.351, which has now been found in people in more than 40 countries.

WHO team rejects lab leak theory

PUBLIC HEALTH | An international team that spent almost a month in Wuhan, China, investigating the origins of COVID-19 on behalf of the World Health Organization (WHO) has firmly rejected the idea that the pandemic resulted from a lab accident. A "detailed discussion" with staff at the Wuhan Institute of Virology, along with other evidence, showed it was "extremely unlikely" that the SARS-CoV-2 virus that causes the disease escaped from the lab, mission chief Peter Ben Embarek, a WHO food safety scientist, said at a 9 February press conference. Allegations of a lab mishap, or even a deliberate release, circulated widely in the past year and soured U.S.-Chinese relations. The WHO team will no longer pursue that notion, but hopes to further investigate other hypotheses about how the virus jumped from animals to humans, Embarek added.

FDA limits use of blood plasma

THERAPEUTICS | Treating people who have COVID-19 with the antibody-rich blood plasma of recovered patients received a controversial boost last year, when the U.S. Food and Drug Administration authorized its emergency use. Some scientists decried the decision, saying the benefits of such convalescent plasma were small and not solidly supported by data. Last week, the agency revised its authorization, requiring any plasma used to meet a minimum threshold of antibodies and restricting treatment to hospitalized patients early in their disease or people who have poor antibody production themselves. Agency officials said the changes came after a review of recent clinical trials—two were stopped because they found no benefits to patients, but several others had more encouraging data. The officials said they still believe convalescent plasma is a low-risk treatment that may help against COVID-19 in certain cases.

Facebook takes on vaccine lies

SCIENCE COMMUNICATION | Facebook this week said it would take new steps to combat misinformation about all vaccines, including those for COVID-19, and about the disease itself. The social media giant and its subsidiary Instagram will remove posts containing certain debunked claims about vaccines, including that they cause autism, as well as the assertion that getting a COVID-19 vaccine is riskier than the disease itself. In addition, when users search the platforms using search terms related to debunked claims about vaccines, Facebook will provide validated information in response.

SCIENCEMAG.ORG/TAGS/CORONAVIRUS
Read *Science*'s online coverage of the pandemic

A protester in Moscow is one of thousands who attended demonstrations for jailed opposition leader Alexei Navalny.

PHOTO: MAXIM SHEMETOV/REUTERS

HUMAN RIGHTS

Russian scientists protest arrest of opposition leader

Researchers are joining a rising chorus of international protest over the jailing of Russian opposition leader Alexei Navalny. More than 840 Russian scholars at home and abroad have signed a letter calling for an "open and honest" investigation into the politician's near-fatal poisoning in Siberia in August 2020. European defense labs determined he'd been exposed to a Novichok nerve agent—a finding the Russian government has rejected. When Navalny returned to Moscow on 17 January, authorities arrested him on an alleged parole violation, and last week he was sentenced to 3.5 years in prison. Russians took to the streets nationwide, and police have arrested more than 11,000. The letter also calls on Russian authorities to end Navalny's "persecution" and "begin an open dialogue with society." The letter "is a warning," says Alexander Kabanov, a pharmaceutical chemist at the University of North Carolina, Chapel Hill, one of the signers. "Brutal suppression of peaceful demonstrators will lead the country to catastrophe."

Amazon monitoring faces cuts

ENVIRONMENT | Brazil's National Institute for Space Research (INPE), the leading source of satellite data about deforestation in the Amazon, is facing staff and budget cuts as the country's military gears up a parallel service. Citing financial constraints, the Brazilian science ministry this month suspended payments for 107 fellowships that supported nearly one-quarter of INPE's technical staff—just as the institute was preparing to launch the country's first homemade satellite. INPE may lose 15% of its total funding under a proposed budget for this year, which also includes a 34% cut for the science ministry. Meanwhile, the Brazilian air force signed a contract in December 2020 to buy a satellite imaging system to help monitor the Amazon and "complement" existing data. Brazilian President Jair Bolsonaro has heavily criticized INPE in recent years for revealing steep increases in deforestation and fire activity in the Amazon, and he fired INPE Director Ricardo Galvão in 2019. Critics fear the recent cuts are an attempt to undermine the institute's influence.

Big cash prize for carbon capture

GLOBAL WARMING | Elon Musk, aerospace and electric vehicle entrepreneur and the world's wealthiest person, announced this week details of a $100 million prize he has funded to spur the development of technology to affordably capture carbon dioxide (CO_2) from the atmosphere and store it. The 4-year contest will be administered by the XPrize Foundation, with by far the largest award of any technical competition it has run. To avoid the worst impacts of global warming, many climate scientists now believe atmospheric CO_2 removal will be necessary, even if rapid reductions in emissions begin this year. A few companies have tested prototype technologies, but none has yet proved affordable and effective enough for mass deployment.

Radio array gains momentum

ASTRONOMY | The Square Kilometre Array Observatory (SKAO), a dream of radio astronomers for nearly 3 decades, drew closer to reality last week as its governing body met for the first time and began to review plans to begin construction this year. Participating nations ratified a treaty for the project late last year. As the world's biggest radio observatory, with thousands of antennas planned at sites in Australia and South Africa, the €2 billion project could see the universe's first stars and galaxies. External experts have green lighted

a design. The SKAO negotiated funding contributions with the six nations signed up so far—Australia, Italy, the Netherlands, Portugal, South Africa, and the United Kingdom—and others expected to join soon; the payments will be based on each country's volume of astronomy research. The additional countries are Canada, China, France, Germany, India, Spain, Sweden, and Switzerland.

Postage stamp honors physicist

HISTORY OF SCIENCE | The U.S. Postal Service this week issued a stamp commemorating Chien-Shiung Wu, a Chinese-American nuclear physicist who proved that the universe knows its right hand from its left. In 1956, Wu and her team studied the emission of electrons in the radioactive decay of cobalt-60. They found that with respect to the nucleus' spin, the electron emerges in one way more often than the other, violating a

mirrorlike symmetry called parity. Parity violation has become a foundation of modern particle theory, and it earned the 1957 Nobel Prize for the two theorists who suggested it might occur—but not for Wu, who died in 1997. Now, she joins a far more selective club than Nobelists: the eight other physicists who have appeared on U.S. stamps.

India eyes creating its own NSF

FUNDING | India's budget for 2021–22 includes money to create a research funding agency, along the lines of the U.S. National Science Foundation, to boost the scientific capacity of universities and colleges not in the country's top tier of institutions. The proposal includes a total of $6.9 billion over 5 years for the new National Research Foundation, some of it newly appropriated and the rest repurposed from other research programs. Supporters hope the new agency will accelerate the nation's science by funding grantees at thousands of institutions that have historically received almost no national research money. The budget aims to sustain current levels of support for India's hundreds of top-tier universities.

Wealth shapes cancer trial results

BIOMEDICINE | Late-stage clinical trials of cancer treatments led by scientists in lower income countries are more likely to identify successful therapies that benefit patients than analogous studies in high-income countries, researchers report. They examined 694 phase 3 oncology trials published from 2014 to 2017, including a subset that reported whether the interventions provided a substantial benefit to patients. Just under half of 21 trials by authors in low- and middle-income countries reported a substantial benefit versus only one-third of 145 conducted by groups in high-income countries. What's more, trials in lower income countries were more likely to report statistically significant benefits, such as a halt to tumor growth, according to the study published 28 January in *JAMA Oncology*. Its authors attribute the differences to "a level of pragmatism" among cash-strapped scientists in lower income countries; funding limitations may encourage them to adopt productive trial designs. The authors call for these studies to receive adequate funding.

Tweeted papers don't always click

SOCIAL MEDIA | Many scientists share tweets containing links to published journal articles, hoping it will speed discussions of the latest science. But half of the links examined in a recent study drew no clicks, and another 22% drew just one or two. Zhichao Fang of Leiden University and colleagues analyzed 1.1 million links on Twitter to scholarly articles published from 2012 to 2017. Only about 10% of links received more than 10 clicks, according to the paper published 23 January in the *Journal of the Association for Information Science and Technology*. The authors could not track the outcome of all the tweeted links, only those created using the bit.ly link-shortening service. The study is one of the first about whether people engage with the content of scholarly tweets beyond clicking Twitter's like and retweet buttons.

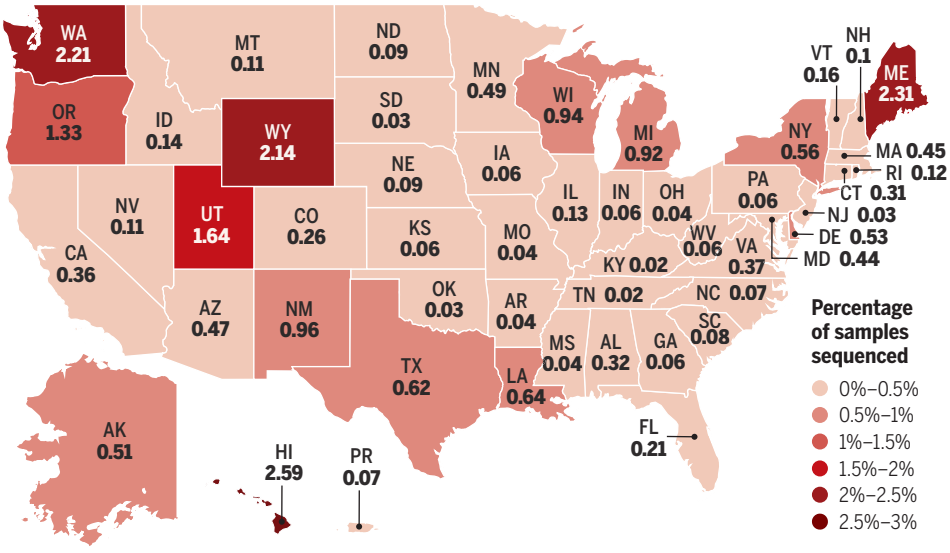
The control room for the United Arab Emirates's mission to Mars.

PLANETARY SCIENCE

UAE's first Mars probe arrives in orbit

The arrivals lounge at Mars begins to fill up this week with two probes checking in, both from nations venturing to the Red Planet for the first time. The United Arab Emirates's Hope mission arrived in orbit on 9 February. The spacecraft will examine martian atmosphere and climate from a higher vantage point than most other satellites to get a global perspective. It will also seek to help scientists understand how hydrogen and oxygen escape from the lower atmosphere into space. China's Tianwen-1, which includes an orbiter, lander, and rover, was due to arrive this week after *Science* went to press. Its orbiter, which aims to study Mars's atmosphere and magnetic field as well as mapping its surface geology, will hold onto its landing cargo for several months. When the rover eventually touches down, it will use ground-penetrating radar to study subsurface features. NASA's Perseverance rover is expected to join the party on 18 February, and head straight to the Jezero crater.

IMAGES: (TOP TO BOTTOM) © 2021 U.S. POSTAL SERVICE; CHRISTOPHER PIKE/BLOOMBERG/GETTY IMAGES



CREDITS: (MAP) K. FRANKLIN/*SCIENCE*; (DATA) GISAID; JOHNS HOPKINS CORONAVIRUS RESOURCE CENTER

Rare and uneven sequencing of coronavirus samples leaves the United States all but blind to emerging viral strains. Scientists are calling for sequencing at least 5% of confirmed COVID-19 cases nationally, a target not currently reached by any state.

COVID-19

United States rushes to fill void in viral sequencing

Some experts say Biden administration plan is too modest to track fast-evolving variants as they arise and spread

By **Meredith Wadman**

Since May 2020, Jeffrey Milbrandt has had his systems fine-tuned to sequence 1000 coronavirus samples a week. The director of a major sequencing center at Washington University in St. Louis (WashU), Milbrandt knew months ago that the United States urgently needed to identify and track emerging variants of SARS-CoV-2, the pandemic coronavirus already spreading across the nation.

But to date, fewer than 100 coronavirus samples have made it to his sequencers at the McDonnell Genome Institute, and the United States remains nearly blind to several coronavirus strains that have recently upended the course of the pandemic. "We have it all worked out but there's not a lot of takers," Milbrandt says of his center's sequencing abilities. "We are getting more inquiries from the press than from people who need the information. … Some of us have pipelines available—they are just not being utilized."

The Centers for Disease Control and Prevention (CDC) plans to rapidly change that by coordinating a national surveillance system for coronavirus variants. Scientists and public health experts welcome the plan. "The hardware is set up, the people are trained, the infrastructure is in place," says genome scientist Chris Mason of Weill Cornell Medicine. "It's just a big question: Who's gonna pay for it?"

Given the magnitude of the problem, some scientists worry current plans don't go far enough. That WashU, one of the flagship centers that sequenced the human genome 2 decades ago, is idling in neutral is just one sign of the massive challenge the Biden administration faces as it tries to keep up with the virus. The latest bit of bad news on that front came on 7 February, when a team of U.S. scientists posted worrisome findings.

Their sequencing identified 212 instances of a highly contagious variant, B.1.1.7, from samples collected in the United States in December and January. They calculated that the variant is doubling in relative frequency in the United States every 10 days and is 35% to 46% more transmissible than the original virus. (Preliminary studies from the United Kingdom, where the variant was first identified, also suggest it may be more lethal.) In Florida, the variant accounted for about 4.5% of new cases by late January. Knowing what variants are present and where can be vital for policymakers. For example, skyrocketing levels of B.1.1.7 in England spurred strict winter lockdowns.

The United States is a sequencing superpower, but as of 7 February it ranks 36th in the world in sequencing SARS-CoV-2, according to data compiled by *The Washington Post*. To date, the United States is sequencing the genomes of just 0.36% of confirmed cases, with state rates ranging from 0.02% to 2.6%, according to *Science*'s analysis of public databases (see map, left). The dismal showing is "totally unacceptable," White House COVID-19 Coordinator Jeffrey Zients said at a recent press briefing. In contrast, Denmark, a leader in monitoring variants, has sequenced more than half of positive cases (*Science*, 5 February, p. 549).

But scaling up variant surveillance involves many players, and sequencing itself isn't the rate-limiting step, says Duncan MacCannell, chief scientist of the CDC program that handles genomic surveillance of coronavirus. "If [we] had a straightforward way to connect the sequencing capacity at Washington University with samples and data, we would likely be using them."

A White House document circulating in Washington, D.C., last week calls for spending $340 million to sequence 5% of samples from confirmed COVID-19 cases, which would be about 39,000 genomes per week at current infection rates. But some scientists support a $2 billion request to sequence up to 15% of cases, introduced in a bill last week by Senator Tammy Baldwin (D–WI).

Other scientists say current efforts need to ramp up faster. "The Biden administration is doing a lot, but … it is not yet sufficient," says Rob Knight, a sequencing expert at the University of California, San Diego (UCSD), and a co-author on the preprint on B.1.1.7 in the United States. "Far more sequencing needs to be done to understand the emergence and spread of new variants [and] to design new targeted probes to detect these variants cheaply and easily without sequencing every sample."

Sequencing all 30,000 bases of SARS-CoV-2's RNA genome allows researchers to detect misspellings, deletions, and other mutations that could give the virus new capabilities, such as spreading more easily, making people sicker, or evading vac-

cines. Two variants first identified in South Africa and Brazil, for example, may escape vaccine-induced antibodies. Like B.1.1.7, both have now been found in the United States, albeit only in a handful of cases.

CDC is responsible for SARS-CoV-2 surveillance, but it must work with a patchwork of state and local public health agencies. And until December, it had limited funding for sequencing as troubles with diagnostic testing consumed attention and the Trump administration showed little interest. "Things were on a low slow burn for a while," says Scott Becker, CEO of the Association of Public Health Laboratories.

Early last fall, CDC awarded small sequencing contracts to seven academic centers. For example, scientists at UCSD and Scripps Research got $2.5 million to sequence samples in California and Mexico. Other small efforts moved ahead without CDC help. A group at Cedars-Sinai Medical Center, for example, last month identified a new variant spreading in California that contains a mutation that may help the virus invade cells more easily.

Then in late December, after the United Kingdom sounded the alarm over B.1.1.7, CDC finally got the go-ahead to support more sequencing. By 1 February, the agency had launched coronavirus-sequencing contracts worth a combined $12.54 million with Illumina, its collaborator Helix, and medical testing giants LabCorp and Quest Diagnostics. They have committed to analyzing 6000 virus genomes per week by midmonth, CDC Director Rochelle Walensky said last week. Combined with sequences coming from academic centers and state public health labs, the United States will soon be sequencing "up to threefold more" samples than today, Walensky predicted on 8 February.

CDC is looking to expand its contracts with companies like LabCorp and Quest, MacCannell says. They "give us the ability to sample from their testing network, which combined spans the entire country. … It has been a fantastic way of turning the lights on in parts of the country where we didn't have a lot of data."

Such partnerships are essential to a coordinated, national, virus-surveillance system, says Patrick Ayscue, a senior biosecurity fellow at the Chan Zuckerberg Biohub, which is playing a major role in coronavirus sequencing in California. "One of the primary bottlenecks in the U.S. is getting that sample onto a sequencer and then those data back to people who can interpret that information and act on it."

Engaging huge diagnostic labs like Quest and LabCorp along with public health labs in every state will help avoid "genomic deserts," says Phil Febbo, chief medical officer at Illumina. For instance, not enough sequencing is being done now in majority Black communities in some southern states, says James Lillard, an immunologist at the Morehouse School of Medicine.

CDC has also begun to require that state and local public health labs submit, collectively, 750 sequences each week, Walensky noted in late January. But such labs have been both underfunded and swamped during the pandemic. At the Utah Public Health Laboratory—one of the most sophisticated and best equipped of the state labs—a shortage of pipette tips drove sequencing rates down from 600 per week in mid-January to 100 last week, says Kelly Oakeson, the lab's chief sequencing scientist.

> **"One of the primary bottlenecks in the U.S. is getting that sample onto a sequencer and then those data back to people who can interpret [them]."**
>
> **Patrick Ayscue**, Chan Zuckerberg Biohub

Other scientists say CDC's goal is too low. "Seven hundred and fifty surveillance strains a week is nothing," says Lea Starita, a genome scientist at the University of Washington, Seattle, and the Seattle Flu Study. Combined with the 6000 viral genomes CDC expects to receive weekly from the companies, that would allow sequencing of just 0.9% of the current weekly average of nearly 800,000 COVID-19 cases.

Many scientists support the 5% goal, including those on a white paper sent on 14 January to the presidential transition team by a coalition of 14 academic labs, companies, and others. They estimated that their group, if given funding today, could actually do much more—more than 60,000 genomes per day within 10 weeks. Febbo estimates that would cost $500 million in 2021.

Other scientists are bullish on the idea of aiming higher than 5% of genomes. David O'Connor, a viral sequencing expert at the University of Wisconsin, Madison, says he and colleague Tom Friedrich, working with local stakeholders to track university-to-community transmission, have already achieved that 5% benchmark in Dane county, funded by a $1.3 million CDC contract issued last fall. "If you can do that in a county like ours … if you professionalize it and coordinate it … you can absolutely get to 15%. It just takes resources and will." ■

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

EVOLUTION

Genes for life on land evolved earlier in fish

Hidden genetic pathways behind land-water transition are found in living fish

By **Elizabeth Pennisi**

Almost 700 years ago, Jacob van Maerlant, a Dutch poet, envisioned a fish all set for life on land: It had sprouted arms to hoist itself ashore. Now, three genetic studies make his fantasy look remarkably prescient. Together, the studies suggest that in terms of genes, the aquatic precursors of four-limbed land animals, or tetrapods, were as well-prepared as the Dutch fantasy fish. They were pre-equipped with genes that could be turned to making limbs, efficient air-breathing lungs, and nervous systems tuned to the challenges of life on land.

"All these studies tell us that the origin of tetrapods was something waiting to happen," says Borja Esteve-Altava, an evolutionary biologist at Pompeu Fabra University in Barcelona, Spain. Genetically, "Everything necessary was already there" before vertebrates came ashore, nearly 400 million years ago.

Fossils reveal the outlines of the story. Fish with fleshy fins supported at their base by a single bone, known as lobe-finned fish, moved into shallow water about 375 million years ago. About 5 million years later, some of those lobe fins crawled onto terra firma. The first fish to set fin on land must have already had at least some of the physical traits and genetic modifications needed to do so, but researchers hadn't worked out how and when they became equipped for the change. "The big question of how such a large morphological shift actually occurred remains very much in play," says Peter Currie, an evolutionary developmental biologist at Monash University.

In the trio of studies published last week in *Cell*, genes in living fish took the place of fossils as a way to peer back in time. One set of clues came from studies of mutagenized zebrafish, a favorite model for studying development. M. Brent Hawkins, then a Harvard University graduate stu-

dent and now a postdoc, was shocked to discover zebrafish mutants with two bones resembling the forelimb bones of land animals in their front fins, complete with muscles, joints, and blood vessels. The finding is "quite spectacular," says Marie-Andrée Akimenko, a developmental bio-logist at the University of Ottawa.

Two mutated genes, *vav2* and *waslb*, were responsible for the transformation. Both genes code for proteins that are part of a pathway controlling the activity of Hox11 proteins, regulatory molecules that guide the formation of the two forearm bones in mammals, among other functions. In fish, other proteins normally suppress Hox11 and prevent the formation of those bones. But the mutations, which Hawkins re-created using the gene editor CRISPR, reactivate the pathway. The "landmark" finding is "changing the paradigm on limb development and evolution," says Renata Freitas, a developmental biologist at the University of Porto in Portugal.

This fish with arms appeared in a medieval Dutch poet's manuscript.

Other genetic clues come from living representatives of ancient fish lineages. Only two groups of the lobe-finned fish are alive today: lungfish and coelacanths. About 400 million years ago, they diverged from the line of lobe-finned fish that gave rise to tetrapods 30 million years later. Today's oceans are mostly populated with species from another group that originated about 420 million years ago: the ray-finned fish, so named because their fins are supported by slender spines.

Evolutionary geneticists Guojie Zhang at the University of Copenhagen and Wen Wang of Northwestern Polytechnical University in Xi'an, China, and their colleagues sequenced the genomes of the African lungfish, which branched off early from other lobe-finned fish. The researchers also sequenced the bichir, an elongated, air-breathing, ray-finned fish that lives in the shallows of tropical African rivers, as well as the American paddlefish, the bowfin, and the alligator gar. All are ray-finned fishes that evolved much earlier than teleosts, the group that dominates the world's waters today (see diagram, below). Knowing when each of those lineages branched away from others, the researchers could infer when and where certain genes first appeared on the fish family tree.

None of the sequenced fish is on the precise branch that led to tetrapods. Yet all have much of the genetic equipment needed for life on land, including most of the genes and regulatory DNA needed to build limbs. For example, all the fish sequenced have a regulatory element that helps form synovial joints, which make fins and limbs flexible and are essential for terrestrial locomotion. The fish also have 11 genes that are needed to build lungs and that work the same way in the bichir's lungs as they do in humans. One is for a pulmonary surfactant, a lubricating secretion that helps lungs expand and contract. Both the ray-finned fishes and the lobe-finned lungfish also apparently have a regulatory element that helps shape the right ventricle of the heart to deliver oxygen more efficiently.

The findings show that "a lot of things we think are just in land animals are also in fish," says Gage Crump, a developmental biologist at the University of Southern California. Finding all those genes in both lobe-finned and ray-finned fish means those genetic pathways must have been present in their common ancestor, some 425 million years ago. "It is surprising that some of these elements are so conserved for such a long evolutionary time," Zhang says. (Teleosts, in contrast, have lost much of the DNA that prepared early fish for life on land, apart from the Hox11 pathway, the team reported.)

The genome of the lungfish offers a glimpse of later adaptations along the path to terrestrial life. It includes additional pulmonary surfactant genes that the ray-finned fishes lack, as well as DNA for specifying five toes, connecting nerves to limb muscles, and for sensitizing the brain to react fast. All those genes were previously thought to be unique to tetrapods.

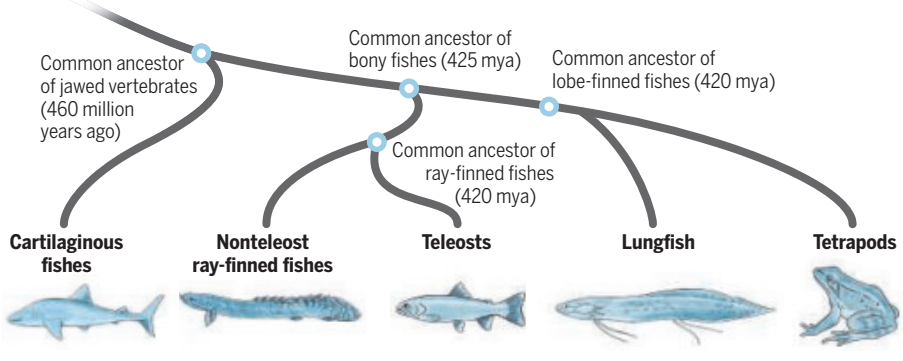
Putting it all together, Wang and Zhang think the transition to land involved three key steps. The ability to breathe air occasionally appeared in the common ancestor to ray- and lobe-finned fish, about 425 million years ago. Then surfactant genes, new nervous system genes, and other innovations enabled lobe-finned fish to leave the water temporarily. Finally, after the African lungfish split off from the lobe fins, the common ancestor of land vertebrates acquired other respiratory and locomotive refinements needed to live out of water.

Rather than building new structures and genetic pathways just when vertebrates moved onto land, evolution apparently was thrifty, using existing genes to adapt to the opportunities offered by terrestrial habitats. "[The studies] show the extent to which the fish-tetrapod transition was achieved by modifying existing molecular systems, rather than creating new ones," says Per Ahlberg, a paleontologist at Uppsala University.

Gaps still remain in understanding how fish made landfall, but the new studies "bring us closer to the living biology of the fish-tetrapod transition," Ahlberg adds. Van Maerlant would be pleased. ■

The long swim to land

The groundwork for terrestrial traits like limbs and lungs was laid deep in the fish family tree. Genes for such traits found in both lobe-finned and ray-finned fishes must also have been present in their common ancestor.



CREDITS: (IMAGE) DER NATUREN BLOEME/NATIONALE BIBLIOTHEEK; (GRAPHIC) X. BI *ET AL.*, *CELL*, 2021. DOI.ORG/10.1016/J.CELL.2021.01.046. ADAPTED BY N. DESAI/*SCIENCE*

COVID-19

Pandemic hit academic mothers hard, data show

New evidence highlights need for policy changes to aid female scientists

By **Katie Langin**

In March 2020, Reshma Jagsi—a radiation oncologist at the University of Michigan, Ann Arbor—wrote an opinion piece predicting female scientists would feel a disproportionate impact from the COVID-19 pandemic. Skeptical journal editors declined to publish it. Since then, though, many commentators have echoed her message. And now the evidence has become clear: The pandemic has exacerbated existing disparities and created additional challenges for women, especially those with children, struggling to maintain their research productivity.

In some fields, studies show, the proportion of female authors on preprints, submitted manuscripts, and published papers dropped during the first few months of the pandemic. Mothers also suffered a 33% larger drop in research hours compared with fathers, according to a global survey of 20,000 Ph.D. holders published as a National Bureau of Economic Research working paper last month. The survey, conducted from May to July 2020, also found that mothers took on more household and child care duties than fathers.

But the news isn't all dire: One funding agency recognized and corrected for a gender disparity early on in the pandemic. In February 2020, the Canadian Institutes of Health Research (CIHR) offered funding for COVID-19 research. The agency wanted to address the fast-growing issue quickly, so researchers were given an unusually short time to submit proposals—just 8 days. This was before lockdowns were imposed in Canada, but the gender disparity in the responses highlight underlying challenges female academics face: Only 29% of the resulting proposals were led by women, a drop of roughly seven percentage points compared with previous comparable funding opportunities.

"As soon as we got the numbers we went, 'Oh my goodness, we did something wrong,'" says Cara Tannenbaum, scientific director of the Institute of Gender and Health at CIHR. When the agency offered a second round of COVID-19 funding 2 months later, it extended the deadline to 19 days and reduced the paperwork requirements. The number of proposals by women jumped to 39%, as Tannenbaum and her co-authors describe in a paper published this week in the *Proceedings of the National Academy of Sciences*.

Holly Witteman, a health researcher at Laval University, saw the benefit of the policy shift. "I looked at the 8-day deadline and I just thought, 'There's no way,'" says Witteman, who has two children. "I can't just cram in extra work by staying up late to write another grant." The extended deadline during the second round afforded her enough time to put together a proposal, which was ultimately funded.

Still, she worries about the repercussions of the gender disparity in the first round, which had a much higher success rate (44% versus 9%). "When you have funding, you get to do studies, and then you get to publish them, and then when you apply for more funding … you're in a much better position; it's this cumulative advantage snowball that just keeps rolling."

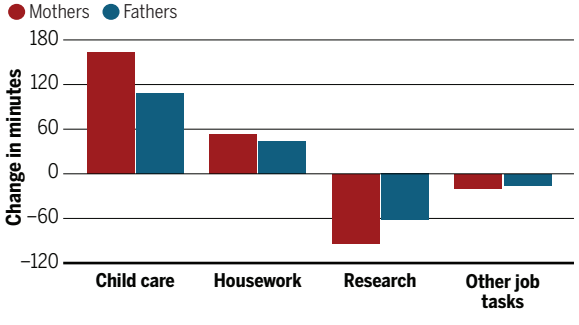
COVID-19 put many existing inequalities in "stark relief," says Robin Nelson, an anthropologist at Santa Clara University whose work hours were slashed by more than half last year when her two children were home from school. "Some people are just simply working against more obstacles than other people." To level the playing field, she adds, "We are going to have to finally start questioning business as usual in the way that we administer grants, … time to tenure, all of these policies."

Robinson Fulweiler, an ecologist at Boston University, says universities and funding agencies should give scientists the option of submitting COVID-19 impact statements going forward, describing how the pandemic impeded their work. Employers should also take steps to ensure all researchers have access to affordable day care, she says. Fulweiler and other scientist-mothers detail those recommendations and many others in an opinion piece in press at *PLOS Biology*. (It appeared as a preprint in September 2020.) "Now is the time to act," she and her coauthors write. "Rather than rebuilding what we once knew, let us be the architects of a new world."

Nelson agrees. "I hope that COVID has … given us some space to actually think critically about what we expect of faculty and whether we can make space for faculty to have full lives that might involve disabilities, chronic illness, care taking of any sorts—just anything outside of work," she says. "Academe has really privileged people who can get other people to do all of that labor for them. And so we need to really start thinking about what does a sustainable academic environment with healthy faculty look like and how can we make space for that without thinking that we're compromising rigor." ■

Motherhood penalty

In the months after the pandemic started, most researchers worked fewer hours per day than usual. But parents—especially mothers—lost the most work time, according to a new study. Mothers also spent more time on child care and housework than fathers.



CREDITS: (GRAPHIC) K. LANGIN/*SCIENCE*; (DATA) DERYUGINA *ET AL.*, NATIONAL BUREAU OF ECONOMIC RESEARCH, 10.3386/W28360 (2021); (PHOTO) ISTOCK.COM/DAMIRCUDIC

The 2014 shooting of Laquan McDonald by Chicago police sparked protests.

PHOTO: ARMANDO L. SANCHEZ/*CHICAGO TRIBUNE*/TRIBUNE NEWS SERVICE/GETTY IMAGES

SCIENCE AND THE LAW

Study: Police diversity matters

Landmark analysis of 7000 Chicago police shows nonwhite and female officers make fewer stops

By **Douglas Starr**

Bocar Ba spent a good part of his youth "scrambling" from police, he says. The University of California, Irvine, economist grew up in the notorious banlieues of Paris, and as a young Black man, he often encountered law enforcement in ways that most scholars do not. Years later, as a doctoral student of public policy at the University of Chicago, he continued to think about police behavior and its impact on civilians.

In 2015, while attending a seminar, he was struck by the assumptions fellow academics were making, based on a few ride-alongs in a police car. And so, like many other academics in the wake of the shootings of Michael Brown and other Black civilians by police, he decided to probe an important social question: Does a police officer's race matter when it comes to the use of force, or even the decision to stop someone in the first place? Although many activists, academics, and even police departments have answered yes—and hired more minority officers to improve community relations—a link has never been proved, and the field has been riddled with contradictory studies.

Now, Ba and his colleagues have added some certainty to that discussion. After combing through millions of police records from 2012 to 2015 and analyzing them for the nature of the action, the time of day, the race of the civilian and the officer, and many other factors, they found that Black, Hispanic, and female officers in Chicago made fewer stops and arrests than their white male counterparts, especially for petty crimes.

The study, the most fine-grained of its kind to date, "is really valuable," says Jeffrey Fagan, a professor of law at Columbia University who was not involved in the work. "A lot of other studies out there have tried to use officer race as a benchmark test for discrimination … [but this one] got much more precise estimates of what the police workplace actually looks like."

Getting the initial data for those estimates was a stroke of luck. In 2014, an investigative group of journalists and lawyers in Chicago called the Invisible Institute had won a court ruling forcing the police department there to release thousands of complaint records from several previous years. Ba approached the organization and suggested that together they build a fuller view of police-community relations by looking for additional data.

Ba studied police union contracts, which helped him work out which city departments might host the records. "I said, 'Give me a couple of lawyers and I'll tell them exactly where to get the data we need,'" he recalls. Soon a pattern developed: Ba would direct the institute toward data, and it would sue for additional records—a process that continued for several years. "Bocar wanted to understand the nuance, so he asked us all sorts of challenging questions," says Chaclyn Hunt, a civil rights attorney with the institute.

In 2019, Ba started to collaborate with Dean Knox at the University of Pennsylvania's Wharton School and Jonathan Mummolo at Princeton University—quantitative social scientists who use massive amounts of data to analyze social policies and behaviors. Together with Ba's collaborator, Roman Rivera, they set out to scrutinize the behavior of Chicago police in the most fine-grained way possible.

To make the most valid comparisons, the researchers first had to "prune" their collection. For example, two officers could have the same badge number if one retired and passed it on, or the same names if they were father and son, so the team also had to use birthdates. To properly compare arrest and stop records, they had to geolocate each incident. They also threw out the records of higher ranking officers and explosives technicians, who tended not to interact with civilians. The result, Mummolo says, "was a real Frankenstein data set." But it allowed the team to refine its list to nearly 7000 officers, who, over 3 million shifts, had engaged in 1.6 million stops, arrests, and uses of force over a period of 3 years.

When they analyzed their data, they found Black officers in Chicago made far fewer stops and arrests, by 29% and 21%, respectively, than their average white counterparts. They used force 32% less frequently. Overall, Black officers stopped 17% fewer white civilians than their white counterparts, and 39% fewer Black civilians. Most of the differences involved discretionary stops for "suspicious" activity or minor violations; the researchers saw little difference in stops and arrests for violent crimes, they report on p. 696. They saw a similar pattern with Hispanic officers. Across all races, female officers made fewer stops for minor violations than male officers.

The results make a strong case that diversifying a police force can reduce conflicts between officers and the community without making any trade-offs in public safety, other researchers say. "These are not small numbers," says Phillip Atiba Goff, professor of African American studies and psychology at Yale University and co-founder of the Center for Policing Equity, whose commentary on the study appears on p. 677. "They imply that you can save lives by increasing the number of Black and women officers."

"The whole question whether diversity makes a difference or not in policing is a huge policy issue," says Christopher Winship, professor of sociology at Harvard University. "A paper like this can have a huge impact on the local level going forward." ■

Douglas Starr is a journalist in the Boston area.

WILDLIFE

Illicit centipede spurs taxonomy journal debate

Editors differ about whether they are obliged to ensure specimens are legally acquired

By **Yao-Hua Law**

In 2018, a new species of centipede graced the pages of the prominent taxonomy journal *Zootaxa*. More than 14 centimeters long, with striking teal-colored legs, it lives in the montane and mossy forests of the Philippines. Now, however, the centipede is in a harsh spotlight. The Philippine government says the Spanish neurologist and amateur biologist who described the species acquired his specimens illegally.

Neither the journal's editors nor its peer reviewers caught the lapse—and the journal has no policy requiring documentation that specimens have been collected with proper permits. Some editors tell *Science* that should change. Others worry about hampering research when undescribed species are vanishing fast. And all agree that journals would struggle to enforce any such rules, given the wide variation in countries' legal requirements. "There is simply no way for a journal to police this," says Maarten Christenhusz, an independent botanist and editor-in-chief of the *Zoological Journal of the Linnean Society*.

Carles Doménech of the University of Alicante in Spain had contacted Filipino collectors after seeing images of the centipede online. One, Michael Andrew Cipat, caught wild centipedes and sold them—dead and alive—to Doménech in 2016 and 2017. Cipat tells *Science* he had collecting permits and that a friend with export permits shipped the specimens. But the Philippine Department of Environment and Natural Resources says it is illegal to sell specimens to a foreign researcher who has not signed an agreement with DENR. "The Philippine government does not tolerate such illegal acts," a representative wrote to *Science*. The collectors could be imprisoned or fined under a Philippine wildlife protection law.

Doménech says he didn't know he needed permits to export the centipedes, calling himself a "newbie" who worked largely alone. After he submitted his paper describing the new species, which he named *Scolopendra paradoxa*, neither *Zootaxa* nor any of the five reviewers of his manuscript asked about permits, he says. "Now I know it [was] a mistake," he wrote in an email. "Now I capture my specimens and don't let anybody do it for me without the corresponding legal permits."

Zhi-Qiang Zhang, editor-in-chief of *Zootaxa*, who studies mites at Landcare Research in New Zealand, says that even though the journal does not impose permit requirements, individual editors may reject manuscripts that lack permits. He says the journal's editors had previously discussed whether it should instruct authors to follow permit requirements and could not agree. "Most editors had negative views about 'permits' or 'legal requirements' for specimens," Zhang says, citing opinions that such regulations curb biodiversity research and conservation (*Science*, 29 June 2018, p. 1405).

A new centipede (reproduced with permission from copyright holder) beguiled taxonomists—but the specimens had murky origins.

One reviewer of Doménech's manuscript, Carlos Martínez, a centipede taxonomist at the University of Turku's Zoological Museum, says he was "really mad" to learn about the origins of the centipede specimens. "As reviewers, we have the right to know if the specimens were illegally obtained," he says. "We have the right to refuse to review the paper." Martínez says he interviewed four of the five Filipino collectors named in the paper and confirmed the specimens were illegal. But he says reviewers can't be expected to routinely probe the legality of specimens. "We reviewers are not supposed to be the police."

Illegal specimens in research have been exposed occasionally, but journal editors disagree about the magnitude of the problem (*Science*, 1 March 2019, p. 914). One editor described it as "insignificant"; another said it is "impossible to know." They also disagree on what to do about it. Louis Deharveng, deputy editor-in-chief of *ZooKeys* and an emeritus arthropod researcher at the National Museum of Natural History in Paris, says an editorial policy on permits "is essential."

But among five respected taxonomy journals, two—the *Zoological Journal of the Linnean Society* and *Zootaxa*—do not instruct authors to abide by specimen-collecting laws. (*Science* will soon add such compliance with legal requirements to its editorial policy.)

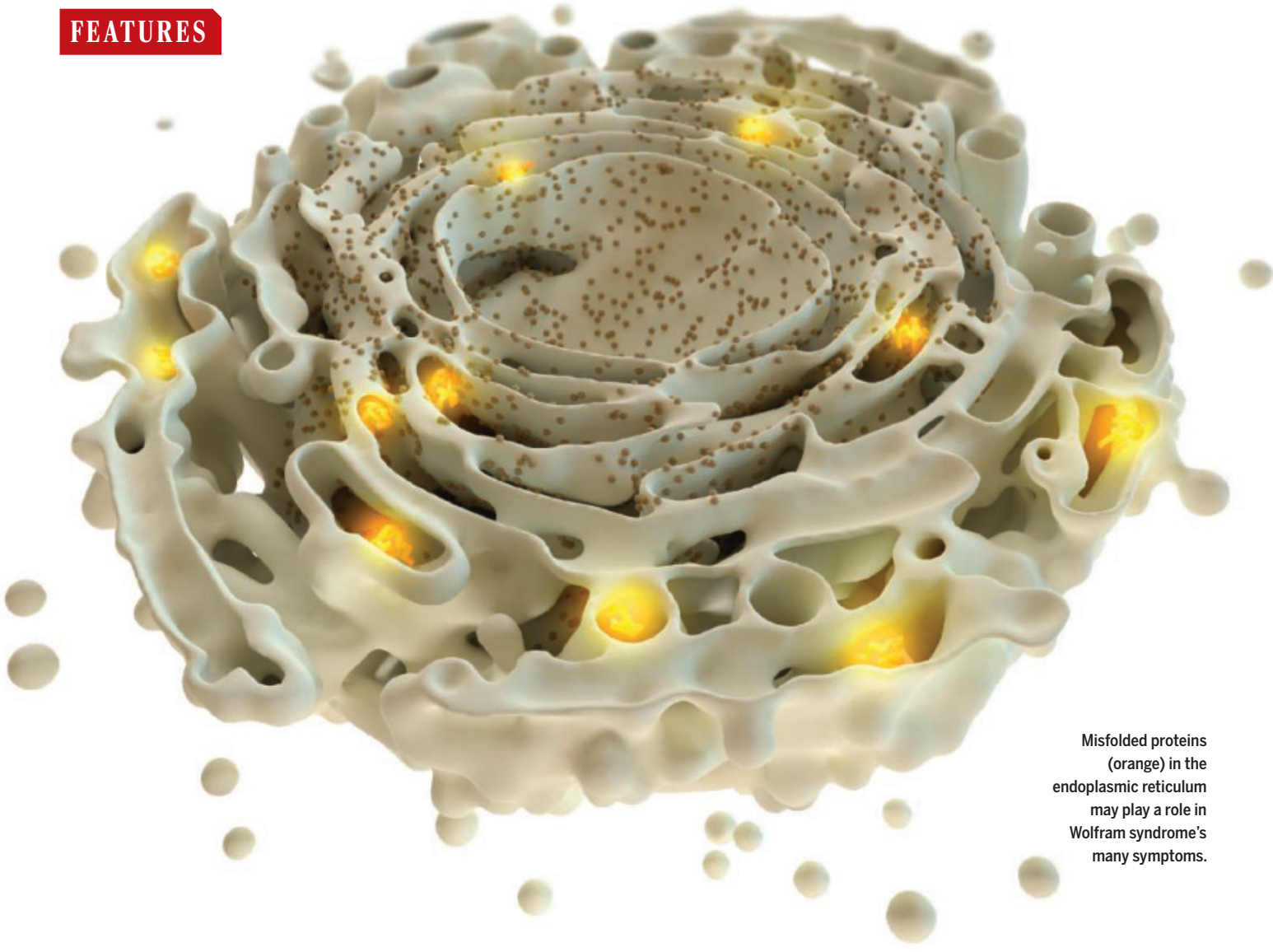
Shaun Winterton, editor-in-chief of *Systematic Entomology* and an entomologist at the California Department of Food and Agriculture, says his journal tells authors to follow the law, but adds, "If we as editors suspected that material was illegally collected, it could be difficult for us to confirm." (He notes that he is speaking on behalf of the journal, not his employer.) The complex and varying legal conditions that countries impose on research are one obstacle. A further complication is the international Nagoya Protocol, which aims to ensure "fair and equitable sharing of benefits arising from the utilization of genetic resources." The agreement may govern the import of some specimens, but whether it applies to taxonomy specimens is unclear.

Gonzalo Giribet, editor-in-chief of *Invertebrate Systematics* and a zoologist at Harvard University, says reviewers can't take on the responsibility either. "They are doing this altruistically," he says. "Journals should have clear statements about the origins of the biological materials and ethics, and the ultimate responsibility [for legality] should lie with the authors."

Clear information about legal requirements would help reviewers, editors, and researchers, says Caroline Fukushima, an arachnologist at the Finnish Museum of Natural History. In June 2020, in *Conservation Biology*, she and colleagues recommended creating a platform for countries' legal requirements for wildlife research. "We are facing habitat destruction, so we must make life easier and faster for scientists," she says. ■

Yao-Hua Law is a science journalist in Kuala Lumpur, Malaysia.

PHOTO: © MAGNOLIA PRESS; C. DOMÉNECH *ET AL.*, *ZOOTAXA* 4483(3), 401 (2018)



Misfolded proteins (orange) in the endoplasmic reticulum may play a role in Wolfram syndrome's many symptoms.

A REVEALING FLAW

A rare disease that cripples a key cellular organelle holds clues to treating more common conditions

By **Mitch Leslie**

Maureen Marshall-Doss says the first sign that her vision was deteriorating came when she misidentified the color of a dress. At a backyard get-together about 20 years ago, the Indianapolis resident pointed out an attractive yellow dress another woman was wearing. "You see that as yellow? She's wearing a pink dress," Marshall-Doss recalls her husband responding.

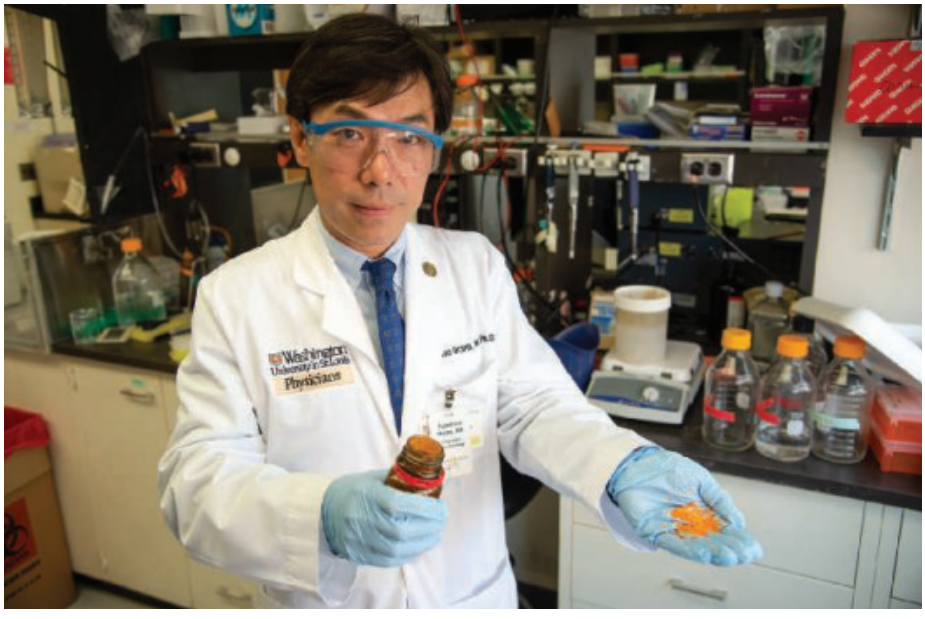
Today, Marshall-Doss is virtually blind. With help from custommade eyeglasses that magnify objects 500 times, "I can see shapes," she says. But she can no longer drive and had to quit the job she loved as a school librarian. Along with her dimming vision, she has type 1 diabetes and has lost her sense of taste and smell.

Marshall-Doss is one of 15,000 to 30,000 people around the world with Wolfram syndrome, a genetic disease. For decades, the condition remained enigmatic, untreatable, and fatal. But in the past few years, insights into its mechanism have begun to pay off, leading to the first clinical trials of drugs that might slow the illness and sparking hopes that gene therapy and the CRISPR DNA-editing tool might rectify the underlying genetic flaws. "Here is a rare disease that the basic science is telling us how to treat," says physiologist Barbara Ehrlich of the Yale School of Medicine.

The research could also aid more than the relatively few patients with Wolfram syn-

ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*

drome. Driving the disease's many symptoms is a malfunction of the endoplasmic reticulum (ER), the multichambered organelle that serves as a finishing school for many cellular proteins. Known as ER stress, the same problem helps propel far more common illnesses, including type 2 diabetes, amyotrophic lateral sclerosis (ALS), Parkinson's disease, and Alzheimer's disease. "Wolfram syndrome is the prototype of an endoplasmic reticulum disorder," says medical geneticist Fumihiko "Fumi" Urano of Washington University School of Medicine in St. Louis. Because Wolfram syndrome is "simpler," says Scott Oakes, a cell biologist and pathologist at the University of Chicago, researchers think it could illuminate the mechanisms of other ER-disrupting diseases, which affect hundreds of millions of people worldwide.



Fumihiko Urano holds dantrolene, a muscle relaxant drug he helped test as a treatment for Wolfram syndrome.

IN THE LATE 1930s, four children with diabetes were going blind, and doctors were stumped. Like many other people in the United States struggling through the Great Depression, the siblings ate a paltry diet, subsisting on potatoes, bread, oatmeal, and a little milk. But after examining three of the children, Donald Wolfram, a physician at the Mayo Clinic in Rochester, Minnesota, and an ophthalmologist colleague ruled out malnutrition as the cause of their puzzling condition. Lead poisoning and syphilis—though common enough—weren't to blame, either. When Wolfram and his partner wrote up the cases in 1938, they concluded that the symptoms could be "manifestations of an hereditary or acquired cerebral lesion."

The physicians were right that the syndrome eventually named for Wolfram is hereditary. Recessive mutations in the gene for a protein called wolframin are responsible for most cases, with glitches in a second gene causing most of the rest. However, the pair was wrong to think the defect lies only in the brain. Instead, the symptoms stem from widespread cell death. "It's definitely a disease that affects the whole body," Marshall-Doss says.

The first sign of the illness, appearing when patients are children, is usually diabetes mellitus, or faulty sugar metabolism, sparked by the demise of insulin-secreting beta cells in the pancreas. Most patients also develop the unrelated condition diabetes insipidus, in which the pituitary gland doesn't dole out enough of a hormone that helps control the body's fluid balance, causing the kidneys to produce huge amounts of urine.

Ellie White, 19, of Centennial, Colorado, who was diagnosed with Wolfram syndrome 12 years ago, says she hasn't had a full night of sleep since she was 3 years old. She gets up again and again to use the bathroom and monitor her blood sugar.

Yet she and other patients say that as disruptive as those problems are, they are not the disease's most dismaying consequence. "The biggest symptom of Wolfram syndrome that affects me the most is my vision," White says. Because neurons in the optic nerve perish, patients usually go blind within 10 years of their first visual symptoms.

Other neurons die as well. As the disease progresses, brain cells expire, and walking, breathing, and swallowing become difficult. Most people with Wolfram syndrome die before age 40, often because they can no longer breathe. At 57, Marshall-Doss is one of the oldest patients; one of her mutated genes may yield a partly functional version of wolframin, triggering a milder form of the disease, Urano says.

TWO ADVANCES have made it possible to begin to tackle those symptoms. The first was Urano's discovery nearly 20 years ago that linked Wolfram syndrome to ER stress.

The ER is where about one-third of a cell's newly made proteins fold into the correct shapes and undergo fine-tuning. Cells can develop ER stress whenever they are under duress, such as when they don't have enough oxygen or when misfolded proteins begin to pile up inside the organelle.

In test tube experiments, Urano and his colleagues were measuring the activity of genes to pinpoint which ones help alleviate ER stress. One gene that popped up encodes wolframin, which scientists had shown in 1998 was mutated in patients with Wolfram syndrome. Following up on that finding, Urano and his team determined that wolframin takes part in what's known as the unfolded protein response, which is a mechanism for coping with ER stress in which cells take steps including dialing back protein production.

Scientists think wolframin plays a key role in the unfolded protein response, though they haven't nailed down exactly how. When wolframin is impaired, cells become vulnerable to ER stress. And if they can't relieve that stress, they often self-destruct, which could explain why so many neurons and beta cells die in the disease.

Defective wolframin may harm cells in other ways. The ER tends the cell's supply of calcium, continually releasing and absorbing the ion to control the amount in the cytoplasm. Changes in calcium levels promote certain cellular activities, including the contraction of heart muscle cells and the release of neurotransmitters by neurons. And wolframin affects calcium regulation.

Beta cells genetically engineered to lack functional wolframin brim with calcium, Ehrlich and colleagues reported in July 2020 in the *Proceedings of the National Academy of Sciences*. When exposed to lots of sugar, the altered cells release less insulin and are more likely to die than healthy beta cells, the team found. The cells share that vulnerability with beta cells from patients with Wolfram syndrome. "We think that excess calcium is leading to excess cell death," Ehrlich says.

ER malfunctions could hamstring other organelles as well. The ER donates cal-

PHOTO: WASHINGTON UNIVERSITY SCHOOL OF MEDICINE IN ST. LOUIS

cium to the mitochondria, the cell's power plants, helping them generate energy. In 2018, a team led by molecular biologist Cécile Delettre and molecular and cellular biologist Benjamin Delprat, both of the French biomedical research agency INSERM, discovered that in cells from patients with Wolfram syndrome, mitochondria receive less calcium from the ER and produce less energy. Those underpowered mitochondria could spur the death of optic nerve cells, the researchers speculate.

The link between ER stress and Wolfram syndrome has been crucial for identifying potential treatments because "otherwise we would have nothing to target," Urano says. But a second development was also key, he says: the advocacy and support of patient organizations, such as the Snow Foundation and the Ellie White Foundation, headed by its namesake's mother. The foundations have stepped up with money for lab research and clinical trials when other sources, including government agencies, didn't come through.

Scientists, patients, and their advocates say Urano also deserves much of the credit. Besides treating patients, he heads the international registry of cases and has taken the lead in organizing clinical trials, screening compounds for possible use as treatments, and devising potential therapies. "Fumi is clearly the driving force," says Stephanie Snow Gebel, co-founder of the Snow Foundation, who about 10 years ago helped persuade him to forgo a plum job as department chair at a Japanese university and take over the Wolfram program at Washington University.

PATIENTS COULD SOON start to reap the benefits. In 2016, Urano and colleagues started the world's first clinical trial for the disease: a phase 1/2 study of dantrolene, an approved muscle relaxant. The molecule was a top performer when they screened 73 potential treatments for their ability to save cells with terminal ER stress. Dantrolene didn't improve vision in the 22 participants, including White, the scientists reported in an October 2020 preprint. But in some patients, beta cells appeared to be working better and releasing more insulin. The drug is safe, but Urano says it will need to be chemically tweaked to target its effects before future trials are warranted.

Researchers are pursuing other possible treatments targeting ER stress or calcium levels. In 2018, U.K. scientists launched a trial that will include 70 patients to evaluate sodium valproate, a therapy for bipolar disorder and epilepsy that, in the lab, prevents cells with faulty wolframin from dying. Last year, another compound that emerged from Urano's screens, the diabetes drug liraglutide, entered a clinical trial. Also last year, an experimental drug developed by Amylyx Pharmaceuticals for Alzheimer's disease and ALS received orphan drug designation from the U.S. Food and Drug Administration for Wolfram syndrome because it curbs ER stress. That designation offers tax breaks and other incentives, and it will get trials started sooner, Urano says.

Ehrlich and her team have a candidate of their own that they have begun to test in rodents: the drug ibudilast, which is approved in Japan to treat asthma. The researchers found it reduces calcium levels in beta cells lacking wolframin and boosts their survival and insulin output. New screening projects may reveal still more candidates.

But Urano knows that even if a treatment receives approval, it would be only "a Band-Aid for Wolfram syndrome." Hoping to develop a genetic cure, he and colleagues are introducing replacement genes into cells from patients and from mice engineered to replicate the disease. The researchers are endowing the cells with healthy copies of the gene for wolframin or the gene for a protein that reduces ER stress to determine whether they restore cellular function and reduce cell death. At INSERM, Delettre and colleagues are also evaluating whether directing a working gene into optic nerve cells can curtail vision loss in mice with faulty wolframin. The scientists are still gathering data, but early results suggest the treatment can halt the deterioration.

Urano and his collaborators have also turned to the genome editor CRISPR, deploying it to correct the gene defect in patients' stem cells and then growing them into beta cells. When the researchers transplanted the revamped cells into mice with diabetes, the animals' blood sugar returned to healthy levels, the team reported in April 2020 in *Science Translational Medicine*.

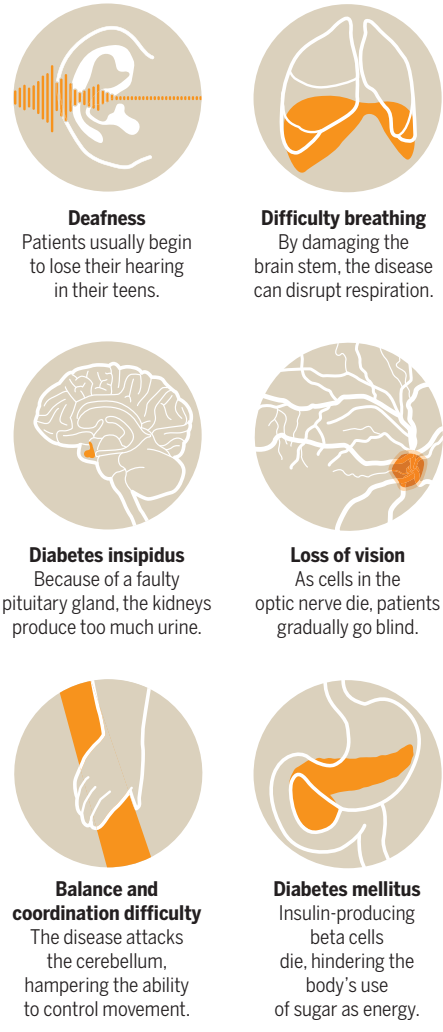
Stem cell biologist Catherine Verfaillie of KU Leuven is collaborating on the CRISPR research. But she notes that because the faulty wolframin gene affects so many tissues, researchers will have to figure out how to deliver the CRISPR components to most cells in large organs such as the brain and liver—a prospect she calls "pretty daunting." Urano agrees, predicting that CRISPR-based Wolfram therapies might take 10 to 20 years to develop. The alternative approach, gene therapy, could reach clinical trials more quickly, in 3 to 10 years, he says, because researchers have more experience with gene therapy and have created several treatments that have already been approved for other illnesses.

Because it stems from a single genetic glitch, Wolfram syndrome could also help scientists tease out the role of the ER in more complex diseases, including neurological conditions, type 2 diabetes, and cancer. The ER also falters in those diseases, causing cells to die, but the mechanism is harder to discern because they stem from myriad genetic and environmental factors. In Alzheimer's disease, for instance, neurons develop ER stress as misfolded proteins accumulate inside and outside the cells.

Besides deepening researchers' understanding of other conditions, the research on Wolfram syndrome might even deliver candidate treatments. "Everyone would be very excited if we can make advances in targeting ER stress in Wolfram syndrome," Oakes says. "It would open up the whole field to doing this in other degenerative diseases." ■

A wide-ranging toll

Mutations in the gene for wolframin disrupt the endoplasmic reticulum and lead to cell death throughout the body, causing a range of symptoms.



ILLUSTRATIONS: V. ALTOUNIAN/*SCIENCE*

TAKING ROOT

A new influx of funds could help realize Africa's Great Green Wall. But will the massive tree-planting effort learn from past failures?

By **Rachel Cernansky**

After returning home from college to northwest Cameroon in 2004, Tabi Joda felt a sense of profound loss. Trees that once bore fruit, provided medicine, and created shade had been cut down. Rich soils had turned to dust. "The land I used to know as a forest was no longer a forest," he recalls.

Joda, a business consultant, got to work, calling on what he'd learned in school and from local knowledge passed down over generations. He collected seeds, started a tree nursery, and launched an agroforestry initiative that enlisted local people in planting trees. They chose species that provided food and timber, supported livelihoods, and helped wildlife thrive. The effort soon spread to nearby communities. And Joda ultimately became a vocal advocate for an even bigger dream: the Great Green Wall, which aims to transform the lives of some 100 million people by planting a mosaic of trees, shrubs, and grasses along a corridor stretching some 8000 kilometers across Africa by 2030 (see map, p. 667).

Since the African Union first launched the Great Green Wall in 2007, the initiative has struggled to make headway. Made up of local efforts across 11 countries, it has reached just 16% of its overall goal to vegetate 150 million hectares. But last month, the project—which analysts estimate will cost at least $30 billion—got a major boost: a pledge of $14 billion in funding over the next 5 years from a coalition of international development banks and governments. The money is meant to accelerate the effort to sustain livelihoods, conserve biodiversity, and combat desertification and climate change, French President Emmanuel Macron said in announcing the pledges on 11 January.

PHOTO: DAVID ROSE/PANOS PICTURES

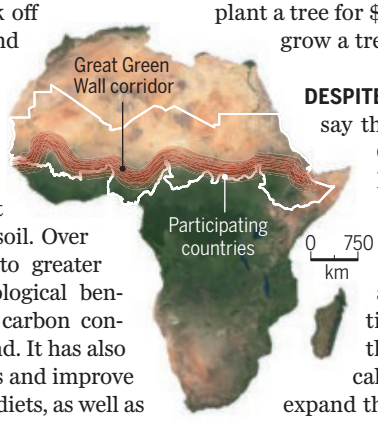
A farmer in Niger tends to his nursery. Researchers say such community-led efforts will be key to completing the Great Green Wall (map, right).

Environmental restoration and community development specialists welcomed the news. But many are also apprehensive. In recent years, research by ecologists, economists, and social scientists has shown that many forestry projects around the world have failed because they didn't adequately address fundamental social and ecological issues. Project leaders often didn't ask communities what kinds of trees they wanted, planted species in places where they didn't belong, and did little to help the saplings survive. "Tree planting is often viewed as the simple act of digging a hole," forest scientists Pedro Brancalion of the University of São Paulo, Piracicaba, and Karen Holl of the University of California, Santa Cruz, noted last year in a review of agroforestry projects in the *Journal of Applied Ecology*. "But this short-term, naïve view has resulted in large quantities of money being spent on … efforts that have failed almost entirely."

It's a problem that Joda knows well. "I have traveled the breadth of Africa and seen it everywhere," he says. "Trees are planted, but they are not taken care of and so they never grow." The question now, he and others say, is whether Great Green Wall projects fueled by the fresh burst of cash will heed those hard-learned lessons.

THERE'S LITTLE DOUBT that well-executed agroforestry initiatives can deliver hefty economic and ecological benefits. In the Brazilian Amazon, for example, Brancalion and other researchers have documented projects that have helped restore native plants, improve soil's ability to store carbon, and boost the quality of drinking water. Local residents report restoration projects have even fostered tourism.

In Africa, the adoption of one approach, called farmer-managed natural regeneration (FMNR), is credited with regreening a notable portion of the Sahel, the semiarid region south of the Sahara. The low-cost, low-tech restoration technique took off in Niger in the 1980s and has since spread to other nations. It relies on farmers to protect and nurture shrubs and trees that sprout from stumps, or grow from existing root systems and seeds in the soil. Over decades, FMNR has led to greater tree cover and other ecological benefits, such as higher soil carbon content, researchers have found. It has also helped increase crop yields and improve the diversity of household diets, as well as boost incomes.



MAP: *NATIONAL GEOGRAPHIC*, ADAPTED BY A. CUADRA AND N. DESAI/*SCIENCE*

But many efforts, particularly those not led by local communities, stumble. Newly planted trees can die of neglect when planners don't engage communities from the start in discussions about which species to plant, as well as whether residents are willing and able to provide the water, fertilizer, and protection from grazing animals that saplings need. Farmers are often busy and have their own priorities; they "will not … manage trees that they do not value," forest specialist Abayneh Derero of Ethiopia's Environment and Forest Research Institute and colleagues concluded in a study of Ethiopian tree-planting projects published last year in *Agroforestry Systems*.

A lack of community input is a leading cause of failure, according to a survey of agroforestry projects in 66 nations published last year in *Forests* by a team led by forest scientist Markus Höhl of the Thünen Institute of Forest Ecosystems. "The most important thing is that the local stakeholders are engaged," Holl says, "and that the right species are located for both the goal of the project and the site."

Other research has highlighted the problematic ecological consequences of flawed projects. For example, many have used easily available, fast-growing species that are not native to project sites, such as eucalyptus and pine. In some cases, the results have been counterproductive: The poorly adapted saplings wither, or become invasive weeds that crowd out native species and suck up scarce water.

"We do see a gap there … in terms of planting the right tree for the right purpose in the right place," says Ian Dawson, a senior scientist at the World Agroforestry Centre. Too often, he says, "The emphasis is on numbers, not quality."

Elvis Paul Tangem, who coordinates the Great Green Wall Initiative for the African Union Commission, agrees. He says promises to plant huge numbers of trees at low cost, for example at $1 per seedling, can distract from the real challenge. "You can plant a tree for $1," he says, "but you cannot grow a tree for $1."

DESPITE SUCH PROBLEMS, analysts say there are promising models. One is in Ethiopia. Called PATSPO, for Provision of Adequate Tree Seed Portfolio, it has scientists working with local institutions and communities to identify desirable tree species that are well adapted to local conditions. The goal is to expand the production, distribution, and planting of high-quality seeds of desired varieties, many of which are now in short supply, says PATSPO's lead investigator, Lars Graudal, a forester at the World Agroforestry Centre.

PATSPO is "an example that we want to see scaled up and scaled out across the whole of the Great Green Wall," Dawson says. And many of the institutions funding the wall—including the World Bank, the United Nations Food and Agriculture Organization, and the U.N. Convention to Combat Desertification—say they are committed to avoiding past missteps.

Those vows will now be tested, as new funding flows to projects along the corridor. Researchers say there's more than money at stake. The ecological consequences of failed projects could be long lasting. And projects that flop could cost small-scale farmers their investments of time and resources, and reduce their trust in future efforts.

"When you make a mistake, it's not the [project funder] that's going to pay for that error," says Jens-Peter Barnekow Lillesø, a forest researcher at the University of Copenhagen who's involved in PATSPO. "It's the tree planter. These small farmers will pay the price." ■

Rachel Cernansky is a science journalist in Denver.

POLICY FORUM

SCIENCE POLICY

Strengthen scientific integrity under the Biden administration

Accountability must extend to highest levels of leadership

By **Jacob M. Carter, Gretchen T. Goldman, Andrew A. Rosenberg, Genna Reed, Anita Desikan, Taryn MacKinney**

At his victory speech on 7 November 2020, U.S. president Joseph R. Biden described that the will of the people was in part to "marshal the forces of science." He declared that his plans to beat the novel coronavirus would be built on a bedrock of science. On 27 January, the Biden administration issued a presidential memorandum to strengthen scientific integrity and evidence-based decision-making (*1*). These are great steps to bring science back to the table, but the administration still has a lot of work ahead to improve the role of science in government decision-making. The records of abuses of the past 4 years and data from surveys provide evidence that under the Trump administration, scientists were censored, scientific information was ignored, and reports and publications were unduly suppressed—all actions that undermined the appropriate use of science in decision-making processes. This occurred despite many federal agencies already having scientific integrity policies, communications policies, and well-established science advisory systems in place to promote an appropriate role for science in agency decision-making. But from these setbacks, we can draw lessons and suggest a road map for moving forward.

SCIENTIFIC INTEGRITY OF THE PAST

The use and misuse, and proper role, of science and scientists in decision-making is a long-standing and debated issue. The process is affected by not only scientific information but also various actors and their values. It is therefore important that all actors involved in this process make transparent their postures (for example, their attitudes or perspectives) to help shield science policy decisions from being overburdened by personal preferences presented as fact. This kind of transparency also allows the public and other officials to observe and engage with the decision process.

Union of Concerned Scientists, Washington, DC, USA.
Email: jcarter@ucsusa.org

ILLUSTRATION: DAVIDE BONAZZI/SALZMANART

In the U.S. federal government, science-based decisions are informed by multiple actors, including appointed political positions (such as the heads of agencies), federal scientists, scientific advisory committees composed of academic scholars and industry experts, and the public. Having multiple levels of input into science-based decisions accommodates that science is complex and can be layered with uncertainty. Although all of these stakeholders are important, federal scientists play a key role in producing science that directly informs decisions.

At the same time, the U.S. federal government is a system with a high concentration of power in the Executive Branch, which can result in heavy influence from political appointees on science-based decisions according to their values, culture, beliefs, or interests. For example, the administrator of the U.S. Environmental Protection Agency (EPA) has a responsibility to consider scientific information provided from scientific advisory committees, stakeholder groups, EPA's own scientific staff, and public opinion, but final decisions on the laws EPA is charged with carrying out (such as The Clean Air Act, the Safe Drinking Water Act, and the Toxic Substances Control Act) often lie with the administrator. Such high levels of authority on decisions sometimes mean that political appointees inappropriately suppress, alter, or otherwise undermine the scientific process or scientific conclusions for political, financial, or ideological reasons—what we refer to as political interference in science-based decision-making.

On 17 December 2010, President Obama's science advisor, John Holdren, issued a memorandum directing federal agencies to develop and implement scientific integrity policies (*2*). The memo was intended to help agencies guard against instances of political interference in science-based decision-making (*3*). Violations of scientific integrity observed during the prior administration (for example, censorship of federal scientists, manipulation of scientific assessments, and career scientists being entirely cut out of the policy process) made clear how vulnerable science was to politicization and why the "Holdren memo" was needed (*4*).

Twenty-eight federal agencies and departments now have scientific integrity policies. One could argue that all parts of the federal government should have scientific integrity policies in place because all parts of the government deal with technical evidence in some capacity; however, existing policies cover the vast amount of federal work involving scientific data collection and research as well as use of science in regulatory processes.

Because the Holdren memo was intended to give agencies flexibility with assigning roles and responsibilities and establishing specific procedural aspects of how the policy is implemented, the scientific integrity policies vary in the comprehensiveness of their protections, and gaps remain (*5*). For example, White House officials and political appointees at federal agencies still have a great deal of protection if they choose to politicize science-based information.

Weaknesses in scientific integrity policies were made glaringly apparent by the Trump administration. The Union of Concerned Scientists has documented more than 180 instances of political interference in science, far more than 98 instances under George W. Bush and 22 under Barack Obama (*6*). The cases range from manipulating scientific data on COVID-19 to censoring scientists from using "evidence-based" and "climate change" in agency documents. In a 2018 survey, federal scientists themselves also detailed the damage to science-based decision-making, reporting widespread political interference in science-based decision-making processes (*7*).

Why did these problems occur despite scientific integrity policies at every science agency? The Trump era has revealed that scientific integrity policies vary in their strength across agencies, and enforcement mechanisms are challenging to implement when senior-level officials violate the policies. Scientific integrity policies and other infrastructure, such as related committees and scientific integrity officials, have been hugely effective at addressing issues before they become formal violations and empowering federal employees with paths for allegations, investigations, and resolution of scientific integrity issues. But recent years have brought to light policy gaps and proven that policy constraints and power dynamics make it challenging to hold agency leaders accountable when they themselves violate the policy. What do you do if science agency heads are hostile to science? Fortunately, lessons learned in the past two decades can provide guideposts. Congress can help bring science back to the forefront of policy decisions, such as through passage of the Scientific Integrity Act, which would codify many of the key provisions of scientific integrity policies. Here, we focus on initial steps the Biden administration could take to address these vulnerabilities and meet the administration's stated goal to build a stronger federal science enterprise.

SCIENTIFIC INTEGRITY OF THE FUTURE

Let scientists speak

The Biden administration should continue to follow through on its campaign promise to "listen to the scientists." This will elevate scientific voices but will not magically solve society's challenges. The process by which science informs policy decisions is a complex, value-laden social process and is affected by both scientists and nonscientist actors. Especially in what some have labeled as the "post-truth" era, the Biden administration will have to figure out how to navigate scientific and societal complexities to effectively make progress on science-based issues.

Allowing federal scientists to inform the public, however, is a good initial step. The Biden administration should grant all federal scientists the right to speak freely to the public and media about their scientific work and expertise. Agencies may choose to designate official spokespeople or ask scientists to include a disclaimer that they are not speaking on behalf of the agency, but federal scientists should not need permission to answer a reporter call or post on their personal social media page. Some argue that such freedoms could lead to the spread of misinformation from rogue scientists, but evidence of such concerns has not materialized. The National Oceanic and Atmospheric Administration

(NOAA), for example, has for years had a permissive media policy without a major issue. NOAA allows its scientists to speak with reporters without preapproval from political staff or public affairs. By contrast, other agencies require scientists to get explicit advanced permission to speak with journalists. Additionally, any dissenting views aired by individuals speaking out could be addressed with transparent decision-making and by policies that explicitly allow for expression of differing scientific opinions by agency scientists, as discussed below.

The novel coronavirus pandemic has made clear the need for timely scientific communication from authoritative government sources and the consequences for public health when science is silenced. Fewer U.S. Centers for Disease Control and Prevention press briefings, Dr. Anthony Fauci being blocked from interviews with major television networks, scientific guidance documents with political manipulation, and blatant misinformation from the U.S. president all underscore the vital importance of access to federal science and scientists, especially in crises. (*8*).

Following its scientific integrity policy review, the White House Office of Science and Technology Policy (OSTP) should direct science agencies to modify existing policies to explicitly grant technical experts the freedom to speak to the media and public without needing approval from political leadership and public affairs staff. And at agencies within larger departments, such as NOAA within the Department of Commerce, that mandate must apply up the chain of command and be adhered to by all political staff, even the heads of agencies and departments (*9*).

Make scientific integrity more powerful

After its review of scientific integrity policies, OSTP should protect government science and scientists by directing federal agencies to strengthen their scientific integrity infrastructure to provide clear and consistent scientific integrity policies at every science agency. Currently, only some agencies have strong scientific integrity infrastructure in place, including a full-time scientific integrity official, a scientific integrity board to help adjudicate potential violations, and scientific integrity liaisons dispersed throughout the agency to help other employees understand the policy. The Trump administration's record of abuses to science, survey responses from thousands of federal scientists, and inconsistencies across policies between agencies provide important data about what works and what does not when it comes to protecting science and scientists from politicization (*10*). A challenge that stands out is agencies' inability to enforce and hold accountable senior-level political officials who violate policies. Although many federal agencies have scientific integrity officers, some do not. And even where scientific integrity officers exist, they often do not have power to hold political officials accountable for missteps.

The Biden administration should follow through on its mandate that all agencies conducting or dealing with scientific work have a full-time scientific integrity officer dedicated to training federal employees, investigating scientific integrity allegations, and enforcing scientific integrity policies. Further, the Biden administration should require that all scientific integrity policies have explicit language to protect scientists and their work from interference from political officials, external stakeholders, or anyone else. Such protections can be tackled at both ends of the problem: by better protecting scientists involved through transparency measures, process changes, and opportunities for reporting and by holding wrongdoers accountable, including and especially senior political officials. The latter point is crucial because many Trump-era issues stemmed from high-level political appointees. The Biden administration must ensure that scientific integrity officers can hold these senior leaders accountable for violations of scientific integrity. This could occur by granting agency scientific integrity officers the authority to work closely with agency inspectors general offices, which currently do not generally coordinate with scientific integrity offices. This would empower scientific integrity officers and increase opportunities for holding senior officials accountable because inspectors general offices operate more independently from agency leadership than do scientific integrity offices in existing structures.

Let scientists dissent

The Trump administration ignored the recommendations of technical experts in numerous cases in which long-standing processes require the input of those experts. This was the case even where the law is clear that science must be the sole basis for decisions (such as the listing and delisting of endangered species). For example, in 2020 the U.S. Fish and Wildlife Service (FWS) moved forward and delisted the gray wolf across the United States, ignoring every scientist who peer reviewed FWS's delisting proposal. Those scientists were left with little recourse to ensure that the administrative record reflected the best available science rather than being restricted to the scientific evidence that was politically expedient. The scientists could provide public comment on the delisting decision, although the Trump administration was notorious for also ignoring such comments. Scientists now have one way to move forward to ensure that this decision reflects the best available science:

ILLUSTRATION: DAVIDE BONAZZI/SALZMANART

fight the decision through the court system.

One tool that the Biden administration could put in place is Dissenting Scientific Opinion policies at federal science agencies. Such policies allow scientists who were directly involved in a policy decision to write a dissenting scientific opinion for the public record. These opinions can prove useful in subsequent litigation and could discourage political leadership from ignoring scientists in the first place. Such protections will also be helpful in the event that political appointees burrow into career positions with the intention to continue hostility toward science from within the government. For example, an EPA Dissenting Scientific Opinion filed by a senior career official in September 2020 revealed the extent to which science and scientists were ignored in the development of an EPA rule that restricts the agency's use of scientific studies relying on nonpublic data. Such information provides new insight on internal scientific integrity challenges and may be an important data point in anticipated litigation around the rule. Only three agencies (the EPA, U.S. Food and Drug Administration, and U.S. Department of Energy) have Dissenting Scientific Opinion policies (*4*).

Hire more scientists

The Biden administration must address the drain of scientific experts from federal agencies. An analysis from *The Washington Post* estimated that 1600 federal scientists left their post under the Trump administration (*11*). According to *The Washington Post*'s analysis, this loss of scientists represents a 1.5% decrease, whereas an 8% increase in scientific staffing was seen during the same time under the Obama administration. Another analysis from the Environmental Data and Governance Initiative reported that the EPA's total workforce was down by 16% in 2018 compared with 2013, and 21% in 2018 compared with an all-time peak in staffing at the agency during 2000 (*12*). That these losses present a problem is corroborated by our survey of federal scientists, in which capacity constraints were deemed a key barrier to making science-based decisions at the agencies (*6*).

The White House should prioritize building scientific capacity across the government, both in its science and technology budget and hiring plans. One way the administration should boost science capacity and recruit talented early-career STEM (science, technology, engineering, and mathematics) experts is by focusing on science fellowships. Hiring of fellows can help to bolster scientific capacity faster than the normal government hiring processes. OSTP should reestablish the STEM track for the Presidential Management Fellowships or expand fellowship awards in federal programs such as those in Sea Grant. The Trump administration's hostility toward science may have long-lasting impacts that deter qualified job candidates from applying. Therefore, the administration should consider additional incentives, such as skills training, fast-track promotions based on merit, and measures aimed at boosting applications from scientists of color and others from underrepresented backgrounds.

Revive independent science advice

The Biden administration must restore independent science advice to federal agencies. Federal science agencies depend on external experts who provide guidance through their service on federal advisory committees. During the past 4 years, many advisory committees have been inactive, and some have been disbanded or stacked with appointees who are unqualified or have conflicts of interest (*13*). Then-President Trump also issued an executive order that required each federal agency to arbitrarily cut the number of its advisory committees by at least one-third, with a limit of 350 committees across the government. President Biden should rescind this executive order immediately so that advisory committees can continue to be formed or disbanded as decided by agencies rather than political mandate (*14*).

President Biden should issue an executive order that improves the integrity and transparency of federal advisory committees. Currently, management of conflicts of interest around advisory committees is not sufficiently transparent and is inconsistently enforced. A Biden executive order should direct the Office of Government Ethics to update its guidelines on conflicts of interest for advisory committee members and federal agencies to provide explicit and transparent guidance describing how they set up advisory committees, vet nominees for conflicts of interest, and decide on member composition (*15*). The order should require agencies to extend disclosure requirements so that committee members provide information on past employers and research funding. The order should also ask agencies to make public the roster of candidates for membership, along with details on which nominees will be representatives of a particular stakeholder group and thus not subject to government ethics requirements, and request comments regarding candidates' potential conflicts of interest. Appointments to advisory committees should also include a more diverse group of independent scientists, including scientists of color and those earlier in their careers.

CONCLUSION

The previous 4 years were unprecedented in terms of obstacles and challenges surrounding science-based decision-making (*2*). The scientific community and the Biden administration must not write this period off as unique. The Biden administration has a lot of work to do to rebuild "better" the scientific basis of public policy decision-making. Fortunately, we have learned lessons and now know how to strengthen science-based processes to make them more resilient to such unprecedented attacks. The administration's 27 January memorandum establishes a working group to review scientific integrity policies and their implementation, reestablishes science advisory committees, and appoints chief science officers and scientific integrity officials at federal agencies to ensure that the government and its research programs are well grounded in science and technology and that decision-making is conducted with integrity. These are all important steps. President Biden said it is the will of the people to marshal the forces of science, and the people should hold him accountable to his word. ■

REFERENCES AND NOTES

1. The White House, "Fact Sheet: President Biden takes executive actions to tackle the climate crisis at home and abroad, create jobs, and restore scientific integrity across federal government" (27 January 2021).
2. J. P. Holdren, "Memorandum for the heads of executive departments and agencies" (Office of Science and Technology Policy, 17 December 2010).
3. E. Berman, J. M. Carter, *Journal of Science Policy and Governance* **13**, 1 (2018).
4. G. T. Goldman *et al.*, *Science* **355**, 696 (2017).
5. T. MacKinney, J. M. Carter, G. Reed, G. Goldman, "Strengthening scientific integrity at federal agencies" (Union of Concerned Scientists, 2020).
6. Union of Concerned Scientists, "Attacks on science" (20 January 2017); www.ucsusa.org/resources/attacks-on-science.
7. G. T. Goldman, J. M. Carter, Y. Wang, J. M. Larson, *PLOS ONE* **15**, e0231929 (2020).
8. A. Desikan, T. MacKinney, G. Goldman, "Let the scientists speak: How CDC experts have been sidelined during the COVID-19 pandemic" (Union of Concerned Scientists, 2020).
9. T. Allen, S. Agrawal, K. Husbands Fealing, E. Robinson, "An independent assessment of allegations of scientific misconduct filed under the National Oceanic and Atmospheric Administration scientific integrity policy" (National Academy of Public Administration, 2020); https://nrc.noaa.gov/Portals/0/SIC/NOAA%20Final%20Report_scanned_061220.pdf?ver=2020-06-15-074029-673.
10. M. Brym *et al.*, "Further efforts needed to uphold scientific integrity at EPA," report no. 20-P-0173 (EPA Office of inspector General, 2020); https://www.epa.gov/sites/production/files/2020-05/documents/_epaoig_20200520-20-p-0173.pdf.
11. A. Gowen, J. Eilperin, B. Guarino, A. Ba Tran, "Science ranks grow thin in Trump administration," *The Washington Post*, 23 January 2020.
12. L. Frederickson *et al.*, "A sheep in the closet: the erosion of environmental enforcement at the EPA" (Environmental Data Governance Initiative, 2019).
13. G. Reed, S. Shulman, P. Hansel, G. Goldman, "Abandoning science advice" (Union of Concerned Scientists, 2018).
14. G. T. Goldman, *Nature* **570**, 417 (2019).
15. M. L. Kelly, G. Reed, G. T. Goldman, J. M. Carter, *Environ. Law Rev.* **51**, 1 (2021).

10.1126/science.abg0533

Humans are exposed to different types of fibers and particles, including microplastics; the potential health effects of microplastics are largely unknown.

TOXICOLOGY

Microplastics and human health

Knowledge gaps should be addressed to ascertain the health risks of microplastics

By **A. Dick Vethaak**[1,2] and **Juliette Legler**[3]

The ubiquity of microplastics (plastic particles <5 mm, including nano-sized plastics <1 μm) in the global biosphere raises increasing concerns about their implications for human health (*1*–*3*). Recent evidence indicates that humans constantly inhale and ingest microplastics; however, whether these contaminants pose a substantial risk to human health is far from understood. The lack of crucial data on exposure and hazard represents key knowledge gaps that need to be addressed to move forward.

Microplastics are created by the weathering and breakdown of plastic objects, car tires, clothing, paint coatings, and leakage of preproduction pellets and powders. They may also be intentionally added to daily-life products (e.g., cosmetics and abrasive cleaners) (*1*, *2*). Microplastics represent a highly diverse class of contaminants spanning five orders of magnitude in size, are of various shapes (e.g., spheres, fragments, fibers), and have a complex composition, including polymeric materials and mixtures of chemicals (residual monomers, additives, and hydrophobic environmental contaminants) (*4*–*6*). Furthermore, biofilms growing on microplastics may be a source of harmful microorganisms (*2*, *7*). Their ubiquity in the environment raises serious concerns about their effects on wildlife and ecosystems (*1*), but what are their effects on human health?

Microplastics may enter the human body through both inhalation and ingestion, potentially causing health effects (see the figure). A parallel can be drawn with particulate air pollution: Small particles (<2.5 μm), such as those from diesel exhaust, are capable of crossing cell membranes and triggering oxidative stress and inflammation, and have been linked with increased risk of death from cardiovascular and respiratory diseases or lung cancer (*3*). This parallel provides ample incentive to gather more information on the potential risk of microplastic particles.

A major issue when determining the risks of microplastics to human health is the lack of information on human exposure. Adequate analytical tools to sample, isolate, detect, quantify, and characterize small microplastics (<10 μm), especially nano-sized plastic particles, are urgently needed. External exposure estimates comprise limited and highly variable data of mainly large particles (>10 to 50 μm), with poor standardization and quality control measures, hampering a comprehensive exposure assessment (*1*, *8*). Nevertheless, a growing body of evidence suggests widespread exposure to microplastics from various foods, drinking water, and air (*1*, *9*, *10*).

Reported concentrations of microplastics in tap and bottled water vary between 0 and 10^4 particles/liter, with generally greater particle counts for small-sized microplastics (*8*). The first atmospheric measurements of larger-sized, predominantly fibrous microplastics indicate that plastic particles are a relevant component of fine dust, with, for example, deposition rates in central London ranging between 575 and 1008 microplastics per square meter per day (*9*). Increased exposure through indoor air, direct swallowing of house dust or dust settling on food (*10*), and direct exposure to particles released from plastic food containers or bottles, such as polypropylene infant feeding bottles (*11*), are of special concern. Larger microplastics are likely excreted through feces, or after deposition in the respiratory tract or lungs through mucociliary clearance into the gut (*1*, *4*). Given the methodological limitations and measurement bias toward larger particles, existing analyses probably underestimate human external exposure and generally do not include the fraction of smaller-sized particles <10 μm, which are likely more relevant to toxicity (*1*, *12*). Notably, internal exposure measurements of plastic particles in human body fluids and tissues are still in their infancy.

A better understanding of the ability of microplastics to cross the epithelial barriers of the airway, gastrointestinal tract, and skin is needed to reduce the current uncertainty in the human risk assessment of microplastics. Limited in vitro and in vivo data suggest that only small fractions of administered microplastics are capable of crossing epithelial barriers of lungs and intestines, with specific uptake profiles and generally increasing uptake efficiency with decreasing particle size (*2*). This low proportion of particle uptake is not necessarily unimportant when considering life-long exposure and because of possible accumulation in tissues and organs. Studies with human cells in culture, and in rodents and aquatic species indicate translocation of microplastics <10 μm from

[1]Department of Environment and Health, Vrije Universiteit Amsterdam, Amsterdam, Netherlands. [2]Department of Marine and Coastal Systems, Deltares, P.O. Box 177, 2600 MH, Delft, Netherlands. [3]Institute for Risk Assessment Sciences, Department of Population Health Sciences, Faculty of Veterinary Medicine, Utrecht University, Utrecht, Netherlands. Email: dick.vethaak@deltares.nl

PHOTO: DICK VETHAAK

the gut cavity to the lymph and circulatory systems, causing systemic exposure and accumulation in tissues including liver, kidney, and brain (*12*). Although the smallest particles (<0.1 μm) may be capable of accessing all organs, crossing cell membranes (*12*), the placenta (*13*), and also the brain (*14*), major knowledge gaps regarding absorption, distribution, metabolism, and excretion (ADME) still exist. Whether there are dose-dependent effects of microplastics in humans also remains unknown.

Once in contact with epithelial linings in the lung or intestine, or after being internalized, microplastics may cause physical, chemical, and microbiological toxicity, which could also act cumulatively. Several in vitro (i.e., human cell culture) and in vivo rodent studies indicate the potential of inhaled or ingested microplastics to cause a variety of biological effects, including physical (particle) toxicity, leading to oxidative stress, secretion of cytokines, cellular damage, inflammatory and immune reactions, and DNA damage, as well as neurotoxic and metabolic effects (*12*). The observed effects are usually triggered at high exposure concentrations of microplastics, and these experiments use a limited number of pristine, commercially available particle types, which are inconsistent with those encountered in the environment. Furthermore, chemical contamination of these test particles cannot always be excluded. Similar to the effects observed in ambient particle exposure studies, epidemiological studies have reported lung injuries, including inflammation, fibrosis, and allergy, among workers in the plastic and textile industry who are exposed to high amounts of plastic fibrous dust (*4*).

Chemical toxicity may be caused by microplastics acting as vectors to transfer exogenous hazardous chemicals, proteins, and toxins present in or on the particles into the body (*1*, *5*, *6*). However, this "Trojan horse" effect is understudied with little knowledge of the role of nanosized microplastics, which are more effective at crossing biological membranes and have increased surface area for chemical reactivity than larger-sized microplastics. Some studies suggest that aquatic microplastics may act as vectors of microbiological toxicity, carrying biofilm-associated opportunistic bacterial pathogens and antibiotic resistance genes that may interact with gut microbiota (*15*). In-depth research on the stability of microbial contaminants within the human body is required to further clarify this. The possibility that microplastics act as carriers of other potential pathogens, such as fungi and viruses, also deserves attention. More research is urgently needed to fully understand the potential toxicity, underlying mechanisms, and long-term effects of microplastics under real life conditions.

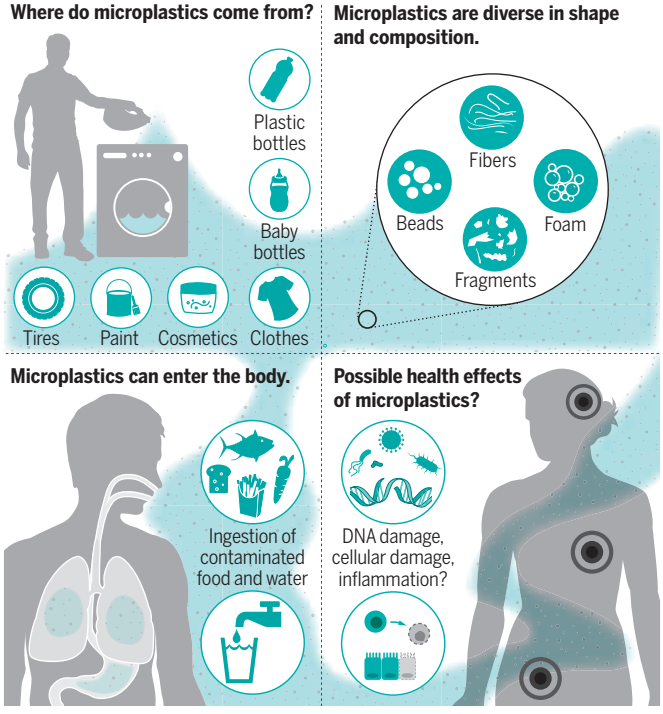
An additional intriguing, yet understudied, but potentially hazardous property of microplastics is the presence of an eco- or biocorona, i.e., biomolecules and other substances on the surface of the plastic particle, which may influence particle uptake, fate, and effects (*7*, *13*). The heterogeneous composition of the eco- or biocorona is determined by the physicochemical properties of the microplastic and complex particle interactions with both the environment (comprising natural matter, biomolecules, chemical contaminants, and microorganisms) and the human body (adsorbed lipid and proteins) (*6*, *7*, *13*). Before crossing the epithelial barriers in the lung and intestine, microplastics are trapped in the mucus layer covering the cells, whereas ingested particles have to pass through acidic conditions in the stomach and intestinal lumen. The role of the changing composition of the eco- or biocorona acquired by microparticles, from the outside to the inside of the body, across tissue barriers, and the underlying mechanisms in mediating uptake and toxicity are poorly understood and deserve more study.

The major knowledge gaps outlined above prevent a thorough assessment of the health risks of microplastic exposure for humans. However, ongoing research can help progress our understanding. Technological advancements for particle analysis of microplastics, especially nano-sized microplastics, in relevant human body fluids and tissues are anticipated within the next few years. In general, microplastics are thought to affect human health as a function of their properties, such as chemical composition, size, shape, and surface charge (*1*, *4*, *6*). Improved characterization of test particles and research that reflects real environments are needed, for example, by examining environmentally weathered flakes and fibers in addition to the pristine polystyrene spheres often used now.

Additionally, given their physicochemical similarities (e.g., poor solubility, high persistence, wide size range, and complex nature), there are important parallels between microplastics and much-studied nanomaterials and particulate air pollution. Therefore, research on plastic particles may build on existing knowledge and lessons learned from research on nanomaterials; and predictions from kinetics, toxicological, and epidemiological data associated with particulate air pollution, notably the effects of exposure to mineral dust particles and soot particles from combustion sources. To assess the extent to which any effects found are specific to microplastics, it would be useful to compare the effects of different standard polymer reference materials with well-studied positive controls, such as soot particles, engineered nonplastic nanomaterials, silica particles, and natural polymers. Furthermore, important knowledge on environmental microplastics can be mined from the use of polymeric particles in drug delivery systems and particles abraded from plastic prosthetic implants (*2*).

What are the effects of microplastics in humans?

Microplastics (plastic particles <5 mm) can come from the breakdown of plastic objects, car tires, and clothing, but also from their use in cosmetics and other applications. They have diverse shapes and encompass a suite of chemical and biological constituents. Microplastics can enter the human body through ingestion and inhalation where they may be taken up in various organs and might affect health, for example, by damaging cells or inducing inflammatory and immune reactions.



GRAPHIC: N. CARY/*SCIENCE*

Every day, humans are exposed to a wide range of natural and manufactured particles, with particulate air pollution recognized as one of the world's leading environmental risk factors for disease. It is crucial to understand the role of microplastics and their contribution to total ambient particle exposure to evaluate their potential contribution to global disease burdens. Owing to their persistence, wide size range, and complex nature, microplastics may exhibit distinct particle properties with a different and broader toxicity profile compared to those of other ambient particles. To date, pressing microplastic-related health issues such as internal exposure; ADME processes, including the effect of the eco- or biocorona; interaction with the immune system; whether nanosized plastics can affect the placenta, fetus, and brain; and how environmental microplastics differ from other ambient natural and engineered nanoparticles are largely unexplored. Pioneering interdisciplinary research programs (such as Microplastics & Health in the Netherlands and the European Union Horizon 2020 research program) are beginning to resolve some of these issues, which are fundamental to innovation, evidence-based policy-making, and strategies to improve risk management. Multidisciplinary research efforts, involving scientists from environmental and medical sectors as well as polymer scientists, are needed to tackle this potential health hazard. Comprehensive risk assessment is still far away, but the major research gaps should be addressed now to support timely decision-making on health policies and mitigation strategies. ■

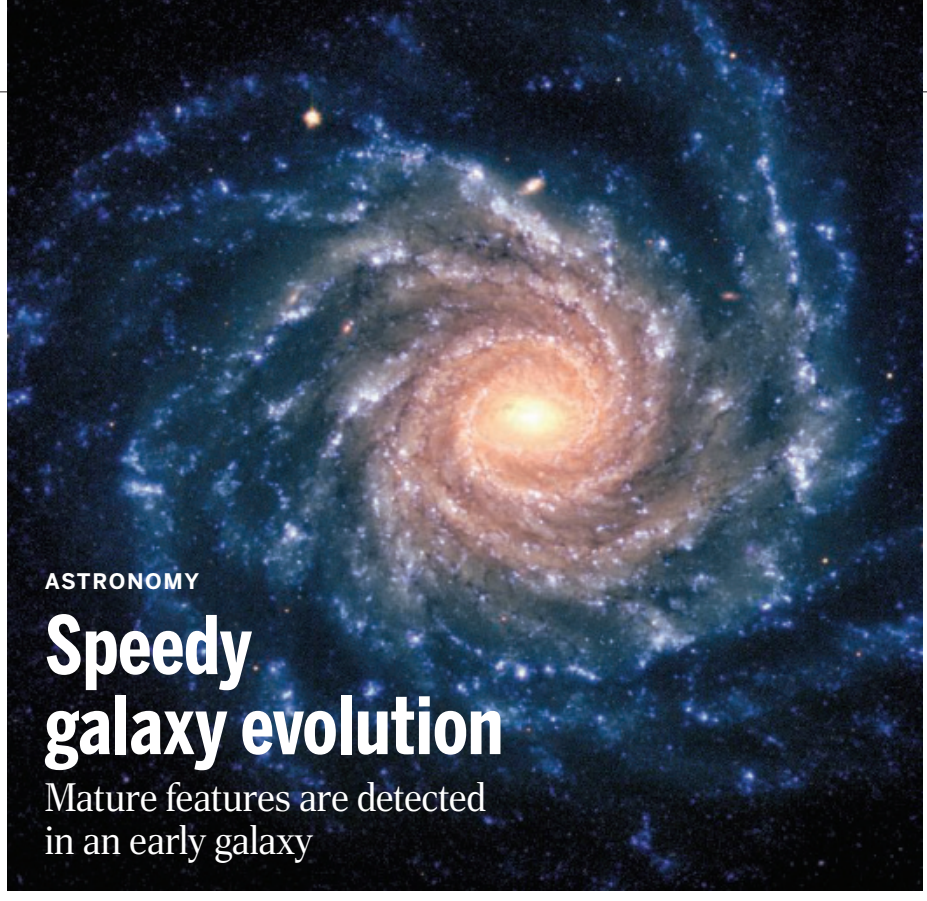
REFERENCES AND NOTES

1. SAPEA, "A scientific perspective on microplastics in nature and society" (2019); https://doi.org/10.26356/microplastics.
2. S. L. Wright, F. J. Kelly, *Environ. Sci. Technol.* **51**, 6634 (2017).
3. F. Kelly, J. C. Fussell, *Philos. Trans. R. Soc. A* **378**, 20190322 (2020).
4. C. M. Rochman *et al.*, *Environ. Toxicol. Chem.* **38**, 703 (2019).
5. A. A. Koelmans *et al.*, *Environ. Toxicol. Chem.* **50**, 3315 (2016).
6. J. Shi, D. Wu, Y. Su, B. Xie, *Sci. Total Environ.* 10.1016/j.scitotenv.2020.142775 (2020).
7. F. Ribeiro *et al.*, *TrAC-Trend Anal. Chem.* **111**, 139 (2019).
8. A. A. Koelmans *et al.*, *Water Res.* **155**, 410 (2019).
9. S. L. Wright *et al.*, *Environ. Int.* **136**, 105411 (2020).
10. A. I. Catarino, V. Macchia, W. G. Sanderson, R. C. Thompson, T. B. Henry, *Environ. Pollut.* **237**, 675 (2018).
11. D. Li *et al.*, *Nat. Food* **1**, 746 (2020).
12. C. Q. Y. Yong, S. Valiyaveetill, B. L. Tang, *Int. J. Environ. Res. Public Health* **17**, 1509 (2020).
13. M. M. Gruber *et al.*, *J. Nanobiotechnology* **18**, 128 (2020).
14. M. Prüst, J. Meijer, R. H. S. Westerink, *Part. Fibre Toxicol.* **17**, 24 (2020).
15. L. Lu *et al.*, *Sci. Total Environ.* **667**, 94 (2019).

ACKNOWLEDGMENTS

The authors receive financial support from The Netherlands Organisation for Health Research and Development (ZonMw). We thank F. Pierik, program leader ZonMw Microplastics & Health program, and all collaborators in this program for helpful discussions.

10.1126/science.abe5041



ASTRONOMY

Speedy galaxy evolution

Mature features are detected in an early galaxy

An image of galaxy NGC 1232 (September 1998) shows spiral arms and central red bulge, the hallmarks of maturity.

PHOTO: ESO

By **Julie Wardlow**

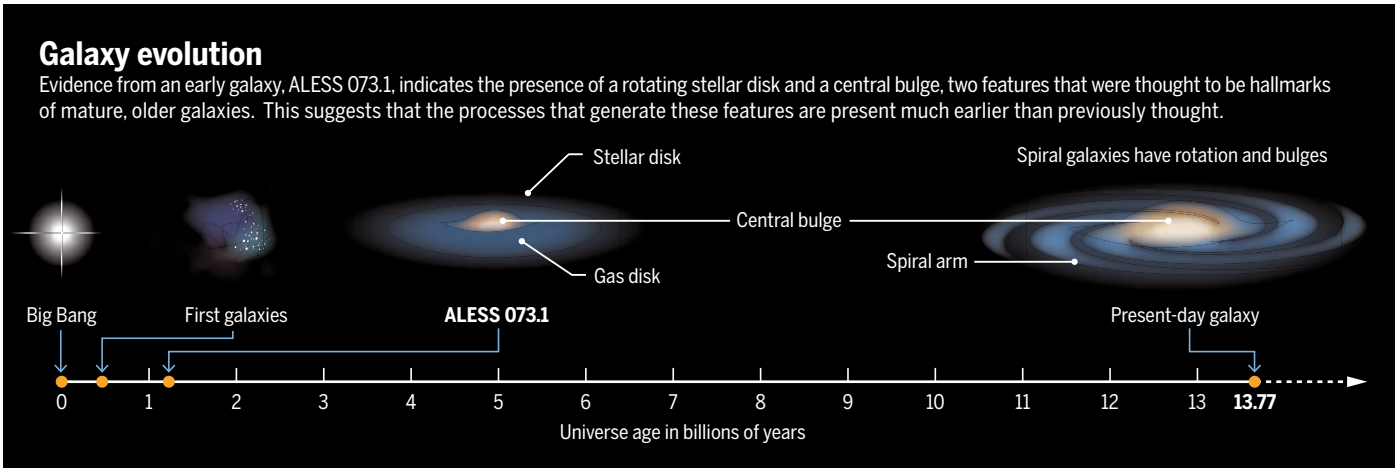
The processes that transformed small, turbulent, relatively unstructured protogalaxies into rotating spiral or giant elliptical galaxies are not well understood. Most galaxies are expected to go through a spiral-like phase, maturing into an elliptical structure. Many local spiral galaxies have a classic rotating disk of young stars as well as a "bulge" of older red stars at their centers; these two features are considered to be signatures of galaxies that have evolved from their original primordial forms. Unfortunately, these features are challenging to directly detect, particularly in the very distant (i.e., early) Universe. On page 713 of this issue, Lelli *et al.* (*1*) report a galaxy that had evolved features (both a disk and a bulge) when only 1.2 billion years had elapsed since the Big Bang (~12.5 billion years ago) (see the figure). This finding suggests that the processes that generate the key features of a mature galaxy arose more rapidly than has been thought.

Theoretical simulations suggest that primordial galaxies started to form shortly after the Big Bang and are expected to merge together through gravity. They form increasingly large but somewhat chaotic structures (*2*) that may have some rotation but must also eventually form bulges of old stars. The mechanism of bulge formation requires additional observations to constrain the model.

There are two main theories for forming a concentrated population of old stars: galaxy mergers and internal instabilities. Mergers cause huge gravitational disturbances that can compress cold gas (the fuel for new stars), triggering huge but short-lived bursts of star formation. Similarly, the distant Universe was a chaotic place, and it is thought that early galaxies were often gravitationally unstable such that internal stochasticity could have triggered similar dynamical disturbances and compressions of gas leading to an intense starburst. In both scenarios, after the gravitational disturbance, the cold gas fuel for star formation is rapidly exhausted, leading to an aging of the central stellar population along with a relaxation of the dynamics. Because both scenarios lead to similar theoretical outcomes, the question of whether galaxy bulges form from stochastic internal processes or as the result of galaxy mergers requires observations to answer the question of how bulges emerge.

Although it is clear that distant galaxies tend to have more irregular and "clumpy" morphologies than local galaxies, these clumps can be interpreted as evidence for both massive star-forming regions (i.e., sec-

Physics Department, Lancaster University, Lancaster LA1 4YB, UK. Email: j.wardlow@lancaster.ac.uk



ular processes) and mergers, depending on their prevalence and mass (*3*, *4*). Additional data are required to untangle the various processes driving the evolution of distant star-forming galaxies.

Instruments that can map the motion of galaxies' gases and stars in two dimensions have recently gained sensitivity and resolution, enabling astronomers to measure the rotation of galaxies farther into the distant Universe. Multiple tracers now allow researchers to separately pinpoint stars, ionized and molecular gases, and dust in distant galaxies (*5*). The spatial and dynamical differences between components can point to different formation mechanisms.

Star formation activity in the Universe was at its peak 10 to 11 billion years ago, and large kinematic surveys of massive star-forming galaxies from that time show that most had morphologies consisting of a disk and bulge (*6*). Their internal kinematics were dominated by rotation, as expected for spiral galaxies, albeit with a mass dependence and with the disks typically harboring more turbulence than their local counterparts (*7*, *8*). Thus, it seems that systems with similarities to today's spiral galaxy population had already evolved 10 to 11 billion years ago. Thus, to observationally identify even earlier stages in galaxy evolution, the measurement of even more distant systems is required.

For detected light that was emitted when the Universe was less than 2 billion years old, the bright emission lines that are used to trace the dynamics of ionized gas in nearer galaxies are redshifted out of the wavelength range of ground-based near-infrared telescopes. Furthermore, these emissions are blocked from detection by Earth's atmosphere. For these most distant galaxies, another approach is needed, and far-infrared observations of carbon, carbon monoxide, oxygen, or nitrogen are fast becoming the key method for tracing their gas (*9*). Such observations have provided evidence for rotation in massive galaxies at these epochs (*10, 11*), suggesting that gas disks formed early in the Universe's history.

Ideally, direct observations of gases, stars, and dust would be combined, providing complementary information about a galaxy's overall structure. To date, few systems have deep, high-resolution observations in multiple tracers, so observations of tracers in different galaxies are used instead to infer a coherent picture. For example, in the most actively star-forming distant galaxies, star formation tends to be substantially more compact than the existing stars, which could indicate growing bulges (*12*). These most active galaxies are thought to represent some of the most massive galaxies at each epoch, and observing them provides an opportunity to explore the limits of galaxy evolution. A handful of these systems now have resolved observations of their cold gas, which typically seems to be more spatially extended than the ongoing star formation. These massive galaxies are also rotating, and substantial fractions of the overall galaxy mass are in the gaseous component (*10, 13*).

Lelli *et al.* have brought together many of these different tracers of galaxy evolution for the galaxy ALESS 073.1, which was an extremely active galaxy when the light that we observe was emitted, only 1.2 billion years after the Big Bang. The authors use a far-infrared map of carbon emission to show that the cold gas in ALESS 073.1 is rotating and combines this information with imaging of the compact dust component to measure the distribution of different mass components. The diversity and quality of the data for a galaxy so far away are rare, and the authors show that ALESS 073.1 not only comprises a rotating cold gas disk, a stellar disk, and a dark matter halo, but also must have already established a substantial central bulge as well. Because ALESS 073.1 is seen only 1.2 billion years after the Big Bang, this study shows that the galaxy evolution mechanisms that build bulges must act on relatively short time scales, and that if a major merger was involved in the evolution of ALESS 073.1, it must have taken place less than 1 billion years after the Big Bang. If this merger scenario is the currect mechanism, then Lelli *et al.*'s observations of ALESS 073.1 also require that multiple massive galaxies must have already formed by this epoch. Additionally, ALESS 073.1 contains an active galactic nucleus (AGN), indicative of a growing central supermassive black hole, which may be associated with the rapid growth of its bulge.

The next step in understanding the buildup of galaxies so soon after the Big Bang is to determine whether ALESS 073.1 is a rarity or whether the presence of bulges is common at this epoch. It also remains unknown whether all such galaxies with early formed bulges contain AGN and/or are undergoing rapid bursts of star formation as seen in ALESS 073.1. Such studies should reveal important constraining factors and the time scales of typical galaxy evolution processes and will thus provide clues as to the role of galaxy mergers in the growth of very early galaxies. ■

GRAPHIC: C. BICKEL/*SCIENCE*

REFERENCES AND NOTES

1. F. Lelli *et al.*, *Science* **371**, 713 (2021).
2. J. Schaye *et al.*, *Mon. Not. R. Astron. Soc.* **446**, 521 (2015).
3. N. Mandelker *et al.*, *Mon. Not. R. Astron. Soc.* **464**, 635 (2017).
4. N. M. Forster Schreiber, S. Wuyts, *Annu. Rev. Astron. Astrophys.* **58**, 661 (2020).
5. C.-C. Chen *et al.*, *Astrophys. J.* **846**, 108 (2017).
6. S. Wuyts *et al.*, *Astrophys. J.* **742**, 96 (2011).
7. J. P. Stott *et al.*, *Mon. Not. R. Astron. Soc.* **457**, 1888 (2016).
8. E. Wisnioski *et al.*, *Astrophys. J.* **886**, 124 (2019).
9. J. A. Hodge, E. da Cunha, *R. Soc. Open Sci.* **7**, 200556 (2020).
10. F. Rizzo *et al.*, *Nature* **584**, 201 (2020).
11. M. Neeleman *et al.*, *Nature* **581**, 269 (2020).
12. J. A. Hodge *et al.*, *Astrophys. J.* **833**, 103 (2016).
13. G. Calistro Rivera *et al.*, *Astrophys. J.* **863**, 56 (2018).

10.1126/science.abg2907

NEUROSCIENCE

Different origins for similar brain circuits

Gene expression profiles indicate separate evolution of cortical circuits in birds and mammals

By **Maria Antonietta Tosches**

More than 320 million years ago, a reptile-like amniote ancestor abandoned aquatic habitats and fully adapted to life on land. This transition was arguably the most impactful event in vertebrate history and catalyzed evolutionary innovations. Soon after, sauropsids (birds and reptiles) diverged from the ancestors of mammals. Racing to survive in their new environments, birds and mammals evolved cognitive abilities unmatched by other vertebrates, such as vocal learning in songbirds and spoken language in humans. The evolution of the brain areas supporting these behaviors has been a dilemma for neuroscientists. On page 695 of this issue, Colquitt *et al.* (*1*) disentangle molecular similarities and differences between the song nuclei of birds and the cerebral cortex of mammals and propose that these brain areas have distinct evolutionary origins.

Vocal learning circuits are part of the pallium, the roof of the most anterior vesicle of the developing neural tube. In mammals, pallial neurons are generally arranged into layers; the neocortex, in the dorsal part of the pallium, has six layers. In birds, there is no layering: The pallium is a collection of nuclei, including a set of large nuclei protruding from the ventrolateral wall of the pallium, called the dorsal ventricular ridge (DVR). Reptiles, such as turtles and lizards, also have a DVR. In the brains of other vertebrates, the pallium is simpler and there is nothing clearly comparable to the neocortex or the DVR. Thus, when the ancestors of mammals and sauropsids parted ways, they each evolved their own brain specializations—the neocortex and DVR—from expansions of the dorsal and the ventral pallium, respectively.

Despite their anatomical differences, the bird DVR and the mammalian neocortex harbor neural circuits with similar organization and function. The DVR and neocortex both have specialized visual, auditory, and somatosensory areas and are involved in high cognitive functions, such as vocal learning, planning, and abstraction. Notably, processing of sensory inputs in the avian DVR follows some of the same characteristic rules that define the "canonical" cortical microcircuit in mammals (*2*). These circuits have three major components: input neurons that receive sensory information relayed by the thalamus, intratelencephalic neurons that process this information locally, and output neurons that project to motor control centers (*3*, *4*) (see the figure). Given these functional similarities, several neuroscientists refer to the DVR as the "bird's cortex," even though its neurons are not arranged in layers.

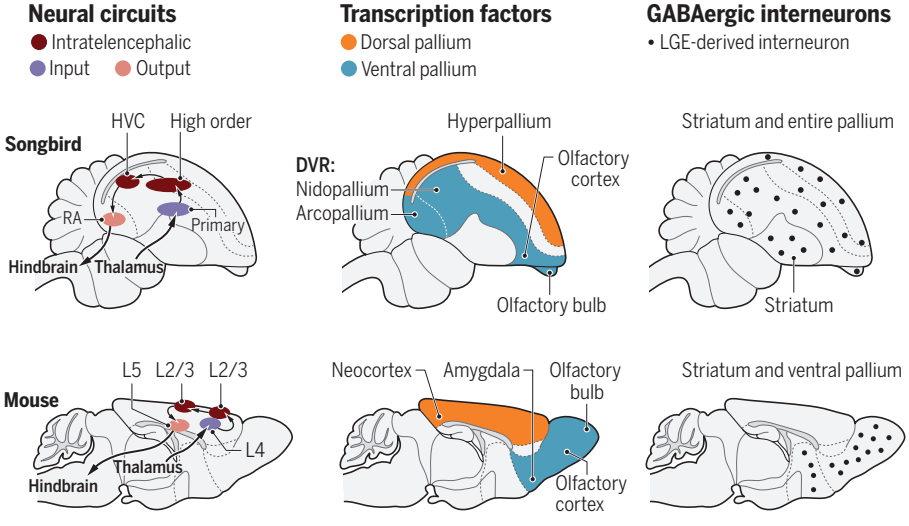
Yet, functional similarities are not sufficient to answer a key question: Do DVR and neocortical circuits trace back to the same neurons in the ancestor of birds, reptiles, and mammals? The equivalent circuit hypothesis claims so and proposes that DVR nuclei are homologous to layers of the mammalian neocortex (*3*, *4*). But this hypothesis is at odds with the observation that the DVR develops from the ventral pallium and the neocortex develops from the dorsal pallium. These are two distinct parts of the pallium that exist in all vertebrates (*5*).

To understand the evolutionary relationships of DVR and neocortical neuron types, Colquitt *et al.* collected single-cell transcriptomics data from two DVR nuclei that form part of the song circuits—HVC (proper name) and RA (robust nucleus of the arcopallium)—in zebra finches and Bengalese finches. Comparison with transcriptomes from the pallia of the red-eared slider turtle (*6*) and mouse (*7*) revealed that DVR nuclei and neocortical layers express similar genes related to neuronal function, as expected from the equivalent circuit hypothesis. However, the expression of these genes is regulated by different sets of transcription factors. DVR neurons express transcription factors enriched in the mouse ventral pallium but not in the neocortex [as also reported in reptiles (*6*)].

Transcription factors specify and maintain cellular identities, and cells that express the same transcription factors have the same core genetic identity and are typically considered homologous (*4*, *8*). For this reason, the results from Colquitt *et al.* strengthen the idea that the DVR and neocortex trace back to separate regions of the ancestral pallium (*5*, *6*). This suggests that complex behaviors, such as vocal learning, require neuronal operations that can be implemented only if neurons are wired up in a certain way, explaining why similar circuits evolved independently in birds and mammals.

Vocal processing circuits in birds and mammals

In birds, neural circuits in the dorsal ventricular ridge (DVR, simplified schematic) resemble circuits in the mammalian neocortex. However, songbird DVR circuits are genetically related to the mammalian ventral pallium and not to the neocortex (dorsal pallium), as indicated by transcription factor expression. Moreover, in songbirds, these circuits include γ-aminobutyric acid (GABA)–ergic interneurons from the lateral ganglionic eminence (LGE), which, in mammals, are restricted to parts of the ventral pallium. RA, robust nucleus of the arcopallium



Department of Biological Sciences, Columbia University, 1212 Amsterdam Avenue, New York, NY 10027, USA.
Email: mt3353@columbia.edu

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

In an unexpected twist, Colquitt *et al*. discover another key difference between the DVR and neocortex. By examining γ-aminobutyric acid (GABA)–ergic interneurons (inhibitory neurons that project locally), they find that the entire songbird pallium is populated by a type of interneuron that is largely absent in the neocortex. These interneurons, born in the lateral ganglionic eminence (LGE), migrate to ventral pallium areas in mammals, such as the amygdala and olfactory bulb. This implies that the cortex-like microcircuits described in the avian DVR may engage an entire class of interneurons that has no counterpart in the mammalian neocortex.

The study by Colquitt *et al*. is a teaser for what is to come. Single-cell genomics, one of the most powerful tools to understand neuronal diversity, is revolutionizing the study of brain evolution. But in single-cell comparative studies, the power and accuracy of bioinformatic analyses depend on the breadth and depth of the data. Colquitt *et al*. sampled a small portion of the DVR in two closely related species of finches. On the mammalian side, ventral pallium single-cell data are scarce. More data—ideally whole-pallium cell-type transcriptomes from multiple species, including amphibians and fish—are needed to clarify in detail the relationships of ventral pallium cell types in vertebrates.

An evolutionary view of the DVR that acknowledges its ventral pallium nature forces reexamination of the functional analogies with the neocortex. The independent expansions of the DVR and neocortex endowed birds and mammals with the capacity to process and integrate sensory inputs in complex ways, but perhaps for different uses. Unlike the neocortex in mammals, the DVR in birds is poorly connected with entorhinal-hippocampal circuits; instead, one of its main projection targets is the hypothalamus, resembling parts of the mammalian amygdala in the ventral pallium (*9*). Did early mammals benefit from an enhanced representation of space to tackle complex navigation tasks? Did the bird brain specialize in computing stimulus valence? A closer look at the differences between the DVR and neocortex is needed to understand the cognitive specializations of birds and mammals. ■

REFERENCES AND NOTES

1. B. M. Colquitt *et al*., *Science* **371**, eabd9704 (2021).
2. A. Calabrese, S. M. N. Woolley, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 3517 (2015).
3. H. J. Karten, *Curr. Biol.* **23**, R12 (2013).
4. S. D. Briscoe, C. W. Ragsdale, *Science* **362**, 190 (2018).
5. L. Puelles *et al*., in *Evolution of Nervous Systems*, J. H. Kaas, Ed. (Elsevier, ed. 2, 2017), pp. 519–555.
6. M. A. Tosches *et al*., *Science* **360**, 881 (2018).
7. A. Zeisel *et al*., *Cell* **174**, 999 (2018).
8. D. Arendt *et al*., *Nat. Rev. Genet.* **17**, 744 (2016).
9. L. L. Bruce, T. J. Neary, *Brain Behav. Evol.* **46**, 224 (1995).

10.1126/science.abf9551

SOCIAL SCIENCE

Asking the right questions about race and policing

When is policing the right tool, and when is it the problem?

By **Phillip Atiba Goff**[1,2]

Upon asking, "How can we fix this problem?" enough times, one wonders whether the problem can ever be fixed. This frustration characterizes an increasingly vocal segment of the United States on the topic of racism in policing, with most residents—including a majority of white respondents—believing that policing needs major change (*1*) and calls to defund the police meriting serious public engagement (*2*). On page 696 of this issue, Ba *et al*. (3) analyze officer demographics and police behavior in Chicago and show that Black officers use force less often than white officers. Such advances in quantitative social science on racism in policing (*4*) should be applauded. But the field needs to tackle questions of police abolition to remain relevant to policy discussions. Gaps remain in the field's reasoning about race that are difficult to address without considering when policing is the right tool—and when it is the problem.

"...the number of officers who identify with vulnerable groups can matter quite a bit in predicting police behavior."

Questions about when and how policing can reduce reported crime now make up one of the largest quantitative literatures on policing (*4*). This research challenged previous scientific consensus that policing could never prevent crime (*5*). Now, the question of policing as crime prevention has been the subject of two National Academies of Sciences reports and volumes of scholarship (*4, 6*). Yet, although the literature produced methodological innovations in establishing causality, it left critical gaps in understanding how policing affected vulnerable communities, particularly historically marginalized Black communities (*4*).

A more recent surge of interest in racial justice has motivated quantitative scholarship around the question of how policing can be made less racially disparate. Though this question does not directly contradict any literature (i.e., there is not a raft of scholarship arguing for policing's inherent egalitarianism), it does highlight major gaps in the previous orthodoxy. Specifically, it highlights the dearth of quantitative social science on racial equity in policing and the tacit assumption in much of the literature that reported crime (rather than, say, public concerns about state violence) is the primary metric by which public safety should be evaluated (*4*). Yet, even as emerging work such as that of Ba *et al*. begins to identify ways to combat seemingly obvious racism in policing, their findings suggest the need to engage a question that poses an even more fundamental challenge to established thinking in the field: whether and when police can be the right solution to social problems.

The assertion that police can cause more harm than good is not new (*7, 8*). But quantitative social scientists have not produced a sizable body of scholarship on the topic, and thus have fallen further behind in their ability to speak to one of the most urgent issues of public safety in the United States. For example, there is no consensus on the influence of training on officer behavior (*9, 10*). There is no consensus on what portion of racially disparate policing outcomes are driven by factors outside of police control (*11*). There is no consensus on what officer characteristics predict biased or burdensome outcomes. And, as Ba *et al*. discuss, there has been no consensus on what the influence of officer demographics is on officer behavior.

Scholars often assert that these questions remain unanswered because of the paucity of data necessary for making precise causal inference (*12*). It can take researchers years to gain access to police data of sufficient granularity to answer basic questions. The consequence is the relative silence of quantitative scientists regarding the most pressing public policy issue in policing. Indeed, a recent National Academies of Sciences report tasked with summarizing the scholarship on racial bias stemming from proactive policing

[1]Department of African American Studies and Department of Psychology, Yale University, New Haven, CT, USA. [2]Center for Policing Equity, Los Angeles, CA, USA.
Email: phillip.goff@yale.edu

concluded, essentially, that the quantitative question of racial bias was unresolved despite strong evidence from historical sources and suggestive analogs from quantitative literatures on bias in other field settings (*4*).

As a result, when scholars identify data that allow for strong inference around race and policing, they are rightly lauded by the field. And, to be sure, Ba *et al.* should be celebrated. Most of the data used to answer questions about how the race and gender of officers influence their behaviors compare officers within or between departments but cannot account for where, when, or whom officers are policing. The result is that the research tended to say "more Black officers is/is not associated with higher rates of force," but the number of potential confounding variables was humbling. The ability to leverage the newly public Chicago Police Department data to match officers with peers on the same shift and in the same neighborhoods is a quantum advance for making strong inferences. Here, Ba *et al.* remove alternative hypotheses (at least in this dataset) that Black officers are simply assigned to higher-crime neighborhoods or that women are sent to less dangerous assignments.

These data reveal that Black officers use force 35% less often than white officers, and that women use force 31% less often than men. The magnitude of the differences provides strong evidence that—at least in some cities—the number of officers who identify with vulnerable groups can matter quite a bit in predicting police behavior. Although this does not settle the matter, the work stands alone in its ability to make apples-to-apples comparisons across officers—regardless of how many may be bad apples.

Although the study by Ba *et al.* advances the field's understanding of the relationship between police demographics and behaviors, some of the most notable findings suggest additional questions that align less with "how can we fix this" and more with "why are we doing this at all."

The study reveals that 38% of the disparities in police stops of pedestrians and motorists between Black and white officers are predicted by white officers' greater use of discretionary stops such as for "vaguely defined 'suspicious behavior,'" which means someone was detained for whatever an officer deemed suspicious. Ba *et al.* frame this as an explanation of how officer demographics predict officer behaviors. But, given that Ba *et al.* find negligible demographic differences in officers' responses to community violence, such a large difference in discretionary stops compels a reader to ask: Are any of those excess stops by white officers necessary? Should a department even be making them, given the demonstrated risk for abuse so evident in vulnerable communities?

Similarly, 83% of the excess use of force by white officers, compared to Black officers, is due to differences in targeting Black residents for force. Likewise, 82% of women's lower rate of using force compared to men is attributable to differences in applying force to Black residents—across racial groups of officers. The magnitude of the finding raises the question: Are any of those excess use of force incidents by white officers necessary? And if the excess force is not necessary for public safety, why does the department target Black communities for so much physical coercion? These questions are difficult to answer outside a broader engagement with the purpose of policing—and its limitations. With violence trending downward the past three decades (*13*), mostly troubling small geographic areas (*4*), and possibly occupying a small portion of police activity (*14*), what should the role of police be? Failing to take seriously the possibility that the answer should be "much less" may end up frustrating both researchers and a public that has been asking the question for far longer than most scientists. And, because policing has been seen as racial terrorism in too many communities and for too long, a failure to name that as a principal problem may render all efforts to fix policing appear to be a frustrating waste of time. ■

REFERENCES AND NOTES

1. S. Crabtree, "Most Americans say policing needs 'major changes,'" Gallup, 22 July 2020; https://news.gallup.com/poll/315962/americans-say-policing-needs-major-changes.aspx.
2. S. Sullivan, R. Bade, "Criticized by moderates and pressured by their base, liberals fight for a voice in the Democratic Party," *The Washington Post*, 29 November 2020; www.washingtonpost.com/politics/democratic-party-future-liberals/2020/11/29/3da05bfe-2ba5-11eb-92b7-6ef17b3fe3b4_story.html.
3. B. Ba *et al.*, *Science* **371**, 696 (2021).
4. National Academies of Sciences, Engineering, and Medicine, *Proactive Policing: Effects on Crime and Communities* (National Academies Press, 2018).
5. R. V. G. Clarke, J. M. Hough, Eds., *The Effectiveness of Policing* (Gower, 1980), pp. 1–16.
6. National Research Council, *Fairness and Effectiveness in Policing: The Evidence* (National Academies Press, 2004).
7. W. E. B. Du Bois, Ed., *Black Reconstruction in America: Toward a History of the Part Which Black Folk Played in the Attempt to Reconstruct Democracy in America, 1860-1880* (Routledge, 2017).
8. A. Y. Davis, *Abolition Democracy: Beyond Empire, Prisons, and Torture* (Seven Stories, 2011).
9. J. J. Sim, J. Correll, M. S. Sadler, *Pers. Soc. Psychol. Bull.* **39**, 291 (2013).
10. G. Wood, T. R. Tyler, A. V. Papachristos, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9815 (2020).
11. P. A. Goff *et al.*, *The Science of Justice: Race, Arrests, and Police Use of Force* (Center for Policing Equity, New York, 2016).
12. P. A. Goff, K. B. Kahn, *Soc. Issues Policy Rev.* **6**, 177 (2012).
13. J. Gramlich, "What the data says (and doesn't say) about crime in the United States," Pew Research Center, 20 November 2020; www.pewresearch.org/fact-tank/2020/11/20/facts-about-crime-in-the-u-s/
14. J. Asher, B. Horwitz, "How Do the Police Actually Spend Their Time?" *The New York Times*, 19 June 2020; www.nytimes.com/2020/06/19/upshot/unrest-police-time-violent-crime.html.

10.1126/science.abf4518

THERMOELECTRICS

Tidying up the mess

Cadmium doping improves the thermoelectric properties of $AgSbTe_2$ through ordering effects

By **Yu Liu** *and* **Maria Ibáñez**

Thermoelectric materials are engines that convert heat into an electrical current. Intuitively, the efficiency of this process depends on how many electrons (charge carriers) can move and how easily they do so, how much energy those moving electrons transport, and how easily the temperature gradient is maintained. In terms of material properties, an excellent thermoelectric material requires a high electrical conductivity σ, a high Seebeck coefficient S (a measure of the induced thermoelectric voltage as a function of temperature gradient), and a low thermal conductivity κ. The challenge is that these three properties are strongly interrelated in a conflicting manner (*1*). On page 722 of this issue, Roychowdhury *et al.* (*2*) have found a way to partially break these ties in silver antimony telluride ($AgSbTe_2$) with the addition of cadmium (Cd) cations, which increase the ordering in this inherently disordered thermoelectric material.

The thermoelectric effect was discovered more than 200 years ago when Volta was conducting experiments with metallic junctions, vessels of water at different temperatures, and dead frogs (*3*). Since then, physicists, chemists, engineers, and material scientists have tried to identify the nature of the effect; understand its microscopic origin; and find, design, and optimize materials with the highest thermoelectric conversion efficiency. The quality of a thermoelectric material is evaluated by using the dimensionless thermoelectric figure of merit, ZT, which includes the relevant properties mentioned above: $ZT = \sigma S^2 T \kappa^{-1}$, where T is the absolute temperature.

Two components contribute to κ: the lattice thermal conductivity κ_{latt}, which represents heat carried by atomic vibrations (phonons), and electronic thermal conduc-

Institute of Science and Technology Austria, Am Campus 1, 3400 Klosterneuburg, Austria. Email: mibanez@ist.ac.at

tivity κ_{ele}, which represents heat carried by electrons. As the number of free charge carriers increases, σ and κ_{ele} also increase, *S* decreases, and κ_{latt} is unaltered. The optimal charge-carrier concentration for most thermoelectric materials is between 10^{19} and 10^{20} carriers per cubic centimeter (*4*). Once the carrier concentration is optimized, further improvements in the *ZT* require modifying the material electronic structure and its phononic interactions (*5*).

In practice, all of these changes are achieved by slightly modifying the compound's chemical composition by using solid solutions or by introducing nanometric secondary phases (*6*). The carrier concentration is commonly tuned by substituting atoms in the crystal structure with other atoms with a different valence. These electronic impurities, or dopants (*7*), change the chemical potential of electrons (the Fermi level) and may also affect the material's electronic structure.

The energy-band structure is linked with the overlapping of atomic orbitals and properties such as the ionization energy, bond energy, and bond length. Thus, the choice of an adequate substitutional atom to modify the electronic band becomes less intuitive (*6*). Introducing atomic defects in the host lattice will also directly affect κ by adding disorder and local strain, which alters the lattice vibrations. Last, given that all of the above is optimized, κ can be further reduced by introducing nano- or microscale features (ideally transparent to electrons) that act as barriers for the propagation of phonons (*8*).

Usually, to maximize *ZT* in a given material, all the above strategies should be cumulatively integrated. In the well-studied material lead telluride (PbTe) (*9*), the optimal number of free carriers is obtained by replacing a small percentage of the double-charged Pb atoms with a single-charged cation. Proper doping with Na creates a p-type material, $Pb_{0.98}Na_{0.02}Te$, which has a carrier concentration of around 10^{20} carriers per cubic centimeter. A solid solution of PbTe with SrTe increases the bandgap and shifts the light and heavy bands closer to each other. Such band conversion increases *S*. Last, secondary SrTe phases are formed in the PbTe matrix material by pushing the content of Sr beyond the solubility limit, which reduces κ_{latt}.

In short, to maximize *ZT* for PbTe, it was necessary to introduce lattice imperfections (Na and Sr at the Pb sites) and nanoprecipitates (SrTe)—that is, increase material structural disorder. Roychowdhury *et al.* embraced the challenge of optimizing the thermoelectric performance of an already highly disordered material, $AgSbTe_2$, a complex material system that can exhibit different structural properties depending on the synthetic conditions (*10*).

"The idiosyncrasy of the strategy...is that the introduction of a foreign atom into the crystal lattice yields a material with a higher ordered state."

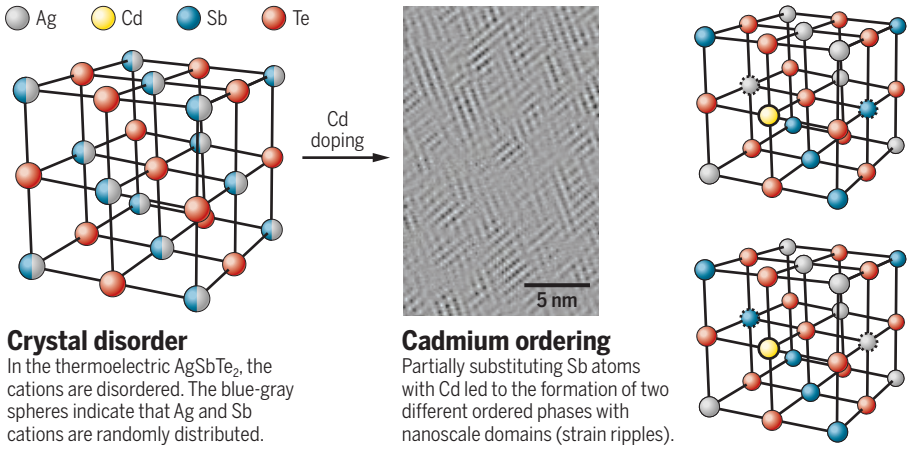
In line with previous reports, Roychowdhury *et al.* showed that polycrystalline $AgSbTe_2$ is inherently disordered because of the random distribution of Ag and Sb over the cubic structure's cationic sites (see the figure) and appearance of Ag_2Te secondary phase (*11*, *12*). To modify material properties, they replace Sb^{3+} atoms with Cd^{2+}. Such elemental substitution has further consequences than the simple increase of free charge carriers. Theoretical calculations showed that Cd preferentially occupies disordered Sb sites, which lowered ordered cationic configurations' formation energy and favored their presence.

Although the disordered phase is cubic, the ordered counterpart can be formed either in a cubic or rhombohedral crystal structure with different cationic configurations at a similar energy cost, which results in Cd-doped $AgSbTe_2$ having different ordered nanoscale domains (see the figure). The strain between the two lattice configuration results in nanoscale superstructures (strain ripples). The promoted changes in the atomic arrangement directly translate into an enhanced σ and reduced κ. Roychowdhury *et al.* explained that the increase in σ is associated with the increase of carriers and the reduction of disorder-induced fluctuations in the lattice potential, which facilitate carrier movement through the material. Two main factors hinder heat transport in this Cd-doped higher-order configuration: The nanoscale domain ripples as well as the local strain around Cd sites.

Overall, Cd's multifaceted role allowed the authors to achieve a *ZT* of up to 1.5 at room temperature, a maximum *ZT* of 2.6 at 573 K, and an average *ZT* of 1.8. These values are among the highest reported to date. The idiosyncrasy of the strategy presented by Roychowdhury *et al.* is that the introduction of a foreign atom into the crystal lattice yields a material with a higher ordered state. Considering that most thermoelectric materials are optimized through added disorder (as in the archetypal PbTe case), this result suggests a new avenue for thermoelectric material design through the optimization of atomic ordering. ■

Thermoelectricity comes to order

Roychowdhury *et al.* improved the thermoelectric performance of silver antimony telluride ($AgSbTe_2$) by partially substituting antimony (Sb) atoms with cadmium (Cd).



Crystal disorder
In the thermoelectric $AgSbTe_2$, the cations are disordered. The blue-gray spheres indicate that Ag and Sb cations are randomly distributed.

Cadmium ordering
Partially substituting Sb atoms with Cd led to the formation of two different ordered phases with nanoscale domains (strain ripples).

GRAPHIC: A. KITTERMAN/*SCIENCE*

REFERENCES AND NOTES

1. A. Zevalkink *et al.*, *Appl. Phys. Rev.* **5**, 021303 (2018).
2. S. Roychowdhury *et al.*, *Science* **371**, 722 (2021).
3. D. Beretta *et al.*, *Mater. Sci. Eng. Rep.* **138**, 100501 (2019).
4. G. J. Snyder, E. S. Toberer, *Nat. Mater.* **7**, 105 (2008).
5. L.-D. Zhao, V. P. Dravid, M. G. Kanatzidis, *Energy Environ. Sci.* **7**, 251 (2014).
6. W. G. Zeier *et al.*, *Angew. Chem. Int. Ed.* **55**, 6826 (2016).
7. M. Ibáñez *et al.*, *Chem. Mater.* **29**, 7093 (2017).
8. K. Biswas *et al.*, *Nature* **489**, 414 (2012).
9. G. Tan *et al.*, *Nat. Commun.* **7**, 12167 (2016).
10. S. V. Barabash, V. Ozolins, C. Wolverton, *Phys. Rev. Lett.* **101**, 155704 (2008).
11. K. T. Wojciechowski, M. Schmidt, *Phys. Rev. B Condens. Matter Mater. Phys.* **79**, 184202 (2009).
12. R. W. Armstrong, J. W. Faust Jr., W. A. Tiller, *J. Appl. Phys.* **31**, 1954 (1960).

10.1126/science.abg0886

CORONAVIRUS

Tracking the UK SARS-CoV-2 outbreak

The UK epidemic was seeded by more than 1000 imported viruses from Europe

By **Martha I. Nelson**

Multiple introductions of SARS-CoV-2 from travelers from Italy, Spain, and France, but not Asia, seeded the epidemic in the UK between January and June 2020.

Following a year of relatively uneventful evolution, the emergence and global spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants signals an urgent need for better genetic tracking (*1*). The United Kingdom (UK) has emerged as a leader in this domain. A £20 million investment in March 2020 established the COVID-19 Genomics UK (COG-UK) Consortium (*2*), which has produced >200,000 SARS-CoV-2 genomes, more than twice the number produced by any other country. Such a large volume of data provides an unprecedented opportunity to trace which human activities drive epidemic growth during a rapidly changing pandemic, but also introduces numerous bioinformatic challenges. On page 708 of this issue, du Plessis *et al.* (*3*) describe a new hybrid phylogenetic approach that integrates genetic data with epidemiological and travel data to uncover the roots of the UK's severe spring epidemic. Notably, they find that the UK epidemic resulted from more than 1000 transmission lineages seeded by travelers from Europe.

The study shows how last winter's control efforts were consistently one step behind the virus, allowing SARS-CoV-2 to permeate national borders. Their analysis of ~26,000 UK sequences from January to June 2020, the largest study of its kind, reveals that the UK epidemic was primarily brought into the country by travelers from European neighbors: first Italy, then Spain and France. Peak viral flow into the UK occurred in March as the virus expanded across Western Europe, but surveillance lags led to restrictions still focusing on travelers arriving from Asia. By capturing large numbers of small transmission lineages that would not be detected at lower levels of virological surveillance, as well as >1600 singleton viruses with no observed progeny, the authors uncovered an unprecedented volume of cross-border virus traffic. Genetic patterns mirrored human movement patterns, as the number of viruses entering the UK rose and then fell after international travel plummeted in March.

The UK is not the only country whose early focus on Asia as the pandemic epicenter allowed viruses to enter from European sources. Genetic data also traced the origins of epidemics in Brazil (*4*), Boston (*5*), and New York City (*6*) back to Europe. Travel restrictions can be highly effective when stringently implemented, but these studies collectively highlight how easily SARS-CoV-2 infection can arise during even small lapses in border control, including the repatriation of Americans from Asia at the beginning of the pandemic (*7*).

There is no magic bullet for triangulating scalability, speed, and statistical rigor as genomic data exceed the capacity of existing platforms. Du Plessis *et al.* confronted the methodological challenges experienced in previous evolutionary analyses of SARS-CoV-2 (*8*), magnified in this case by a substantially larger dataset. These challenges include low phylogenetic signal among genetically similar viruses, exceeding the capacity of standard phylogenetic software, as well as biases that emerge when other countries sequence different numbers of viruses relative to national case counts. The authors pursue a new approach that uses genetic data to infer the timing and number of virus introductions but uses epidemiological metadata to infer the country of origin. Better integration of genomic and epidemiological data will continue to improve outbreak responses but can be cumbersome without open-access data repositories—for example, for fluctuating global air travel volumes. Epidemiologists increasingly turn to crowd-sourced digital and cellular data to trace human movements and social contact patterns (*9*).

Contact tracing has been effective in controlling early COVID-19 outbreaks, such as Europe's first outbreak in Munich, Germany (*10*), and providing key insights into community transmission and the role of superspreading (*11*). But contact tracing is laborious and is often abandoned as epidemics grow. Genetic data can add a new dimension to these efforts by efficiently determining whether two cases belong to the same transmission lineage despite gaps in sampling among individuals in the chain. Du Plessis *et al.* did not explore heterogeneities in transmission at a city level (*5*), but their observations reveal the growth and size-dependent extinction of hundreds of co-circulating lineages as the national epidemic was brought under control by nonpharmaceutical interventions (NPIs).

The study of du Plessis *et al.* made use of a fraction of the UK sequences generated to date. The risk of new variants emerging increases as SARS-CoV-2 populations surge globally, spilling into immunocompromised, chronically infected, or even nonhuman hosts where they encounter different selection pressures. As SARS-CoV-2 becomes more evolutionarily dynamic, the UK's well-sampled data provide a resource for the global community. Denmark, Australia, and other countries also have intensive SARS-CoV-2 sequencing operations. But the UK is currently the only country with more than 1 million COVID-19 cases that sequences more than 1% of SARS-CoV-2 genomes (the UK sequences ~5%).

The most vexing evolutionary questions require broad population-level analyses based on continuous representative national sampling, with randomized selection of viruses to be sequenced (*12*). A centrally coordinated sampling strategy is a highly advantageous feature of the UK's virological monitoring program, even if it is less quantifiable than speed or volume (*2*). The United States has generated the second-

Laboratory of Parasitic Diseases, Division of Intramural Research, National Institute of Allergy and Infectious Diseases, Bethesda, MD, USA. Email: nelsonma@mail.nih.gov

PHOTO: TOBY MELVILLE/REUTERS

largest number of SARS-CoV-2 genomes, but the proportion of cases sequenced varies markedly among cities and states owing to differences in resources. Large-scale studies become methodologically challenging when datasets are amassed from smaller studies originally designed to address other research questions, introducing biases. At times it has been difficult to assess intriguing hypotheses, such as whether SARS-CoV-2 containing the spike protein D614G mutation spread globally because of fitness advantages or random chance (*13*).

Variants that arise in one country quickly become a threat to neighbors. Countries must reciprocate each other's virological monitoring efforts in a rapidly changing global viral landscape. The UK ARTIC network actively shares resources and protocols for SARS-CoV-2 sequencing. NextStrain provides a user-friendly visual platform for tracking SARS-CoV-2 evolution in almost real time. Numerous open-access bioinformatic tools were developed to analyze SARS-CoV-2 sequences (*14*). But a lesson from the UK is the importance of sustained government investment in scalable national infrastructure. Intrepid academic researchers can build popular tools but struggle to scale up as the amount of genomic data explodes. Global coordination would also be helpful, including the universal adoption of a single nomenclature for SARS-CoV-2 lineages.

The COVID-19 pandemic has galvanized long-overdue investments in promising research areas at the frontiers of technology and big data. Over the past two decades, faster, cheaper, and more portable sequencing technologies and flexible bioinformatic platforms have laid the groundwork for real-time genomic epidemiology. Leaps in progress tend to be spurred by public health crises, including outbreaks of influenza, Ebola, and Zika (*15*). The COVID-19 leap has begun. ■

REFERENCES AND NOTES

1. E. Volz *et al.*, *medRxiv* 10.1101/2020.12.30.20249034 (2021).
2. COVID-19 Genomics UK (COG-UK) Consortium, *Lancet Microbe* **1**, e99 (2020).
3. L. du Plessis *et al.*, *Science* **371**, 708 (2021).
4. D. S. Candido *et al.*, *Science* **369**, 1255 (2020).
5. J. E. Lemieux *et al.*, *Science* **371**, eabe3261 (2020).
6. A. S. Gonzalez-Reiche *et al.*, *Science* **369**, 297 (2020).
7. M. Worobey *et al.*, *Science* **370**, 564 (2020).
8. P. Lemey *et al.*, *Nat. Commun.* **11**, 5110 (2020).
9. K. H. Grantz *et al.*, *Nat. Commun.* **11**, 4961 (2020).
10. C. Roth *et al.*, *N. Engl. J. Med.* **382**, 970 (2020).
11. K. Sun *et al.*, *Science* **371**, eabe2424 (2021).
12. E. Volz *et al.*, *Cell* **184**, 64 (2021).
13. B. Korber *et al.*, *Cell* **182**, 812 (2020).
14. T. Hu *et al.*, *Brief. Bioinform.* bbaa386 (2021).
15. N. D. Grubaugh *et al.*, *Nat. Microbiol.* **4**, 10 (2019).

ACKNOWLEDGMENTS

The content does not necessarily reflect the views or policies of the Department of Health and Human Services, nor imply endorsement by the U.S. Government.

10.1126/science.abg2297

CORONAVIRUS

Single-domain antibodies make a difference

A double hit with one antibody construct may avoid viral escape

By **Xavier Saelens**[1,2] and **Bert Schepens**[1,2]

The spike protein of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of COVID-19, mediates attachment of the virion to the host cell, binding to the receptor, and fusion of the viral and host cell membranes. This releases the viral genomic RNA into the host cell cytoplasm, which is the start of virus replication. Antibodies that interfere with SARS-CoV-2 spike function—in particular, antibodies that prevent the interaction between the receptor binding domain (RBD) of spike and human angiotensin-converting enzyme 2 (ACE2), the canonical receptor on the host cell surface—can neutralize the virus in vitro and are associated with protection from infection in vivo (*1*). On page 691 of this issue, Koenig *et al.* (*2*) describe four SARS-CoV-2–neutralizing single-domain antibodies, or VHHs, and combinations thereof that can disable spike function. This extends the growing list of reports on SARS-CoV-2 spike–specific single-domain antibodies that have been proposed as potential therapeutics for COVID-19 patients.

"...some single-domain antibodies can reach sites in spike that are more difficult to access by conventional antibodies."

VHHs are the variable domains of heavy chain–only antibodies, which lack a light chain and the first constant domain of the heavy chain that occur in conventional antibodies. VHHs are present in camelid species, such as dromedaries (Arabian camels) (*3*). Koenig *et al.* immunized a llama and an alpaca with recombinant spike RBD antigen and inactivated SARS-CoV-2 virions to select high-affinity VHHs directed against SARS-CoV-2 spike. Structural analysis of the VHHs in complex with RBD or prefusion-stabilized spike revealed their epitopes, explained their capacity to hinder ACE2 binding, and showed that two of the VHHs captured spike in its so-called up-conformation in which the RBD is flipped upward, which is essential for the virus to engage with ACE2. The closed state is the predominant form of spike on the virion (*4*). ACE2 binding combined with proteolytic activation of spike triggers membrane fusion. The authors propose that some of their VHHs—including biparatopic ones that bind two different, nonoverlapping epitopes on spike—can destabilize the prefusion conformation of the spike protein and thereby inactivate the virus fusion machinery.

Several other studies have generated single-domain antibodies against spike: Some also started off with an immunized llama or alpaca (*5*–*8*), whereas others obtained neutralizing single-domain antibodies from synthetic libraries (*9*–*12*). VHH72 cross-neutralizes SARS-CoV, a betacoronavirus that caused the SARS outbreak in 2002–2003, and SARS-CoV-2 through binding of a highly conserved epitope in the spike RBDs of these viruses (*8*). VHH72 was fused with the Fc domain of human immunoglobulin G1 (IgG1), which imparts an IgG-like long circulatory half-life, enhances potency by making the antibody bivalent, and allows tuning of effector function, for example, to reduce or promote the binding to Fc receptors expressed by different immune cells. A further potency-enhanced engineered derivative of VHH72 is now entering clinical trials to treat COVID-19 patients.

Another way of increasing the valency of VHHs, which increases the binding strength to their target, is by putting them in tandem repeat fusion proteins. This is exemplified by a homotrimeric VHH construct (*5*) that, with a median inhibitory concentration (IC_{50}) of 5.4 pM, is one of the most potent SARS-CoV-2–neutralizing VHH constructs reported to date. Another potently neutralizing VHH construct was derived from a yeast-displayed synthetic VHH library, which was screened for pre-

[1]VIB-UGent Center for Medical Biotechnology, VIB, Ghent, Belgium. [2]Department of Biochemistry and Microbiology, Ghent University, Ghent, Belgium. Email: xavier.saelens@vib-ugent.be; bert.schepens@vib-ugent.be

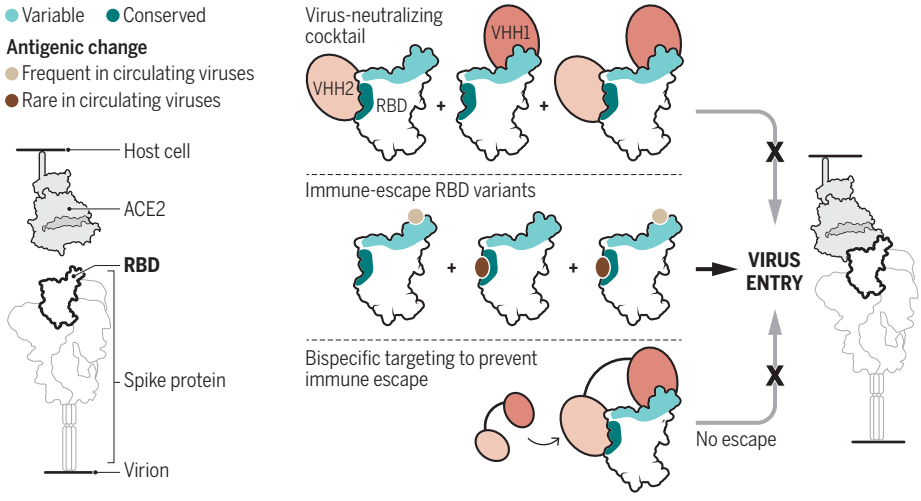
fusion-stabilized SARS-CoV-2 spike binders (*9*). One of the hits from this screen was further affinity matured by a saturating mutagenesis method and eventually converted into a homotrimeric fusion construct that neutralized SARS-CoV-2 in vitro with an IC_{50} of 56 pM. Yet, such trimeric VHH constructs will likely have limited circulatory half-lives because their limited size would still permit renal clearance and most likely would require repeated systemic administration to maintain virus-inhibitory concentration.

Although VHH humanization is now standard practice to reduce immunogenicity, an alternative is to start from a synthetic phage-displayed human IgG1 heavy-chain variable domain (V_H) library, which has also led to the isolation of a SARS-CoV-2–spike RBD-specific binder (*10*). When fused to the Fc domain of human IgG1, this V_H-Fc neutralized SARS-CoV-2 in vitro with an IC_{50} of 390 pM. Notably, the V_H-Fc was stable, did not aggregate after prolonged storage at 37°C, and was not polyreactive, which are key attributes to move a candidate biologic further into clinical development. Furthermore, this V_H-Fc protected mice and hamsters from SARS-CoV-2 infection (*10*).

Clinical trials using intravenously administered neutralizing monoclonal antibodies that target SARS-CoV-2 spike indicate benefit in outpatients (more rapid decline of virus load and lower chance of developing severe disease) and in patients with low antibody titers and high viral loads at the onset of treatment (*13*, *14*). How could VHH-based biologics contribute to the control of COVID-19? The region of the spike RBD that interacts with ACE2 is a hotspot for neutralizing human antibodies; hence, it is expected that SARS-CoV-2 variants may arise with mutations in this immunogenic ACE2-binding RBD region. Indeed, recently there has been rapid spread of SARS-CoV-2 variants, such as the B.1.1.7 variant that emerged in the United Kingdom and the 501Y.V2 variant that was found in South Africa, that are capable of reducing neutralization potency of some monoclonal antibodies that bind this region and, worryingly, even of human convalescent plasma (*15*).

Linked single-domain antibodies prevent virus escape

The spike protein of SARS-CoV-2 contains the RBD that binds to ACE2 on host cells, leading to cell entry. RBD-ACE2 binding can be prevented by single-domain antibodies, or VHHs. Combined treatment with VHH1 and VHH2, which bind nonoverlapping regions of the RBD, prevents infection until escape mutants arise. This can be overcome when the VHHs are covalently linked in a bispecific molecule.



ACE2, angiotensin-converting enzyme 2; RBD, receptor-binding domain; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

The emergence of variants urges the need to develop neutralizing antibodies that bind to epitopes that are under reduced selection pressure by the human antibody response. For this, single-domain antibodies are particularly suited. With their small size, some single-domain antibodies can reach sites in spike that are more difficult to access by conventional antibodies. An additional way of reducing the chance of viral escape from immunity is by merging different VHHs into biparatopic constructs. Using a recombinant vesicular stomatitis virus engineered to use spike for cell entry, but that is more error-prone than SARS-CoV-2 and thus evolves faster in vitro, Koenig *et al.* showed that tandem repeats of two neutralizing VHHs that bind nonoverlapping epitopes strongly reduced the chance that mutant viruses that escape neutralization were selected in vitro (see the figure).

Single domain–based antibodies can also be produced in robust economic microbial expression systems that allow fast-turnaround fermentation runs, which is important to ensure global access at an affordable price for a COVID-19 biologic. Because single-domain antibodies are typically very stable, they can be nebulized for direct pulmonary delivery (*5*, *7*, *9*). As clinical trials with neutralizing antibodies are now shifting toward outpatient and even prophylactic (prevention) settings in high-risk patients, alternative routes of administration such as subcutaneous injection and long half-lives of the antiviral biologics are being explored.

The recent rollout of COVID-19 vaccines in many countries has brought hope that the pandemic will be gradually curbed. For now, however, the spread of SARS-CoV-2 is not yet slowing down: As of now, more than 10,000 COVID-19 patients die every day globally, with many more requiring hospital admission. Thus, there is a pressing need for antiviral drugs, such as neutralizing single-domain antibodies, that can prevent or reduce SARS-CoV-2 replication in exposed patients with progressive disease. This need will likely persist even after the widespread use of COVID-19 vaccines to treat those for whom the vaccine provided no protection, was not used, or was not available, or, in a most alarming scenario, to provide treatment if vaccine-escape mutants arise. Perhaps, in the future, a positive rapid SARS-CoV-2 test outcome will go hand in hand with an easily administered, affordable, subcutaneous injection or nebulized inhalation of an antibody targeting highly conserved epitopes not recognized by the human immune system. ■

REFERENCES AND NOTES

1. S. J. Zost *et al.*, *Nature* **584**, 443 (2020).
2. P.-A. Koenig *et al.*, *Science* **371**, eabe6230 (2021).
3. C. Hamers-Casterman *et al.*, *Nature* **363**, 446 (1993).
4. Y. Cai *et al.*, *Science* **369**, 1586 (2020).
5. Y. Xiang *et al.*, *Science* **370**, 1479 (2020).
6. L. Hanke *et al.*, *Nat. Commun.* **11**, 4420 (2020).
7. T. J. Esparza, N. P. Martin, G. P. Anderson, E. R. Goldman, D. L. Brody, *Sci. Rep.* **10**, 22370 (2020).
8. D. Wrapp *et al.*, *Cell* **181**, 1004 (2020).
9. M. Schoof *et al.*, *Science* **370**, 1473 (2020).
10. W. Li *et al.*, *Cell* **183**, 429 (2020).
11. J. Huo *et al.*, *Nat. Struct. Mol. Biol.* **27**, 846 (2020).
12. T. F. Custódio *et al.*, *Nat. Commun.* **11**, 5588 (2020).
13. D. M. Weinreich *et al.*, *N. Engl. J. Med.* **384**, 238 (2021).
14. P. Chen *et al.*, *N. Engl. J. Med.* **384**, 229 (2021).
15. K. Leung, M. H. Shum, G. M. Leung, T. T. Lam, J. T. Wu, *Euro Surveill.* **26**, 2002106 (2021).

ACKNOWLEDGMENTS

We thank N. Callewaert for helpful discussions. The authors are named as inventors on patent applications related to the use of single-domain antibodies for the prevention and treatment of viral infectious diseases, including disease caused by SARS-CoV-2.

10.1126/science.abg2294

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

The bizarre and rambling nature of dreams may be key to their role in memory processing.

PHOTO: ISTOCK.COM/BRIGHTSTARS

BOOKS *et al.*

NEUROSCIENCE

Dream interpretation meets modern science

Nighttime narratives could be critical to consolidating memories, argue two sleep researchers

By **Michelle Frazer**[1] *and* **Gina Poe**[2,3]

In *When Brains Dream*, sleep scientists Antonio Zadra and Robert Stickgold detail the latest research that seeks to understand what occurs in our brains when we dream, and they present theories about what purposes dreaming may serve. The book takes the reader from humanity's early religious understanding of dreams, through our initial attempts to study the psychology of dreaming, to current experiments on the neurophysiology of the sleeping brain, providing relatable and often humorous anecdotal evidence from the authors' own lives and work along the way.

Humans have contemplated the purpose of dreams throughout recorded history. Ancient texts including the Sumerian *Iškar Zaqīqu* and Artemidorus's *Oneirocritica* reveal a prevalent belief that dreams convey important and often supernatural messages, necessitating skilled interpretation. Even Aristotle weighed in on the nature and use of dreams, although he concluded that they were likely just the result of our organs shifting during sleep.

The reviewers are at the [1]UCLA Interdepartmental Program in Neuroscience, [2]Department of Integrative Biology and Physiology, and [3]Department of Psychiatry, University of California, Los Angeles, Los Angeles, CA 90095, USA. Email: ginapoe@ucla.edu

As the world became more secularized in the 19th century, psychologists began to focus on applying their efforts to understanding the sleeping brain. Early research explored the link between waking experiences and dream content, examined the abstract nature of symbols in dreams, and even began applying statistical principles to quantify data gathered in dream journals. By the time Freud published his seminal treatise on dreams in 1899, lesser-known scientists had already gathered evidence that remains remarkably relevant to sleep research today, including the observation that vivid dreams occur most frequently during early morning sleep and that physiological changes occur as sleep progresses.

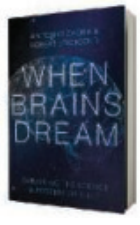
Since the discovery of rapid eye movement (REM) sleep in 1950, however, most studies have approached sleep from a biological perspective. We now know that sleep appears to play a crucial role in clearing waste from the brain, that it regulates hormone levels, and that it helps boost the immune system. There is also a large body of evidence linking sleep to learning and memory.

But are dreams necessary to these functions? Or are they a meaningless consequence of random neural firing? In the latter half of the book, the authors present the ambitious theory that dreams themselves, not just the dream sleep state, play a critical role in the consolidation of memories—a theory they refer to as network exploration to understand possibilities (NEXTUP).

On the surface, the idea that the infrequently remembered and often bizarre narratives we construct during sleep could contribute to memory consolidation seems unlikely. But Zadra and Stickgold argue that the disjointed nature of dreams is actually key to their role in memory processing. They maintain that dreams, rather than merely repeating the events of the day to cement them into long-term storage, allow our brains to freely explore memories that have been filed away over time, extracting information and developing a narrative based on associations.

According to the NEXTUP theory, dreams primarily feature weakly associated themes and memories, an idea supported by previous work in the Stickgold lab, which found that weakly associated word pairs are preferentially activated during REM sleep (*1*). Coactivation of weakly associated items could underlie the strange and often unexpected twists that unfold during dreaming.

But why do dreams take on a narrative structure at all? The authors suggest that the narrative allows the dreamer to explore and evaluate possible scenarios, providing a mechanism by which a verdict can be rendered. Dreams that elicit strong emotions, they argue, may cue the brain about the association's potential utility, which may in turn lead to a strengthening of that association.



When Brains Dream
Antonio Zadra and Robert Stickgold
Norton, 2021. 336 pp.

The NEXTUP theory appears to be supported by animal studies. Rats trained on a maze later appear to dream about maze routes they had never taken (*2*). REM sleep also appears to be a crucial time for zebra finches to "write" their songs (*3*). The brain is highly plastic during REM sleep, allowing the erasure of erroneous and weak connections as well as the establishment of new ones (*4*). These lines of evidence suggest that dreams may allow the brain to explore potential results that would be dangerous or not possible during wakefulness. ■

REFERENCES AND NOTES

1. R. Stickgold, L. Scott, C. Rittenhouse, J. A. Hobson, *J. Cogn. Neurosci.* **11**, 182 (1999).
2. A. S. Gupta, M. A. A. van der Meer, D. S. Touretzky, A. D. Redish, *Neuron* **65**, 695 (2010).
3. A. S. Dave, D. Margoliash, *Science* **290**, 812 (2000).
4. G. R. Poe, *J. Neurosci.* **37**, 464 (2017).

10.1126/science.abf7609

CLIMATE POLICY

Bill Gates's guide to avoiding climate catastrophe

The Microsoft mogul offers a pragmatic approach to achieving greenhouse gas targets

By **Miriam R. Aczel**

Bill Gates is not a climate scientist. In his new book, *How to Avoid a Climate Disaster*, he explains that he came to climate science indirectly, via issues of energy poverty that arose as part of his health and development work with the Bill and Melinda Gates Foundation. "I am aware that I'm an imperfect messenger on climate change," he writes: "The world is not exactly lacking in rich men with big ideas about what other people should do, or who think technology can fix any problem."

Yet Gates clearly has an appreciation for the technical aspects of climate change and its potential solutions. His expertise, gleaned from interactions with climate scientists, glaciologists, energy scholars, agriculture experts, and others, is apparent in the book's lucid explanations of the scientific aspects of climate change. The solutions he outlines are pragmatic and grounded in forward-thinking economic reasoning. Although he does not avoid the hard truths we must face as our climate changes, Gates remains optimistic and believes that we have the ability to avoid a total climate disaster.

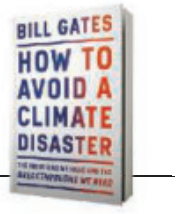
Gates presents himself as an imperfect but highly motivated climate champion.

The book's five logically organized sections establish a road map for moving forward. The first chapter explains why we need to get to net-zero greenhouse gas emissions, rendering the abstract concept of 1° or 2°C warming tangible with a description of the devastating effects such a temperature increase would have on a relatively prosperous farmer in Nebraska and a subsistence farmer in rural India. The farmers are hypothetical, but the scenarios Gates describes—crop failures, economic ruin, a mass exodus from agrarian occupations—are all too real.

Reaching net-zero emissions will not be easy. Not only is the global population increasing, people are living longer and standards of living are improving, leading to greater demand for energy and materials. It would, of course, be "immoral and impractical to try to stop people who are lower down on the economic ladder from climbing up," argues Gates. But can we find a way to support economic development without adding carbon emissions to the atmosphere?

Gates divides greenhouse gas–emitting activities into five sources: "making things"; "plugging in," or generating electricity; "growing things," including agriculture and livestock farming; "getting around"; and "keeping warm and cool." He uses this framework to evaluate various emissions-lowering strategies, assessing the "Green Premium" that makes lower-emissions solutions more costly than fossil fuel technologies and explaining how government policies and incentives can help amortize these costs.

How to Avoid a Climate Disaster: The Solutions We Have and the Breakthroughs We Need
Bill Gates
Knopf, 2021. 272 pp.



Gates argues that there are three key components necessary for reducing emissions: robust climate policies, new technologies and companies to develop zero-emissions solutions, and markets consisting of the financial institutions and investors that support these companies. These components, he maintains, must work in a complementary manner. Policies to stimulate research and development expenditure, for example, can lead to development of new technologies. Investors and markets can ensure that the technologies are scaled up. At the same time, it is critical for policies to be shaped by new technologies and to ensure that regulations keep up with technological advancements.

In the book's final chapter, Gates lays out specific actions that each of us can take to help mitigate climate change, from being a more engaged citizen to being a more informed consumer, with specific advice on reducing household emissions, switching to an electric vehicle, and setting up an internal carbon tax for businesses.

Gates includes an afterword on COVID-19, making the case that many of the lessons we have learned from the pandemic—the need for international cooperation, the importance of letting science guide our actions, the understanding that solutions should be designed to meet the needs of those who suffer most—also apply to climate change. We are at a critical juncture, he argues, and must commit to investing in new policies, market structures, and technologies over the next decade if we are to eliminate greenhouse gas emissions by 2050. "It's hard to think of a better response to a miserable 2020 than spending the next ten years dedicating ourselves to this ambitious goal," he concludes. ■

10.1126/science.abf9618

The reviewer is at the California Institute for Energy and Environment, University of California, Berkeley, Berkeley, CA 94720, USA, and the Centre for Environmental Policy, Imperial College London, London SW7 1NE, UK. Email: miriam.aczel14@imperial.ac.uk

PHOTO: MIKE LAWRENCE/GETTY IMAGES FOR GATES ARCHIVE

LETTERS

Père David's deer (*Elaphurus davidianus*) remain at risk after rewilding.

PHOTO: ZHENYU ZHONG

Edited by **Jennifer Sills**

Reversing extinction in China's Père David's deer

Père David's deer (*Elaphurus davidianus*) became extinct in the wild in China in the late 19th century (*1*), but after reintroduction three decades ago, the population has grown to more than 8000 individuals, with at least 2420 living in the wild in China (*2, 3*). This species reintroduction is considered one of the most successful stories in the world (*4*); it provides a watershed example of biodiversity conservation best practices for the upcoming UN Decade on Ecosystem Restoration. However, the species needs continued support to thrive.

Heavily endangered as a result of overhunting and habitat loss (*1*), the last of China's Père David's deer had been pillaged by Europeans and sent to zoos across Europe before the Qing Dynasty fell in 1912 (*1*). In 1985, nearly a century later, 38 deer were donated back to China for re-wilding (*5*). After the reintroduction of an additional 39 deer from zoos in England in 1986 (*6*), the population in Dafeng Milu National Nature Reserve grew to 5681 individuals (*2*). Père David's deer populations now cover all habitats in which they lived before the species' extinction in the wild.

The conservation of Père David's deer, however, faces ongoing challenges. The populations within China lack genetic diversity and are therefore susceptible to multiple risks, such as high miscarriage rates, reduced life span, and diseases (*4*). Wild populations are also limited by environmental constraints, keeping population numbers low and hindering stability. Moreover, there is no master plan for Père David's deer conservation at the national level and thus no coordinated monitoring platform or guidelines for overcoming inbreeding and environmental obstacles. Finally, there is a lack of international cooperation between researchers in China and abroad. National and international cooperation to strengthen monitoring of Père David's deer populations and to develop a shared database and germplasm databank between all countries with wild populations would be a major step forward for securing the long-term conservation of this species.

Yuanyuan Zhang[1,2], Jiade Bai[1,2]*, Annah Zhu[3], Ruishan Chen[4], Dayuan Xue[5], Zhenyu Zhong[1,2], Zhibin Cheng[1,2]

[1]Beijing Biodiversity Conservation Research Center, Beijing, 100076, China. [2]Beijing Milu Ecological Research Center, Beijing, 100076, China. [3]Environmental Policy Group, Wageningen University, 6706 KN Wageningen, Netherlands. [4]East China Normal University, Shanghai 200042, China. [5]College of Life and Environmental Sciences, Minzu University of China, Beijing, 100081, China.
*Corresponding author.
Email: baijiade234@aliyun.com

REFERENCES AND NOTES

1. G. Jiang *et al.*, *Acta Zool. Sin.* **47**, 53 (2001) [in Chinese].
2. J. Kou, "The number of Père David's deer population in China exceeds 8000," people.cn (2020); http://society.people.com.cn/n1/2020/1109/c1008-31923108.html [in Chinese].
3. P. F. Li et al., *Chin. J. Wildl.* **39**, 041 (2018) [in Chinese].
4. Y. Y. Du, *China Dev. Rev.* **8**, 50 (2018) [in Chinese].
5. C. Q. Yu, *Biodivers. Sci.* **4**, 130 (1996) [in Chinese].
6. K. B. Wang, H. Jiang, Y. T. An, *Chin. J. Wildl.* **41**, 806 (2020) [in Chinese].

10.1126/science.abg6237

A marine biodiversity plan for China and beyond

Although China has taken important steps to protect biodiversity, including a 10-year fishing ban in inland waters (*1*) and a summer fishing moratorium in oceans (*2*), the country's marine biodiversity protection needs a plan to better ameliorate the threats to marine life, including the deterioration of ecosystems and loss of biodiversity (*3*). By 2019, China had established 271 protected areas that include oceans (*4*, *5*). They cover 12.4 million hectares of ocean but represent only 4.1% of China's maritime area (*4*, *5*), far below the 10% goal set by Aichi Target 11 (*6*). In contrast, China has established protected areas covering 18% of its total terrestrial habitats and inland waters, more than the 17% goal of Aichi Target 11 (*6*, *7*). Nearly all of China's marine protected areas (MPAs) are in coastal and continental-shelf regions (*5*), and their ability to protect migratory fishes and mammals is likely hindered by their fragmentation (*4*). The current MPA system cannot keep up with the demand for effective marine biodiversity protection. China needs to accelerate the optimization and integration of existing MPAs and establish new MPAs that use the national

parks model (*8*) to strengthen the protection of marine biodiversity.

Given the connectivity of the ocean and the migratory ability of animals, MPAs should target protection of their key habitats such as breeding or spawning sites. Because shortfalls in effective management can diminish protected areas' ability to stem biodiversity loss (*9*, *10*), MPA policies such as fishing intensity limits and seasonal fishing moratoriums need to be better enforced. In China's protected areas management system, the national park model emphasizes the integrity of the ecosystem and prioritizes the integration of one or more complete ecosystems. Designating areas as national parks can break through administrative obstacles. The current MPA system should be expanded to add more and larger national parks in key ecosystems as determined by scientific data (*8*). These areas likely include the Changshan islands, the Nanji and Beiji islands, the estuaries of the Yangtze and Yellow rivers, the Beibu Gulf, Nanpeng islands, and the coral reef islands in the South China Sea.

Scientists are calling for 30% of the world's oceans to be fully protected by 2030 (*11*). However, 61% of the world's oceans are high seas, which hold the largest global biodiversity richness. In order to protect marine biodiversity in international waters, countries will have to cooperate (*11*). In April 2019, Chinese President Xi Jinping proposed that the international community work together to ensure a shared future for the world's oceans (*12*). To achieve this goal, greater efforts are required by China and all countries.

Wenliang Zhou[1], Meng Wang[2], Mingpan Huang[2], Fuwen Wei[1,2,3]*

[1]Center for Evolution and Conservation Biology, Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou 511458, China. [2]Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China. [3]Endangered Species Scientific Commission of the People's Republic of China, Beijing 100101, China.
*Corresponding author. Email: weifw@ioz.ac.cn

REFERENCES AND NOTES

1. "Notice of the Ministry of Agriculture and Rural Affairs on the scope and time of fishing ban in key waters of the Yangtze River Basin" (2019); www.cjyzbgs.moa.gov.cn/tzgg/201912/t20191227_6334009.htm [in Chinese].
2. "Notice of the Ministry of Agriculture and Rural Affairs to adjust the marine fishing moratorium system in summer" (2018); www.moa.gov.cn/nybgb/2018/201803/201805/t20180528_6143235.htm [in Chinese].
3. J. Y. Liu, *PLOS One* **8**, e50719 (2013).
4. L. L. Zhao *et al*., *Mar. Dev. Manag*. **36**, 3 (2019). [in Chinese].
5. *Fish. Sci. Technol. Info.* **47**, 350 (2020) [in Chinese].
6. D. Coates, "Strategic Plan for Biodiversity (2011–2020) and the Aichi Biodiversity Targets" in *The Wetland Book*, C. M. Finlayson *et al*., Eds. (Springer, Dordrecht, Netherlands, 2016).
7. F. W. Wei *et al*., *Natl. Sci. Rev.* 10.1093/nsr/nwaa279 (2020).
8. F. L. Tang, *Trop. For.* **46**, 5 (2018). [in Chinese].
9. K. McVeigh, "Auditors decry 'marine protected areas' that fail to protect ocean," *The Guardian* (2020).
10. S. L. Maxwell *et al*., *Nature* **586**, 217 (2020).
11. C. Roberts *et al*., "30×30: A blueprint for ocean protection: How we can protect 30% of our oceans by 2030" (Environment Foundation Greenpeace, 2019).
12. G. S. He, "The destiny of the oceans is linked to the destiny of mankind," *Guangming Daily* (2019); https://epaper.gmw.cn/gmrb/html/2019-06/06/nw.D110000gmrb_20190606_1-16.htm [in Chinese].

10.1126/science.abg7976

Worst climate outcomes are still possible

In his news item "Global warming forecasts sharpen" in the Breakthrough of the Year news package (18 December, http://scim.ag/BOTYwarming), P. Voosen wrote that new research on climate sensitivity "rule[s] out" the most dire climate scenarios. Although this work is a major advance, its scope is insufficient to eliminate the worst possible outcomes.

The research to which Voosen refers constrains "climate sensitivity"—the warming response to a given increase in the atmospheric CO_2 concentration (*1*). However, climate sensitivity does not account for the complex relationship between human CO_2 emissions and changes in atmospheric CO_2. That relationship depends on carbon cycle feedbacks, which currently act to stabilize climate but could weaken or reverse as warming progresses (*2*, *3*). Even uncontrolled warming is potentially possible (*4*); paleoclimate evidence suggests that carbon cycle feedbacks may have amplified past natural warming trends (*5*–*7*). Of particular concern are greenhouse gas emissions from thawing permafrost, which are substantial now and will only increase as warming progresses (*8*).

Because of limited scientific understanding, carbon cycle feedbacks are incompletely represented in many climate models, including those used in the recent Intergovernmental Panel on Climate Change special report, "Global warming of 1.5°C" (*9*). Every ton of carbon emitted by permafrost is one less ton that humans can emit while limiting warming to 1.5° or 2°C. Thus, unaccounted-for carbon cycle feedbacks increase the challenge of meeting any top-line climate goal.

The new constraints on climate sensitivity that Voosen highlights are welcome, but they must be carefully interpreted to avoid creating a false sense of security about the possible severity of the warming response to human emissions of greenhouse gases. Even with perfect knowledge of climate sensitivity, we would still need better understanding of carbon cycle feedbacks to determine how much humans can emit and still meet top-line climate policy goals of limiting global warming to 1.5° or 2°C—a question of fundamental importance to our collective future.

Philip B. Duffy[1]* and Margaret L. Duffy[2]

[1]Woodwell Climate Research Center, Falmouth, MA 20540, USA. [2]Department of Earth, Atmospheric, and Planetary Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
*Corresponding author.
Email: pduffy@woodwellclimate.org

REFERENCES AND NOTES

1. S. Sherwood *et al*., *Rev. Geophys.* **58**, e2019RG000678 (2020).
2. P. Friedlingstein, *Phil. Trans. R. Soc. A* **373**, 20140421 (2015).
3. Z. Hausfather, R. Betts, "Analysis: How 'carbon-cycle feedbacks' could make global warming worse" (Carbon Brief, 2020).
4. W. Steffen *et al*., *Proc. Natl. Acad Sci. U.S.A.* **115**, 8252 (2018).
5. C. Agnini *et al*., *Paleoceanography* **24**, PA2209 (2009).
6. R. M. DeConto *et al*., *Nature* **484**, 87 (2012).
7. J. M. Cotton *et al*., *Am. J. Sci*. **315**, 337 (2015).
8. S. M. Natali *et al*., *Nat. Clim. Change* **9**, 852 (2019).
9. Intergovernmental Panel on Climate Change, "Global warming of 1.5°C" (2018).

10.1126/science.abg2720

TECHNICAL COMMENT ABSTRACTS

Comment on "Structural evidence for a dynamic metallocofactor during N_2 reduction by Mo-nitrogenase"

John W. Peters, Oliver Einsle, Dennis R. Dean, Serena DeBeer, Brian M. Hoffman, Patrick L. Holland, Lance C. Seefeldt

Kang *et al*. (Reports, 19 June 2020, p. 1381) report a structure of the nitrogenase MoFe protein that is interpreted to indicate binding of N_2 or an N_2-derived species to the active-site FeMo cofactor. Independent refinement of the structure and consideration of biochemical evidence do not support this claim.

Full text: dx.doi.org/10.1126/science.abe5481

Response to Comment on "Structural evidence for a dynamic metallocofactor during N_2 reduction by Mo-nitrogenase"

Wonchull Kang, Chi Chung Lee, Andrew J. Jasniewski, Markus W. Ribbe, Yilin Hu

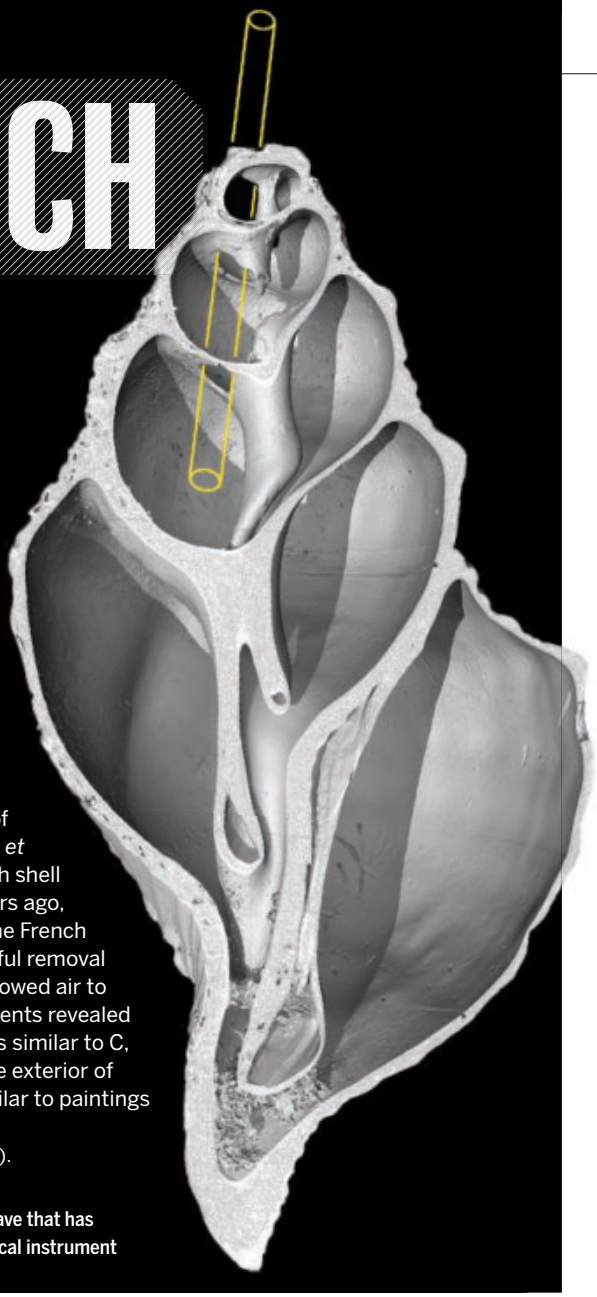
Peters *et al*. comment on our report of the dynamic structure of the nitrogenase metallocofactor during N_2 reduction. Their claim that their independent structural refinement and consideration of biochemical data contradict our finding is incorrect and is strongly refuted by our biochemical and structural data that collectively and conclusively point to the binding of dinitrogen species to the nitrogenase cofactor.

Full text: dx.doi.org/10.1126/science.abe5856

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**



Cross section of the conch shell from Marsoulas Cave that has been modified with drilled holes to serve as a musical instrument

CREDIT: FRITZ *ET AL.*, *SCI. ADV.* 10.1126/SCIADV.ABE9510 (2021)

ARCHAEOLOGY

Sounds from a Paleolithic seashell horn

Although flutes and whistles made of bone have been found at Upper Paleolithic archaeological sites in Europe, musical instruments made of other materials are uncommon. Fritz *et al.* have identified the earliest known conch shell in the world, dating from about 17,000 years ago, which was used as a wind instrument at the French Magdalenian site of Marsoulas Cave. Careful removal of the apex of the shell in ancient times allowed air to flow through it to produce sound. Experiments revealed that playing the shell produces three notes similar to C, C-sharp, and D. Image enhancement of the exterior of the shell revealed dots of red pigment similar to paintings found on the cave walls.
—MSA *Sci. Adv.* 10.1126/sciadv.abe9510 (2021).

GEOLOGY

A boring billion for mountains

Earth's crust has changed over time as supercontinents formed and broke apart. Tied into this cycle are the building and erosion of high mountains, which are tied to collisions between tectonic plates. Tang *et al.* use europium anomalies in zircons to estimate the mean thickness of crust over Earth's history. This proxy shows that mountain building has not always been as active as it is today or as it was very early in Earth's history. Mountain building, and the subsequent erosion, was less intense for about a billion years, roughly correlated with a so-called "boring billion" period of biological evolution. —BG
Science, this issue p. 728

SUPERCONDUCTIVITY

A superconducting interface

Interfaces between materials can harbor quantum states that belong to neither of the materials. A classic example is the superconducting interface between two insulating oxides, $LaAlO_3$ and $SrTiO_3$, with a critical temperature of around 200 millikelvin. Liu *et al.* observed superconductivity at a different interface—formed between $KTaO_3$ as a substrate and an overlayer of either EuO or $LaAlO_3$—at a considerably higher temperature of about 2 kelvin. Transport measurements displayed anisotropy, which may indicate an unusual superconducting state. —JS
Science, this issue p. 716

CORONAVIRUS

Fighting zoonotic coronaviruses

In the past 20 years, three betacoronaviruses thought to have originated in bats have caused devastating disease in humans. The global pandemic caused by the latest such virus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), highlights the need to protect against other strains that could present a threat to humans. Cohen *et al.* constructed nanoparticles displaying the protein domain that binds the host cell receptor (receptor-binding domain or RBD), either a homotypic SARS-CoV-2 particle or mosaic particles displaying RBDs from four or eight different betacoronaviruses. In mice, antibodies to the SARS-CoV-2 RBD were elicited just as well by mosaic particles as by homotypic nanoparticles. The mosaic nanoparticles elicited antibodies that, beyond recognizing the strains displayed, also recognized mismatched strains. —VV
Science, this issue p. 735

NUCLEOTIDE SYNTHESIS

A P catalyst for stereogenic P(III)

Phosphodiesters, the key linkage in DNA, RNA, and bioactive oligonucleotide and cyclic nucleotides, are typically synthesized from phosphoramidite precursors. Phosphorothioate linkages, which are more stable, can also be produced this way but have the complication of chirality at the P center, which is synthetically problematic, especially for multiple linkages. Featherston *et al.* found that two different chiral phosphoric acid catalysts provide stereodivergent synthesis of a stereogenic phosphite intermediate that can be oxidatively modified to the desired stereopure nucleotide derivative. They used this approach in the synthesis of a phosphorothioate derivative of the cyclic dinucleotide 2′,3′-cyclic guanosine monophosphate adenosine monophosphate, which is an important human immune signaling molecule. —MAF
Science, this issue p. 702

CANCER

An "Ama"-zing antibody-drug conjugate

Antibody-drug conjugates (ADCs) targeting HER2 are an attractive candidate for treating patients with HER2$^+$ breast cancer. Li *et al.* developed an ADC that is especially effective at killing breast cancer cells that express HER2 at low levels. HER2-low breast cancer cells often have loss of chromosome 17p, which the authors showed sensitizes tumor cells to the drug α-amantin. The authors treated mice bearing HER2-low tumors with an ADC of trastuzumab and α-amantin (T-Ama), and found that this combination induced immunogenic cell death of tumor cells, which enhanced the response to immune checkpoint blockade. —CSM

Sci. Transl. Med. **13**, eabc6894 (2021).

OCEANOGRAPHY

Structure from a whale song

Probing the structure of the ocean crust requires a wave source. The most common source is an air gun, which is effective but potentially harmful for ocean life and not easy to use everywhere. Kuna and Nábělek found that fin whale songs can also be used as a seismic source for determining crustal structure. Fin whale vocalizations can be as loud as large ships and occur at frequencies useful for traveling through the ocean floor. These properties allow fin whale songs to be used for mapping out the density of ocean crust, a vital part of exploring the seafloor. —BG

Science, this issue p. 731

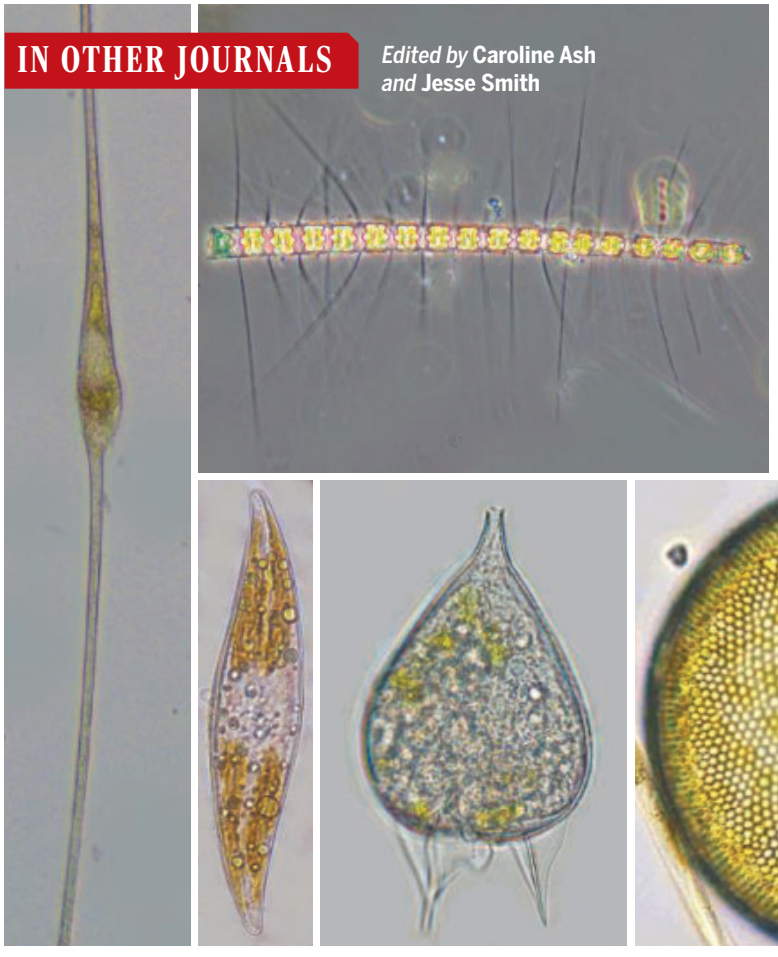
CRIMINAL JUSTICE

Diversity in policing

In the wake of high-profile police shootings of Black Americans, it is important to know whether the race and gender of officers and civilians affect their interactions. Ba *et al.* overcame previous data constraints and found that Hispanic and Black officers make far fewer stops and arrests and use force less than white officers, especially against Black civilians. These differences are largest in majority-Black neighborhoods in the city of Chicago (see the Perspective by Goff). Female officers also use less force than male officers. These effects are supportive of the efficacy of increasing diversity in police forces. —TSR

Science, this issue p. 696; see also p. 677

Fin whales produce vocalizations that can be used to study the structure of the Earth's subocean crust.



IN OTHER JOURNALS

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

CANCER

Painting portraits of tumors

Detailed assessment of tumor genomics, epigenomics, and transcriptomics is becoming increasingly common. As these methods generate ever-increasing collections of data, it can be difficult to analyze all of this information and determine how it applies to individual patients or rare cancer subtypes. To help interpret the accumulating data, Zhou *et al.* designed a web-based tool called GenomePaint, which can visualize a variety of omics data in cohorts of patients and in individual samples. The authors also used GenomePaint to identify previously unreported and likely pathogenic alterations in several tumor types, demonstrating the potential power of this method. —YN

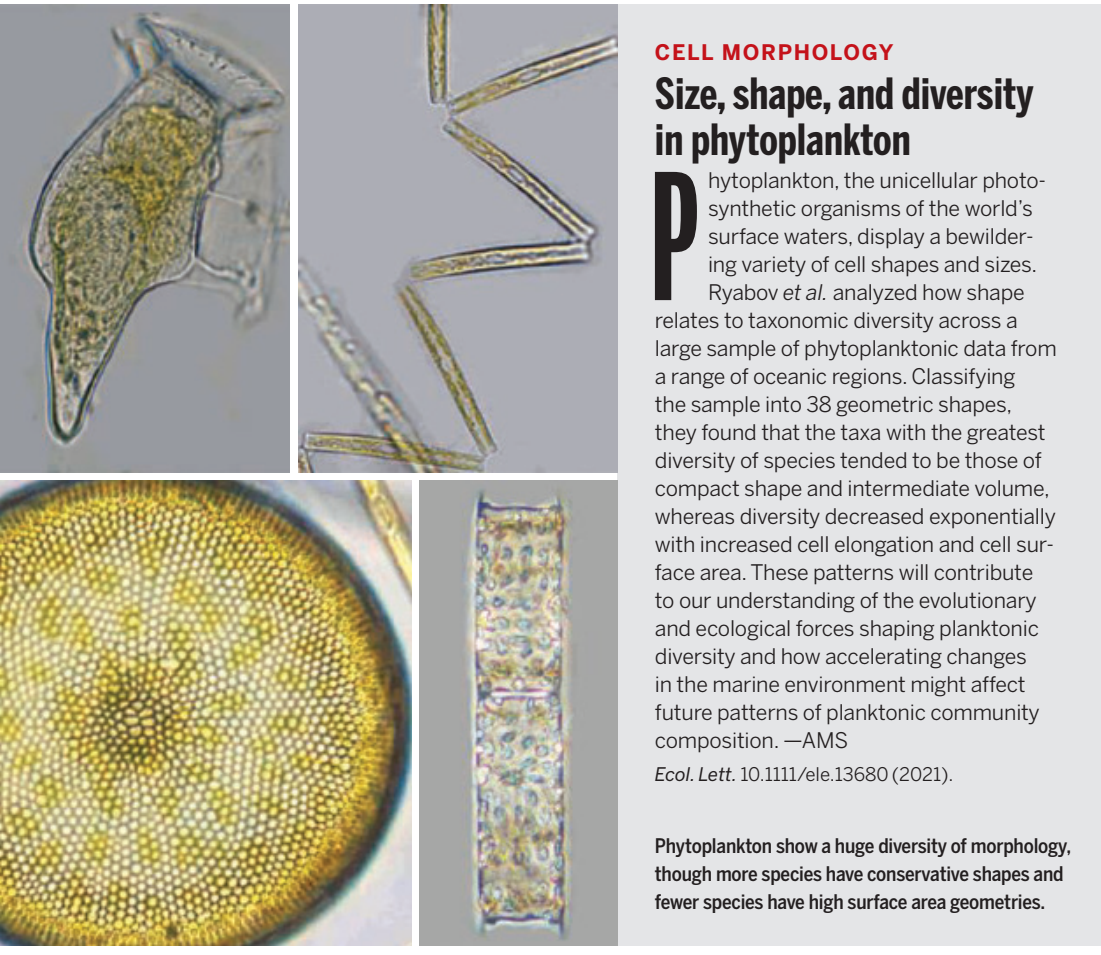
Cancer Cell **39**, 83 (2021).

NEUROSCIENCE

Brain disease and network reorganization

Parkinson's disease (PD) offers an opportunity to study the interplay between compensatory functional and molecular mechanisms. Although postsynaptic dopamine D2 receptors are up-regulated in PD, their importance in neural activity has never been identified. To investigate the relationship between reorganization of dopaminergic networks and functional reorganization of brain activity patterns, Rebelo *et al.* combined PET and fMRI scans of the brain areas involved in the execution of saccades. They found a tight link between functional activation and synaptic changes at the molecular level, reflecting network reorganization in PD. The association between D2 receptor binding and reorganization of the saccadic cortical network

CREDITS (FROM LEFT): TUI DE ROY/MINDEN PICTURES; RYABOV *ET AL., ECOL. LETT.* 10.1111/ELE.13680 (2021)



CELL MORPHOLOGY

Size, shape, and diversity in phytoplankton

Phytoplankton, the unicellular photosynthetic organisms of the world's surface waters, display a bewildering variety of cell shapes and sizes. Ryabov *et al.* analyzed how shape relates to taxonomic diversity across a large sample of phytoplanktonic data from a range of oceanic regions. Classifying the sample into 38 geometric shapes, they found that the taxa with the greatest diversity of species tended to be those of compact shape and intermediate volume, whereas diversity decreased exponentially with increased cell elongation and cell surface area. These patterns will contribute to our understanding of the evolutionary and ecological forces shaping planktonic diversity and how accelerating changes in the marine environment might affect future patterns of planktonic community composition. —AMS

Ecol. Lett. 10.1111/ele.13680 (2021).

Phytoplankton show a huge diversity of morphology, though more species have conservative shapes and fewer species have high surface area geometries.

reflects parietostriatal rerouting in response to a progressive frontostriatal dopamine deficit as the disease develops. —PRS

Proc. Natl. Acad. Sci. U.S.A. **118**, e2013962118 (2020).

NEUROINFLAMMATION

Elucidating synapse loss in MS

Synapses appear to provide the initiation sites of neuronal damage in human multiple sclerosis (MS), but how this synapse loss occurs, how it affects circuit activity, and how it could be prevented are all unknown. Jafari *et al.* describe a mouse model of cortical MS pathology designed to address these questions. They used structural and functional in vivo imaging to reveal the pathogenesis, functional consequences, and therapeutic targeting of immune-mediated synapse loss in inflamed cortical gray matter. Low-dose myelin protein immunization and stereotactic injection of cytokines implicated in MS were used to target MS-like neuroinflammatory lesions to mouse cortex. Single-cell and ultrastructural analysis revealed the loss of one-third of all synapses across both cortical hemispheres. Synapse pathology was completely reversible within a few weeks, after cortical inflammation had subsided. Localized calcium accumulations primed individual synapses for pruning by invading macrophages and activated microglia. Furthermore, an anti–colony-stimulating factor 1 receptor antagonist inhibited synapse pruning and cortical synapse loss. Thus, synaptoprotective therapeutic approaches may be feasible in MS-like neuroinflammatory disease. —SMH

Nat. Neurosci. 10.1038/s41593-020-00780-7 (2021).

QUANTUM NETWORKS

A dynamic quantum memory

Photons, being robust and fast, are ideal carriers of quantum information in long-distance quantum networks. For a practical communication platform, the photons also need to be manipulated at the local level at nodes or exchange points across the network, and this requires information to be reliably encoded, stored, and retrieved from the photonic qubits. Craiciu *et al.* demonstrate such functionality using an ensemble of erbium ions embedded within a solid-state matrix of yttrium orthosilicate. With the ions operating as a dynamic quantum memory for photons at communication wavelengths and the approach being compatible with integrated photonics, the results offer promise for possible integration into large-scale optical fiber networks. —ISO

Optica **8**, 114 (2021).

OZONE LOSS

The role of prevention

The Montreal Protocol, which restricts the manufacturing and use of a host of atmospheric ozone-depleting chemicals, is arguably the most successful environmental agreement ever implemented. Its work is not complete, however, because new dangerous compounds continue to be made and used. Vollmer *et al.* report the detection of three previously unrecognized species of ozone-destroying hydrochlorofluorocarbons (HCFCs). The compounds, with no known end uses, are probably intermediates formed in the production of other HCFCs. Identifying these types of sources before they become a more costly and less manageable problem will help make the Montreal Protocol more effective. —HJS

Proc. Natl. Acad. Sci. U.S.A. **118**, e2010914118 (2021).

MAGNETISM

A close look into a very thin magnet

Single layers of several van der Waals materials have recently been shown to be magnetic. To use these extremely thin magnets for practical purposes, researchers must first learn about the nature of the magnetism that they possess. Abdul Wahab ***et al.*** studied one such material, CrI_3, simulating its properties in both bulk and monolayer form and comparing these computational results with experimental data. The study showed that the magnetic interactions in CrI_3 cannot be described by either Ising or Heisenberg models and instead require accounting for higher-order terms. The monolayer form of CrI_3 was revealed to have metastable magnetic domains with very narrow domain walls, which could be exploited for future applications. —JS

Adv. Mater. **33**, 2004138 (2020).

REVIEW SUMMARY

PHYSIOLOGY

Communicating clocks shape circadian homeostasis

Kevin B. Koronowski* and Paolo Sassone-Corsi†

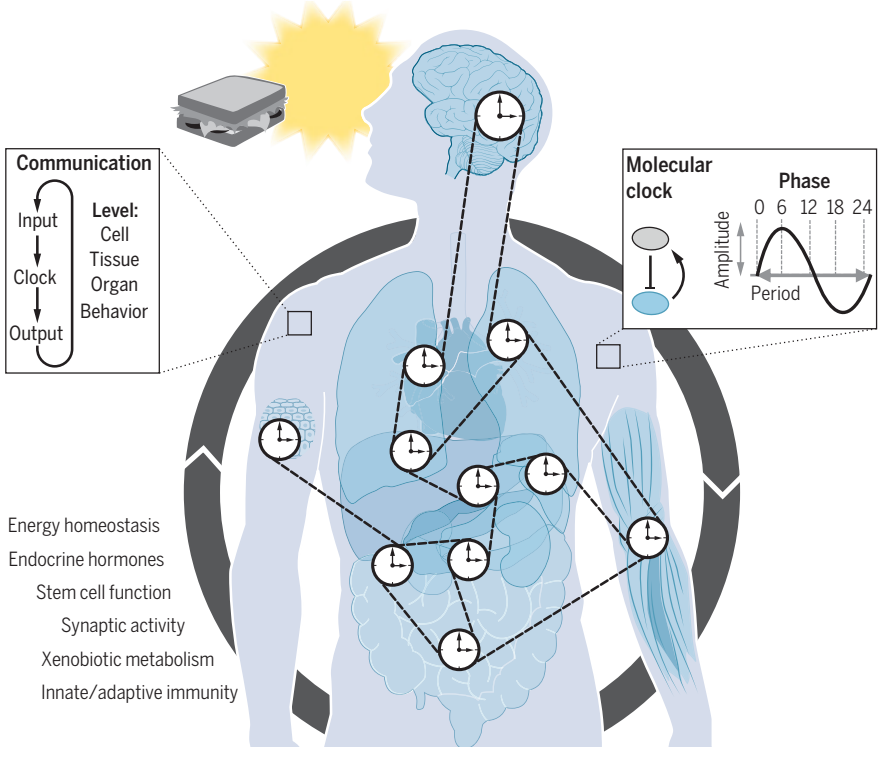
BACKGROUND: Life-forms ranging from bacteria to humans are programmed by circadian clocks—mechanisms that impose an ~24-hour rhythmicity on biology in harmony with geophysical time. Our understanding of circadian rhythms was transformed by the identification of clock genes and the discovery that these genes encode a molecular machinery that oscillates autonomously. With a genetic basis for the clock, complex organisms can consolidate timekeeping in specialized cells and anatomical regions, or they can disperse this task to all cells through ubiquitous expression. Studies in plants, flies, and mice have revealed a vast array of organizations of circadian clock systems across species, all of which rely on the passage of circadian information between cells. In propelling daily cycles of homeostatic processes, the mammalian system functions as coupled cell and tissue clocks that span both the brain and peripheral organs. Recent advances have shed light on how constituent clocks communicate to generate intricate rhythms at every level of physiology.

ADVANCES: In the brain, activity of the central clock (also known as the pacemaker) in the suprachiasmatic nucleus is driven by both neurons and astrocytes. Real-time luciferase and calcium imaging techniques have revealed that astrocytes harbor their own molecular clock, which oscillates in antiphase to neurons and is markedly sufficient on its own to propel rhythms in mice. This feat hinges on neurotransmitter interplay that couples the two cell types. In the forebrain, the sleep-wake cycle controls the daily accumulation and phosphorylation of synaptic proteins, adding an additional layer of posttranscriptional circadian regulation to neuronal function.

Studies in peripheral organs have demonstrated how cellular clocks achieve temporal coherence. Pancreatic islets time the release of insulin, glucagon, and somatostatin to determine the phase relationships of resident α, β, and δ cells, thereby establishing a basal layer of synchrony. Single-nucleus sequencing of isolated liver cell populations shows how clock disruption in hepatocytes influences the molecular rhythms of neighboring endothelial and immune cells, which suggests that circadian programming can be passed from one cell type to another, perhaps to temporally integrate different functional niches. Peripheral clocks also act systemically on distal clocks—a growing list of tissues secrete bona fide synchronizing factors into the circulation, including skeletal muscle, intestine, liver, and adipose tissues.

Complementing tissue-specific loss-of-function experiments, tissue-specific reconstitution of the clock in otherwise clockless mice demonstrates that peripheral clocks are only sufficient to drive a small fraction of local rhythms and thus rely heavily on incoming circadian signals. Extrinsic transcriptional control stems from the cooperation of the molecular clock with lineage-specific transcription factors at gene promotors and enhancers. Through interactions with clock proteins, nuclear receptors regulate specific sets of genes in response to hormone and metabolite fluctuations generated by the clock in other tissues.

OUTLOOK: In modern society, we make conscious decisions, often out of necessity, to override our clock programming. As a result, our rhythms can be dissonant with the environment and, if left uncorrected, can cause adverse health effects. Circadian misalignment, wherein eating and sleeping patterns oppose their natural inclination from the light-dark cycle, disrupts homeostasis and leads to internal imbalance—a feature of diseases ranging from metabolic syndrome to cancer. By contrast, proper alignment and internal synchrony have been demonstrated to combat tissue dysfunction and promote well-being. Thus, our innate circadian biology presents both challenges and opportunities. Given that clock disruption is also a consequence of disease, a task at hand for researchers is to identify the suite of mechanisms through which clock-to-clock communication is achieved and then understand why those mechanisms fail. Re-establishing intertissue timing and coordination could serve as a promising avenue for therapeutic interventions. ■



Cellular to organismal timekeeping: Communication between clocks. The mammalian circadian clock is a coupled network of cell and tissue clocks. Light and food are predominant cues—pushing and pulling on the phase, enhancing or attenuating the amplitude, and activating or inhibiting functional rhythms. In a tissue-specific manner, clocks receive input signals and convert them into timed functional outputs, many of which, in turn, act as inputs and effectively connect the network.

Center for Epigenetics and Metabolism, U1233 INSERM, Department of Biological Chemistry, University of California, Irvine, CA 92697, USA.
*Corresponding author. Email: kkoronow@uci.edu (K.B.K.)
†Deceased.
Cite this article as K. B. Koronowski and P. Sassone-Corsi, *Science* 371, eabd0951 (2021). DOI: 10.1126/science.abd0951

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abd0951

CREDIT: N. CARY/*SCIENCE*

RESEARCH ARTICLE SUMMARY

CORONAVIRUS

Structure-guided multivalent nanobodies block SARS-CoV-2 infection and suppress mutational escape

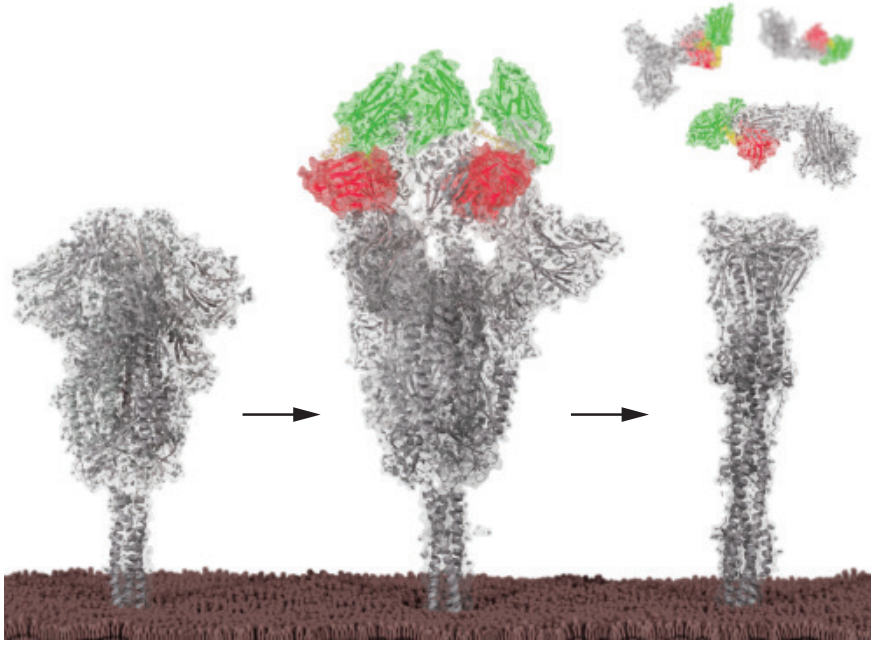
Paul-Albert Koenig*, Hrishikesh Das†, Hejun Liu†, Beate M. Kümmerer, Florian N. Gohr‡, Lea-Marie Jenster‡, Lisa D. J. Schiffelers‡, Yonas M. Tesfamariam‡, Miki Uchima‡, Jennifer D. Wuerth‡, Karl Gatterdam, Natalia Ruetalo, Maria H. Christensen, Caroline I. Fandrey, Sabine Normann, Jan M. P. Tödtmann, Steffen Pritzl, Leo Hanke, Jannik Boos, Meng Yuan, Xueyong Zhu, Jonathan L. Schmid-Burgk, Hiroki Kato, Michael Schindler, Ian A. Wilson, Matthias Geyer, Kerstin U. Ludwig, B. Martin Hällberg*, Nicholas C. Wu*, Florian I. Schmidt*

INTRODUCTION: The global scale and rapid spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pose unprecedented challenges to society, health care systems, and science. In addition to effective and safe vaccines, passive immunization by antibody-related molecules offers an opportunity to harness the vertebrate immune system to fight viral infections in high-risk patients. Variable domains of heavy-chain–only antibodies (VHHs), also known as nanobodies, are suitable lead molecules in such efforts, as they are small, extremely stable, easy to engineer, and economic to produce in simple expression systems.

RATIONALE: We engineered improved multivalent nanobodies neutralizing SARS-CoV-2 on the basis of two principles: (i) detailed structural information of their epitopes and binding modes to the viral spike protein and (ii) mechanistic insights into viral fusion with cellular membranes catalyzed by the spike.

RESULTS: Nanobodies specific for the receptor binding domain (RBD) of SARS-CoV-2 spike were identified by phage display using nanobody libraries from an alpaca and a llama immunized with the RBD and inactivated virus. Four of the resulting nanobodies—VHHs E, U, V, and W—potently neutralize SARS-CoV-2 and SARS-CoV-2–pseudotyped vesicular stomatitis virus. X-ray crystallography revealed that the nanobodies bind to two distinct epitopes on the RBD, interfaces "E" and "UVW," which can be synergistically targeted by combinations of nanobodies to inhibit infection. Cryo–electron microscopy (cryo-EM) of trimeric spike in complex with VHH E and VHH V revealed that VHH E stabilizes a conformation of the spike with all three RBDs in the "up" conformation (3-up), a state that is typically associated with activation by receptor binding. In line with this observation, we found that VHH E triggers the fusion activity of spike in the absence of the cognate receptor ACE2. VHH V, by contrast, stabilizes spike in a 2-up conformation and does not induce fusion. On the basis of the structural information, we designed bi- and trivalent nanobodies with improved neutralizing properties. VHH EEE most potently inhibited infection, did not activate fusion, and likely inactivated virions by outcompeting interaction of the virus with its receptor. Yet evolution experiments revealed emergence of escape mutants in the spike with single–amino acid changes that were completely insensitive to inhibition by VHH EEE. VHH VE also neutralized more efficiently than VHH E or VHH V alone; stabilized the 3-up conformation of spike, as determined by cryo-EM; and more strongly induced the spike fusogenic activity. We conclude that the premature activation of the fusion machinery on virions was an unexpected mechanism of neutralization, as enhanced neutralization could not be attributed simply to better blocking of virus-receptor interactions. Activation of spike in the absence of target membranes likely induces irreversible conformational changes to assume the energetically favorable postfusion conformation without catalyzing fusion per se. Simultaneous targeting of two independent epitopes by VHH VE largely prevented the emergence of resistant escape mutants in evolution experiments.

CONCLUSION: Our results demonstrate the strength of the modular combination of nanobodies for neutralization. Premature activation of spike by nanobodies reveals an unusual mode of neutralization and yields insights into the mechanism of fusion. ■



Bivalent nanobodies neutralize by inducing postfusion conformation of the SARS-CoV-2 spike. On virions, SARS-CoV-2 spike trimers are mostly in an inactive configuration with all RBDs in the down conformation (left). Binding of bivalent nanobody VE stabilizes the spike in an active conformation with all RBDs up (middle), triggering premature induction of the postfusion conformation, which irreversibly inactivates the spike protein (right).

The list of author affiliations is available in the full article online.
*Corresponding author. Email: pakoenig@uni-bonn.de (P.-A.K.); martin.hallberg@ki.se (B.M.H.); nicwu@illinois.edu (N.C.W.); fschmidt@uni-bonn.de (F.I.S.)
†These authors contributed equally to this work.
‡These authors contributed equally to this work.
This is an open-access article distributed under the terms of the Creative Commons Attribution license (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Cite this article as P.-A. Koenig *et al.*, *Science* 371, eabe6230 (2021). DOI: 10.1126/science.abe6230

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abe6230

RESEARCH ARTICLE SUMMARY

MOLECULAR MOTOR

Germanium nanospheres for ultraresolution picotensiometry of kinesin motors

Swathi Sudhakar, Mohammad Kazem Abdosamadi, Tobias Jörg Jachowski, Michael Bugiel, Anita Jannasch, Erik Schäffer*

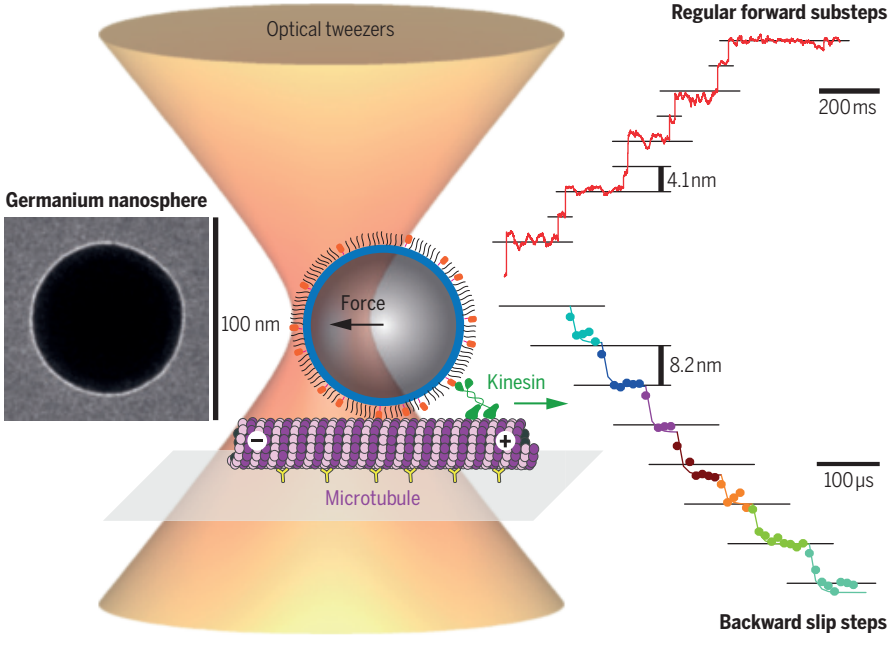
INTRODUCTION: Cytoskeletal motors drive many essential mechanical processes inside cells. For example, kinesin motors are key for cell division or vesicle transport. Kinesin-1 transports cargo along microtubules by coupling adenosine 5′-triphosphate (ATP) hydrolysis to perform mechanical work against piconewton loads. This force generation and overall transport distance are limited by motor detachment. However, how kinesins walk and detach is still unclear.

To simultaneously measure the nanoscale motion and forces of molecular machines, optical tweezers are often used. In the tight, mostly infrared laser focus of optical tweezers, small dielectric particles can be trapped and used as handles for sensitive position and force measurements. Because optical forces scale with the particle volume, piconewton force measurements with nanoparticles require a high laser power. This high power leads to excessive heating and precludes biological measurements. Therefore, for biological single-molecule measurements, micrometer-sized probes are used. However, such probes have a large hydrodynamic drag and therefore lack the spatiotemporal resolution to unravel important fast or small details in the mechanochemistry of molecular machines. These details remain hidden in the storm of Brownian motion.

RATIONALE: To overcome this practical resolution limit of optical tweezers and resolve so-far hidden conformational changes of proteins, we sought to compensate the volume scaling of optical trapping forces by the use of probe materials with a very high refractive index and low light absorption. This compensation should allow the use of nanometer-sized probes and the generation of piconewton optical forces without detrimental heating but with improved temporal response and spatial precision. Promising materials include silicon and germanium that become transparent in the near-infrared, with very high refractive indices exceeding 4. However, efficient methods to fabricate such semiconductor nanospheres suitable for optical trapping do not exist.

RESULTS: We developed a solution-based method to synthesize germanium nanospheres. With a diameter of roughly 70 nm, they are about an order of magnitude smaller as compared with commonly used microspheres and still allowed piconewton force measurements. To find out how kinesin works mechanically, we developed an in vitro reconstituted assay. To this end, we coated the nanospheres with a lipid bilayer [the white rim in the transmission electron microscopy (TEM) image] to mimic vesicles and to be roughly their size inside cells. When we bound kinesin-1 to these "vesicles" and measured the interaction of single motors with microtubules under piconewton tension, we discovered that each hydrolysis cycle is broken up into two 4-nm center-of-mass substeps. The durations of these substeps alternated in their force and ATP dependence, with the duration of one of the substeps being nearly independent of both parameters. Furthermore, when subjected to hindering loads, motors never detached from the microtubule. Instead, motors slipped along the microtubule in 8-nm steps on microsecond time scales. These slip steps are consistent with a bond-rupture model that involves protein friction between the motor and its track. Unexpectedly, motors usually did not detach after a slip event but reengaged in motility that rescued cargo transport.

CONCLUSION: Germanium nanospheres are promising for bioimaging, sensing, optoelectronics, nanophotonics, and energy storage. For optical trapping, the nanospheres open a new temporal window by which to uncover hidden dynamics in molecular machines. The direct observation of load-bearing kinesin substeps resolves a long-standing controversy. Slipping and rescues should allow load distribution and synchronization when motors operate in teams. Understanding their mechanochemistry is important for a better understanding of cellular transport and other essential molecular functions of kinesins, with implications, for example, for neurodegenerative diseases and cancer. ■



Ultraresolution kinesin traces with optically trapped germanium nanospheres. Kinesins are molecular machines that transport vesicles along microtubules inside cells. Membrane-coated germanium nanospheres (TEM micrograph, left) improved the spatiotemporal resolution of optical tweezers and allowed the measurement of substeps during the normal kinesin stepping cycle. Under load, kinesins did not detach but slipped along the microtubule, which led to the discovery of rescues for vesicle transport.

Eberhard Karls Universität Tübingen, Cellular Nanoscience (ZMBP), Auf der Morgenstelle 32, 72076 Tübingen, Germany.
***Corresponding author. Email: erik.schaeffer@uni-tuebingen.de**
Cite this article as S. Sudhakar *et al.*, *Science* 371, eabd9944 (2021). DOI: 10.1126/science.abd9944

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abd9944

RESEARCH ARTICLE SUMMARY

HEART VALVE DISEASE

Network-based screen in iPSC-derived cells reveals therapeutic candidate for heart valve disease

Christina V. Theodoris, Ping Zhou, Lei Liu, Yu Zhang, Tomohiro Nishino, Yu Huang, Aleksandra Kostina, Sanjeev S. Ranade, Casey A. Gifford, Vladimir Uspenskiy, Anna Malashicheva, Sheng Ding, Deepak Srivastava*

INTRODUCTION: Determining the gene-regulatory networks that drive human disease allows the design of therapies that target the core disease mechanism rather than merely managing symptoms. However, small molecules used as therapeutic agents are traditionally screened for their effects on only one to several outputs at most, from which their predicted efficacy on the disease as a whole is extrapolated. In silico correlation of disease network dysregulation with pathways affected by molecules in surrogate cell types is limited by the relevance of the cell types used and by not directly testing compounds in patient cells.

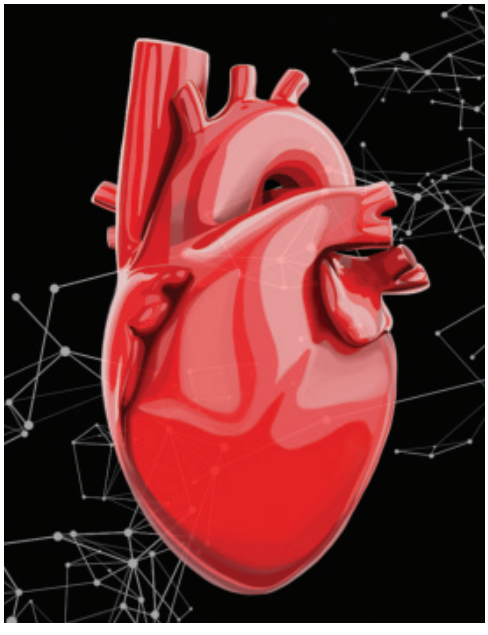
RATIONALE: In principle, mapping the architecture of the dysregulated network in disease-relevant cells differentiated from patient-derived induced pluripotent stem cells (iPSCs) and subsequent screening for small molecules that broadly correct the abnormal gene network could overcome this obstacle. Specifically, targeting normalization of the core regulatory elements that drive the disease process, rather than correction of peripheral downstream effectors that may not be disease modifying, would have the greatest likelihood of therapeutic success. We previously demonstrated that haploinsufficiency of NOTCH1 can cause calcific aortic valve disease (CAVD), the third most common form of heart disease, and that the underlying mechanism involves derepression of osteoblast-like gene networks in cardiac valve cells. There is no medical therapy for CAVD, and in the United States alone, >100,000 surgical valve replacements are performed annually to relieve obstruction of blood flow from the heart. Many of these occur in the setting of a congenital aortic valve anomaly present in 1 to 2% of the population in which the aortic valve has two leaflets (bicuspid) rather than the normal three leaflets (tricuspid). Bicuspid valves in humans can also be caused by NOTCH1 mutations and predispose to early and more aggressive calcification in adulthood. Given that valve calcification progresses with age, a medical therapy that could slow or even arrest progression would have tremendous impact.

RESULTS: We developed a machine-learning approach to identify small molecules that sufficiently corrected gene network dysregulation in NOTCH1-haploinsufficient human iPSC-derived endothelial cells (ECs) such that they classified similar to *NOTCH1*$^{+/+}$ ECs derived from gene-corrected isogenic iPSCs. We screened 1595 small molecules for their effect on a signature of 119 genes representative of key regulatory nodes and peripheral genes from varied regions of the inferred NOTCH1-dependent network, assayed by targeted RNA sequencing (RNA-seq). Overall, eight molecules were validated to sufficiently correct the network signature such that *NOTCH1*$^{+/-}$ ECs classified as *NOTCH1*$^{+/+}$ by the trained machine-learning algorithm. Of these, XCT790, an inverse agonist of estrogen-related receptor α (ERRα), had the strongest restorative effect on the key regulatory nodes *SOX7* and *TCF4* and on the network as a whole, as shown by full transcriptome RNA-seq.

Gene network correction by XCT790 generalized to human primary aortic valve ECs derived from explanted valves from >20 patients with nonfamilial CAVD. XCT790 was effective in broadly restoring dysregulated genes toward the normal state in both calcified tricuspid and bicuspid valves, including the key regulatory nodes *SOX7* and *TCF4*.

Furthermore, XCT790 was sufficient to prevent as well as treat already established aortic valve disease in vivo in a mouse model of Notch1 haploinsufficiency on a telomere-shortened background. XCT790 significantly reduced aortic valve thickness, the extent of calcification, and echocardiographic signs of valve stenosis in vivo. XCT790 also reduced the percentage of aortic valve cells expressing the osteoblast transcriptional regulator RUNX2, indicating a reduction in the osteogenic cell fate switch underlying CAVD. Whole-transcriptome RNA-seq in treated aortic valves showed that XCT790 broadly corrected the genes dysregulated in Notch1-haploinsufficient mice with shortened telomeres, and that treatment of diseased aortic valves promoted clustering of the transcriptome with that of healthy aortic valves.

CONCLUSION: Network-based screening that leverages iPSC and machine-learning technologies is an effective strategy to discover molecules with broadly restorative effects on gene networks dysregulated in human disease that can be validated in vivo. XCT790 represents an entry point for developing a much-needed medical therapy for calcification of the aortic valve, which may also affect the highly related and associated calcification of blood vessels. Given the efficacy of XCT790 in limiting valve thickening, the potential for XCT790 to alter the progression of childhood, and perhaps even fetal, valve stenosis also warrants further study. Application of this strategy to other human models of disease may increase the likelihood of identifying disease-modifying candidate therapies that are successful in vivo. ■



Network-correcting therapeutic candidate for heart disease. A gene network–based screening approach leveraging human disease-specific iPSCs and machine learning identified a therapeutic candidate, XCT790, which corrected the network dysregulation in genetically defined iPSC-derived endothelial cells and primary aortic valve endothelial cells from >20 patients with sporadic aortic valve disease. XCT790 was also effective in preventing and treating a mouse model of aortic valve disease.

ILLUSTRATION: CHRISTINA V. THEODORIS

The list of author affiliations is available in the full article online.
*Corresponding author. Email: dsrivastava@gladstone.ucsf.edu
Cite this article as: C. V. Theodoris *et al.*, *Science* **371**, eabd0724 (2021). DOI: 10.1126/science.abd0724

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abd0724

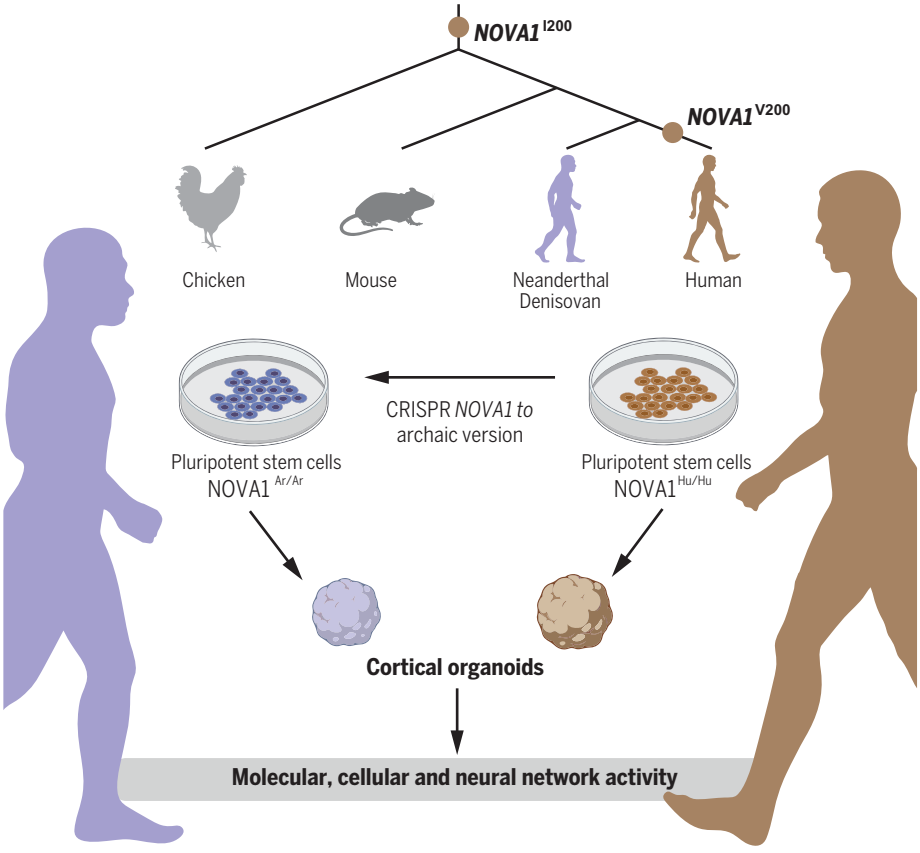
RESEARCH ARTICLE SUMMARY

HUMAN EVOLUTION

Reintroduction of the archaic variant of *NOVA1* in cortical organoids alters neurodevelopment

Cleber A. Trujillo, Edward S. Rice, Nathan K. Schaefer, Isaac A. Chaim, Emily C. Wheeler, Assael A. Madrigal, Justin Buchanan, Sebastian Preissl, Allen Wang, Priscilla D. Negraes, Ryan A. Szeto, Roberto H. Herai, Alik Huseynov, Mariana S. A. Ferraz, Fernando S. Borges, Alexandre H. Kihara, Ashley Byrne, Maximillian Marin, Christopher Vollmers, Angela N. Brooks, Jonathan D. Lautz, Katerina Semendeferi, Beth Shapiro, Gene W. Yeo, Stephen E. P. Smith, Richard E. Green, Alysson R. Muotri*

INTRODUCTION: Current views of human evolution, as supported by the fossil record, indicate that many hominin lineage branches arose, but only one survived to the present. Neanderthals and Denisovans, two of these extinct lineages, are our closest evolutionary relatives and therefore provide the most subtle genetic and phenotypic contrast to our species. Comparison of Neanderthal, Denisovan, and extant human genomes has shown that many humans today carry genes introduced through past admixture events and has allowed enumeration of human-specific genetic differences that may have been important for recent human evolution. Neuro-oncological ventral antigen 1 (*NOVA1*) includes one of the few protein-coding differences between modern human and archaic hominin genomes that could affect human neurodevelopment.



***NOVA1* archaic variant in cortical organoids affects cellular, molecular, and neural network activity profiles.** During human evolution, modern humans acquired a specific nucleotide substitution in the RNA binding domain of *NOVA1*. Using genome editing technology, the archaic version of *NOVA1* was introduced in human iPSCs and differentiated into cortical organoids. The changes could be observed in different levels, from altered proliferation, different gene expression, and splicing profiles to modified glutamatergic synapsis and neuronal network connectivity. $NOVA1^{I200}$, *NOVA1* gene containing isoleucine archaic variant. $NOVA1^{V200}$, *NOVA1* gene containing valine human variant; $NOVA1^{Ar/Ar}$, human cell line with homozygous reintroduction of the *NOVA1* archaic variant; $NOVA1^{Hu/Hu}$, human cell line with *NOVA1* human variant. Figure created with BioRender.com.

RATIONALE: NOVA1 regulates alternative splicing in the developing nervous system and is a master regulator of splicing genes responsible for synapse formation. Altered NOVA1 splicing activity in humans is associated with neurological disorders, underscoring the role of NOVA1 in neural function. Using CRISPR-Cas9 genome-editing technology in human induced pluripotent stem cells (iPSCs), we replaced the modern human allele of the *NOVA1* gene with the ancestral allele found in Neanderthals and Denisovans, which contains a single-nucleotide substitution at position 200 that causes an isoleucine-to-valine change. To investigate the functional importance of this amino acid change in humans, we followed iPSC neural development through functional cortical organoids.

RESULTS: The reintroduction of the archaic version of *NOVA1* into a human genetic background causes changes in alternative splicing in genes involved in neurodevelopment, proliferation, and synaptic connectivity. These changes co-occur with differences in organoid morphology and neural network function, suggesting a functional role for the derived human-specific substitution in *NOVA1*. Furthermore, cortical organoids carrying the archaic *NOVA1* displayed distinct excitatory synaptic changes, which may have led to the observed alterations in neural network development. Collectively, our data suggest that expression of the archaic *NOVA1* leads to modified synaptic protein interactions, affects glutamatergic signaling, underlies differences in neuronal connectivity, and promotes higher heterogeneity of neurons regarding their electrophysiological profiles.

CONCLUSION: A subset of genetic changes may underly the phenotypic traits that separate our species from these extinct relatives. We developed a platform to test the impact of human-specific genetic variants by reintroducing the archaic form found in Neanderthals and Denisovans and measuring its effects during neurodevelopment using human brain organoids. Our results suggest that the human-specific substitution in *NOVA1*, which became fixed in modern humans after divergence from Neanderthals, may have had functional consequences for our species' evolution. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: muotri@ucsd.edu
Cite this article as C. A. Trujillo *et al.*, *Science* **371**, eaax2537 (2021). DOI: 10.1126/science.aax2537

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.aax2537

RESEARCH ARTICLE SUMMARY

NEUROSCIENCE

Cellular transcriptomics reveals evolutionary identities of songbird vocal circuits

Bradley M. Colquitt*, Devin P. Merullo*, Genevieve Konopka†, Todd F. Roberts†, Michael S. Brainard†

INTRODUCTION: The mammalian neocortex, with its distinctive six-layered structure, is thought to enable advanced cognitive functions not seen in other animals. Yet birds, which have a remarkably different brain organization, display a range of complex motor and cognitive abilities, such as tool use and problem-solving, that are comparable to those of many mammals. Although portions of the avian brain are often compared to the neocortex, especially regions involved in the learning and production of vocalizations, it has remained unclear whether these regions are truly homologous with the neocortex (that is, whether they share a common evolutionary origin) or instead are examples of evolutionary convergence.

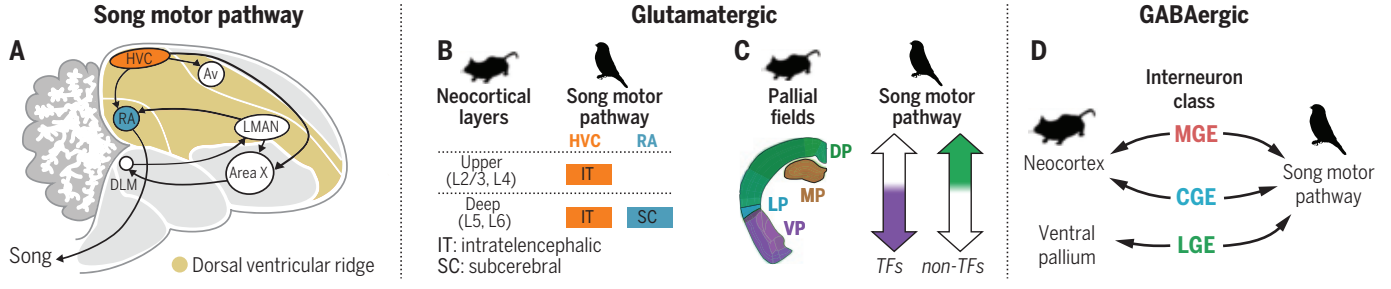
RATIONALE: The nature of the similarities and differences in brain organization between mammals and birds has implications for the evolutionary mechanisms that underlie the emergence of advanced behaviors. In mammals, the six-layered neocortex occupies most of the pallium (the outermost portion of the brain), whereas in birds, most of the pallium consists of a distinct unlayered structure called the dorsal ventricular ridge (DVR). The DVR contains multiple interconnected groups of neurons, often referred to as nuclei, that are necessary for complex avian behaviors, including vocal learning in songbirds. Two prevailing viewpoints offer opposing interpretations for how both the mammalian neocortex and avian DVR enable complex behaviors despite their structural differences. One view proposes that the DVR is homologous to the neocortex and that the nuclei in the DVR correspond to distinct layers of the neocortex and thus represent a rearrangement of a conserved ancestral circuit. A second hypothesis argues that the neocortex and DVR develop from distinct embryonic regions of the pallium (dorsal and ventral, respectively) and are therefore nonhomologous structures that separately evolved to serve similar functions. To test these models, we used single-cell transcriptomics to characterize the cell types and gene expression patterns of two regions in the songbird DVR that are necessary for learning and producing birdsong: HVC (proper name) and RA (robust nucleus of the arcopallium). For each type, we characterized the expression profiles of transcription factors, which reflect the cellular identities and regional origins of neurons, and effector genes, which specify neuronal cell properties and function. We compared these profiles with those of neurons previously described in mammals and reptiles to clarify how individual neuronal types are related across amniotes.

RESULTS: We identify a variety of excitatory cell classes that are different between HVC and RA, and inhibitory classes that are shared across regions, similar to organizational patterns of cell types in mammals and reptiles. We show that excitatory neurons in both HVC and RA have transcription factor profiles that bear strong similarity to the mammalian ventral pallium, which includes the olfactory bulb, piriform cortex, and pallial amygdala, but not to the neocortex, which develops from the dorsal pallium. However, when examining only effector genes, we find that excitatory neurons exhibit greater similarity to neocortical neurons from multiple layers and less similarity to the ventral pallium. We also find that songbird inhibitory neurons bear a considerable resemblance to neuron classes in both mammals and turtles, indicating that the major classes of inhibitory neurons are conserved and likely present in ancestral amniotes. We report that, consistent with the interpretation that song-control regions have ventral pallial origins, the most abundant inhibitory neuron type in the songbird DVR is similar to inhibitory neurons that are enriched in mammalian ventral pallial derivatives and absent from the neocortex.

CONCLUSION: Our findings indicate that the avian DVR and the neocortex derive from different neurodevelopmental regions employing distinct transcription factor expression patterns and therefore are not homologous structures. However, we find that excitatory neurons in the DVR have evolved similar properties to the neocortex by engaging overlapping patterns of effector genes. Such overlapping transcriptional profiles may account for the evolution of similar complex motor and cognitive abilities in mammals and birds, including vocal learning, and suggest that the DVR may perform neural computations in a way that is functionally analogous to the neocortex. By addressing a long-standing controversy regarding the relationship between avian and mammalian brains, these results provide insight into the evolution and diversification of neural cell types and structures that enable advanced behaviors. ■

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: genevieve.konopka@utsouthwestern.edu (G.K.); todd.roberts@utsouthwestern.edu (T.F.R.); michael.brainard@ucsf.edu (M.S.B.)
Cite this article as B. M. Colquitt *et al.*, *Science* **371**, eabd9704 (2021). DOI: 10.1126/science.abd9704

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abd9704



Cellular transcriptomics of a songbird vocal circuit. (**A**) Schematic of the song motor pathway (SMP). HVC, proper name; RA, robust nucleus of the arcopallium; DLM, medial portion of the dorsolateral thalamic nucleus; Av, mesopallial auditory nucleus Avalanche; LMAN, lateral magnocellular nucleus of the anterior nidopallium. (**B**) Transcriptional similarities between glutamatergic neurons in the SMP and the mouse neocortex. (**C**) Pallial biases of transcription factor (TF) versus effector gene (non-TF) expression profiles. DP/MP/LP/VP, dorsal/medial/lateral/ventral pallium. (**D**) Diversity and origins of γ-aminobutyric acid–releasing (GABAergic) neurons in the SMP. LGE/MGE/CGE, lateral/medial/caudal ganglionic eminence.

RESEARCH ARTICLES

CRIMINAL JUSTICE

The role of officer race and gender in police-civilian interactions in Chicago

Bocar A. Ba[1], Dean Knox[2]*, Jonathan Mummolo[3]*, Roman Rivera[4]

Diversification is a widely proposed policing reform, but its impact is difficult to assess. We used records of millions of daily patrol assignments, determined through fixed rules and preassigned rotations that mitigate self-selection, to compare the average behavior of officers of different demographic profiles working in comparable conditions. Relative to white officers, Black and Hispanic officers make far fewer stops and arrests, and they use force less often, especially against Black civilians. These effects are largest in majority-Black areas of Chicago and stem from reduced focus on enforcing low-level offenses, with greatest impact on Black civilians. Female officers also use less force than males, a result that holds within all racial groups. These results suggest that diversity reforms can improve police treatment of minority communities.

Racial disparities in police-civilian interactions and high-profile incidents of excessive force continue to fuel allegations of abusive and discriminatory policing (*1*, *2*). Central to these critiques are the fact that throughout the history of policing in the United States, many police forces have been nearly all white and male (*3*). In turn, some of the most frequently proposed reforms aimed at reducing inequities and police brutality have centered on hiring more nonwhite (*4*) and female (*5*) officers. One agency that has undergone substantial diversification in recent decades is the Chicago Police Department (CPD), transforming from a mostly white and nearly all male force to one in which half of sworn officers are minorities and over one-fifth are female. This heterogeneity across race and gender lines, combined with newly acquired data on officers' daily patrols and enforcement activities, allows a thorough assessment of the practical consequences of diversity in law enforcement. Although we cannot directly infer the future impact of further diversification, we can examine the Chicago case in depth to provide the most credible microlevel evidence to date on the treatment civilians can expect when encountering officers of varied racial, ethnic, and gender identities.

Theories of social distance and intergroup relations in a range of contexts (*6*–*9*) imply that diversifying police agencies may improve the treatment of minorities (*3*, *8*). Individuals rely on stereotypes when evaluating members of social groups (*10*) and are thought to be less likely to engage in harassment toward in-group members (*11*). However, research on organizational culture and bureaucratic politics suggests that officers of different social backgrounds may ultimately behave similarly because of self-selection into service and socialization during training and on the job (*3*, *12*–*16*). To succeed and advance, women and minorities may also face pressure to adopt conventional enforcement practices (*3*, *17*, *18*).

Rigorous evaluation of the effects of police diversity has been stymied by a lack of sufficiently fine-grained data on officer deployment and behavior that makes it difficult or impossible to ensure that officers being compared are facing common circumstances while on duty. Studies typically rely on coarse geographic units, like agency- or precinct-level data (*19*–*21*), which forced previous scholars to invoke the strong assumption that, for example, "white and nonwhite officers are randomly assigned to neighborhoods" (*20*, p. 389). Furthermore, most policing data sets contain records of enforcement events only [e.g., logs of stops or arrests (*22*–*24*)]; events in which officers choose to take no action are unobserved, potentially distorting inferences. Other studies that make valid comparisons are often limited in scope to particular activities, like ticketing during traffic accident investigations (*25*). And although some prior work has leveraged the timing of diversity reforms to estimate agency-level effects (*26*–*29*), those aggregated approaches are by design unable to examine details of police-civilian interactions. Findings with regard to racial diversity in particular have been decidedly mixed: In an exhaustive review of the empirical literature, one prominent legal scholar concluded, "[t]he fairest summary of the evidence is probably that we simply do not know" (*30*).

To assess the impact of diversity in law enforcement, we draw on newly collected data, assembled through years of open-records requests, that allow us to overcome long-standing limitations. These include officer demographics, language skills, daily shift assignments, and career progression. We link these files to time-stamped, geolocated records of the same officers' decisions to stop, arrest, and use force against civilians. After aggressively pruning data to maximize analytic validity, we compile a panel of 2.9 million officer shifts and 1.6 million enforcement events by nearly 7000 officers covering the years 2012 through 2015. Most notably, we leverage fine-grained information on daily patrol assignments, which vary exogenously on the basis of fixed rules and preassigned rotations, to examine how officers of different groups behave when faced with comparable circumstances and civilian behaviors.

The deployment effects that we estimate are a critical first step in the systematic evaluation of widely proposed personnel reforms, which have historically focused on increasing racial and gender diversity among officers. If officers of different demographic profiles do not behave differently when faced with the same conditions, there is little hope that diversifying police agencies will yield tangible differences in the treatment of marginalized civilians. Indeed, we demonstrate that deploying officers of different demographic profiles to comparable environments does produce large differences in how police treat civilians. However, we caution that these deployment effects do not directly generalize to future effects of hiring reforms, for several reasons. Chief among these are that (i) the nature of police-civilian interactions is changing rapidly; (ii) racial, ethnic, and gender differences in current officers' behavior may not map perfectly to those of future cohorts; (iii) deployment patterns will necessarily change as more officers are hired from marginalized groups; and (iv) diversification reforms may exert additional, potentially powerful second-order effects, e.g., through agency culture.

Chicago as a case study

Our focus on one city provides unusually detailed data at the expense of geographic scope. Chicago is a large and racially diverse metropolis, with roughly half of residents identifying as nonwhite. Chicago is also heavily segregated, has a history of racial tensions between residents and police, and has come under recent scrutiny for controversial aggressive policing tactics such as "stop and frisk" (*31*). The agency received national attention for the 2014 killing of 17-year-old Laquan McDonald, an attempted cover-up, and ensuing social unrest (*32*). The CPD was condemned for its "code of silence" (*33*), and then-superintendent Garry McCarthy received widespread criticism for "encouraging

[1]Department of Economics, University of California Irvine, Irvine, CA, USA. [2]Operations, Information and Decisions Department, Wharton School, University of Pennsylvania, Philadelphia, PA, USA. [3]Department of Politics, School of Public and International Affairs, Princeton University, Princeton, NJ, USA. [4]Department of Economics, Columbia University, New York, NY, USA.
*Corresponding author. Email: dcknox@upenn.edu (D.K.); jmummolo@princeton.edu (J.M.)

the kind of aggressive cop culture under which McDonald's shooting took place" (*34*).

As (*26*) recounts, before a series of lawsuits in the 1970s, the CPD was slightly less than 20% Black, in a city that was one-third Black in 1970. The Afro-American Patrolmen's League (AAPL) filed a discrimination suit "on hiring, promotion, assignment, and discipline" (*35*), with the Department of Justice (DOJ) soon joining it (*26*). In the early 1970s, Black hiring shares were around 10%. In 1974, hiring quotas were imposed, and Black hiring shares increased to 40% by 1975 (*26*). These reforms had broader effects on CPD's composition of the department; women made up a larger proportion of Black recruits, with white women lagging in hiring and promotion at first (*36*). As of December 2016, roughly 22% of officers identified as Black, 23% as Hispanic, and 3% Asian; and 22% are female, a stark change from its 99% male workforce in 1970. (Text S1.1 discusses racial and ethnic classification of CPD officers; text S2.1 and fig. S1 provide additional details about the CPD's evolution.)

From one standpoint, it may be difficult to extrapolate from Chicago to settings lacking these racial tensions and history of diversification. But in other ways, it is these very conditions that make Chicago an important test case: Among major departments nationwide, it is arguably one in which reform has historically been sorely needed. A single case study cannot be the final word in an important debate. But Chicago offers an invaluable opportunity to study diversity in policing using unusually fine-grained data, in a setting where concerns over racial inequity are pronounced.

Data

We submitted a series of open-records requests and appeals to the CPD, the city's Department of Human Resources, and the Illinois Office of the Attorney General over 3 years, seeking data on officer demographics and behavior. The resulting records include the name, race, gender, language skill, unit assignments, and appointment date of each officer (*37*, *38*). We also obtained officers' stops, arrests, and uses of force, which we merged with daily patrol assignments and U.S. Census data, per text S1.2 and S1.3. Table 1 reports aggregate counts. Owing to sparse data on other groups, our analysis is limited to Black, Hispanic, and white officers (97% of officers in the sample). Stops and arrests are recorded in officer-shift data once per officer contributing to enforcement. In stop records, one is listed as "first" officer, suggesting a leading role, although arrest records contain no such labeling. (Text S3.9 and fig. S2 conduct additional analyses of first officers only, yielding a highly similar pattern of results.)

Figure 1 depicts a small slice of the data's temporal and geographic coverage: a 3-month window in CPD's Wentworth District (District 2), a highly segregated 7.5-square-mile territory on Chicago's South Side that is 95% Black and consistently ranks among the city's most violent districts in per-capita crime. The district spans 15 patrol areas, shaded according to racial composition. Points indicate geolocated stops, arrests, and uses of force during this period. The figure also offers a detailed portrait of four anonymized CPD officers working in District 2 in this time. For example, "Officer A" is female, Black, does not speak Spanish, and joined the CPD in 1994; "Officer C" is a white male who joined the CPD in 2006 and does not speak Spanish. The figure shows officers' specific patrol slots and each officer's behavior while on assignment.

Identifying racial, ethnic, and gender disparities in policing

Although the CPD has diversified over time, officer groups face substantially different working conditions. Figure 2 displays the average characteristics of districts—22 geographic regions delineated by the CPD—to which officer groups are assigned. Differences associated with officer race and ethnicity are most stark. In general, Black officers work in districts with 47% higher per-capita violent crime and large co-racial populations—on average 68% co-racial, far higher than the average 26 to 30% co-racial and co-ethnic districts where white and Hispanic officers serve. However, white officers are generally overrepresented relative to the resident population; 20% of the 95%-Black Wentworth District officers are white; in Austin (District 15), where residents are 93% Black, officers are 55% white. (Text S2.2 discusses district organization. Figures S3 to S5 present district-level data; assigned officer demographics somewhat track those of district residents, but officers are disproportionately white.) Even within districts, text S2.3 and figs. S6 and S7 demonstrate that marginalized groups are tasked with patrolling different beats, compared to white or male colleagues. (All p-values < 0.001.)

These patterns underscore a central difficulty in evaluating how officer behavior varies across demographic groups. Namely, white officers work in different environments from minority officers, on average. Men and women also work during different hours of the day (text S2.3 and figs. S9 and S10). This means that after aggregating to large geographic units and time periods, observed behavioral differences may simply reflect differing patrol environments, rather than differences in policing approaches.

To make valid comparisons, we assemble a panel dataset in which rows represent officer-shifts—roughly 8-hour patrol periods—and characterize officers' actions and their context. (Text S1.2 describes these datasets; text S1.4 and S1.5 elaborate on preprocessing.) In each of these 2.9 million patrol assignments, we measure officers' stops, arrests, and uses of force, whether they engaged in any of these activities or not. We compared officers of different demographic profiles working in the same specific combination of month and year (e.g., January 2012), day of week, shift time, and assigned "beat" (a patrol task, typically corresponding to small geographic areas less than one square mile; see text S1.6 for a detailed discussion of beat assignments)—a narrow slice of time and space that we abbreviate "MDSBs" (month, day of week, shift, beat). CPD also assigns officers to "day-off groups," which determine who works on rotating dates according to a scheme set late in each calendar year for the following year, representing a large exogenous source of variation in the officers that are available to serve in a particular patrol assignment on any given date. This procedure greatly mitigates threats from self-selection (e.g., officers choosing to take

Table 1. Summary of data on officer behavior (counts), 2012-2015. Summary statistics after pruning officers, shifts, and event records aggressively to ensure common circumstances in our behavioral analysis.

	Black officers	Hispanic officers	White officers	Female officers	Male officers
Stops	253,576	356,493	729,000	264,526	1,074,543
Arrests	47,396	65,581	132,272	43,625	201,624
Uses of force	1,355	2,081	4,513	1,125	6,824
Shifts	829,818	689,091	1,413,771	740,015	2,192,665
Officers	1,834	1,674	3,439	1,785	5,162

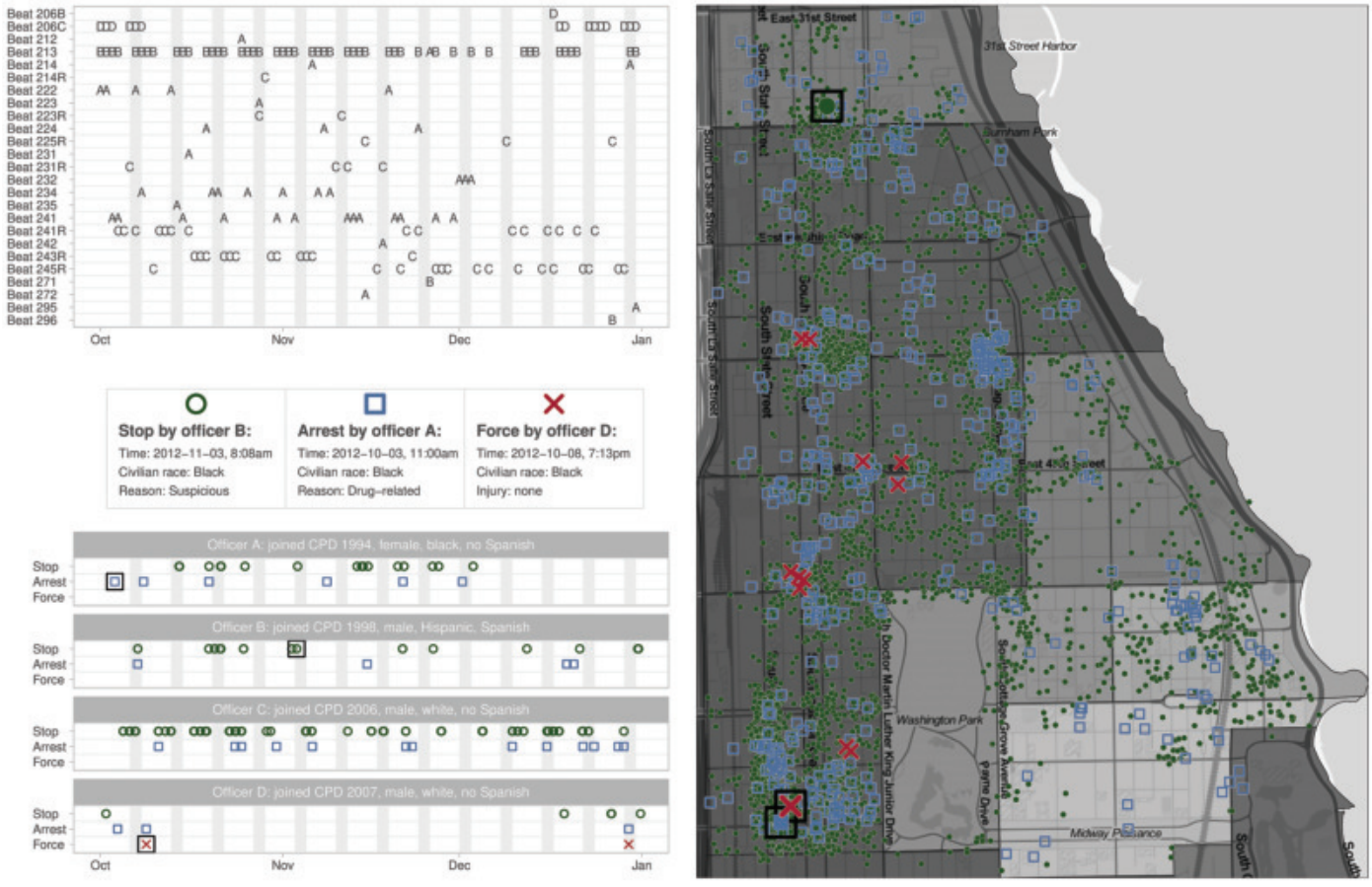


Fig. 1. Detailed view of the data. The right panel maps police activity in a single CPD district (Wentworth, District 2), with green circles, blue squares, and red crosses respectively indicating the locations of stops, arrests, and uses of force. Polygons represent geographic beats and are shaded by their proportion of minority residents. Lower left panels chart the behavior of four anonymized officers over a 3-month period, with panel headers indicating the year officers joined CPD, gender, ethnicity or race, and language ability. Boxed incidents are described further in the left middle panel, which reports civilian and incident specifics. Finally, the top left panel indicates how the four selected officers are assigned to patrol beats over dates and times, with vertical gray bars indicating weekends.

days off when crime spikes), helping to ensure that the officers we compare are facing common circumstances, on average. This design also ensures that comparisons of officer activity have common denominators, as those working in comparable places and times have the same opportunity to take enforcement action. (See text S1.6 and figs. S11 and S12 for further details on CPD shift assignment procedures and related analytic strategy. Text S3.6 and fig. S13 demonstrate how common alternative approaches in prior work can mislead analysts about the magnitude or even the sign of these effects. Text S3.3 examines clock-in/out times by officer race and finds no meaningful differences in shift duration.)

Because our analytic approach relies on comparisons between officers deployed to the same MDSBs, our inferences are limited to MDSBs in which cross-group comparisons are feasible, e.g., in which both Black and white officers are both deployed. (See texts S3.1 and S3.2 for discussions of the data-generating process and statistical estimand, as well as how this analytic strategy circumvents the threats to inference posed by unobserved differences in patrol environments.) Thus, our estimates do not necessarily generalize to every officer, time, or location in Chicago. (Text S3.4 and fig. S8 provide details on the roughly 40% of patrol assignments where cross-group, within-MDSB comparisons are unavailable—typically smaller patrol tasks with fewer assigned officers—because assigned officers are all from the same demographic group. Feasibility does not meaningfully vary with resident racial composition, and nearly every officer rotates through patrol tasks with feasible comparisons.) Officers from different demographic groups also differ in unobserved ways. We therefore estimate the average difference in officer behavior resulting from deploying an officer of one demographic profile—and all the associated traits of that demographic label—relative to another, holding environmental conditions constant. Our results do not reflect the hypothetical effect of changing an officer's race or gender while holding their other traits fixed. Rather, they reflect average differences in treatment that civilians can expect when police commanders assign officers of one demographic group to their temporal and geographic vicinity, compared to another officer group, holding circumstances equal.

We present differences estimated using ordinary least squares with MDSB fixed effects, though our results are robust to several other estimators, including the addition of flexible controls for experience (see text S3 for estimation details and additional results). All statistical inferences are based on officer-level block bootstrap confidence intervals (CIs) that are robust to unobserved officer-specific peculiarities.

Results

Figure 3 displays average differences in the number of stops, arrests, and uses of force by Black and Hispanic officers (relative to white officers) and female officers (relative to male) working in the same MDSBs. Turning first to Black officers, Fig. 3 shows that when faced

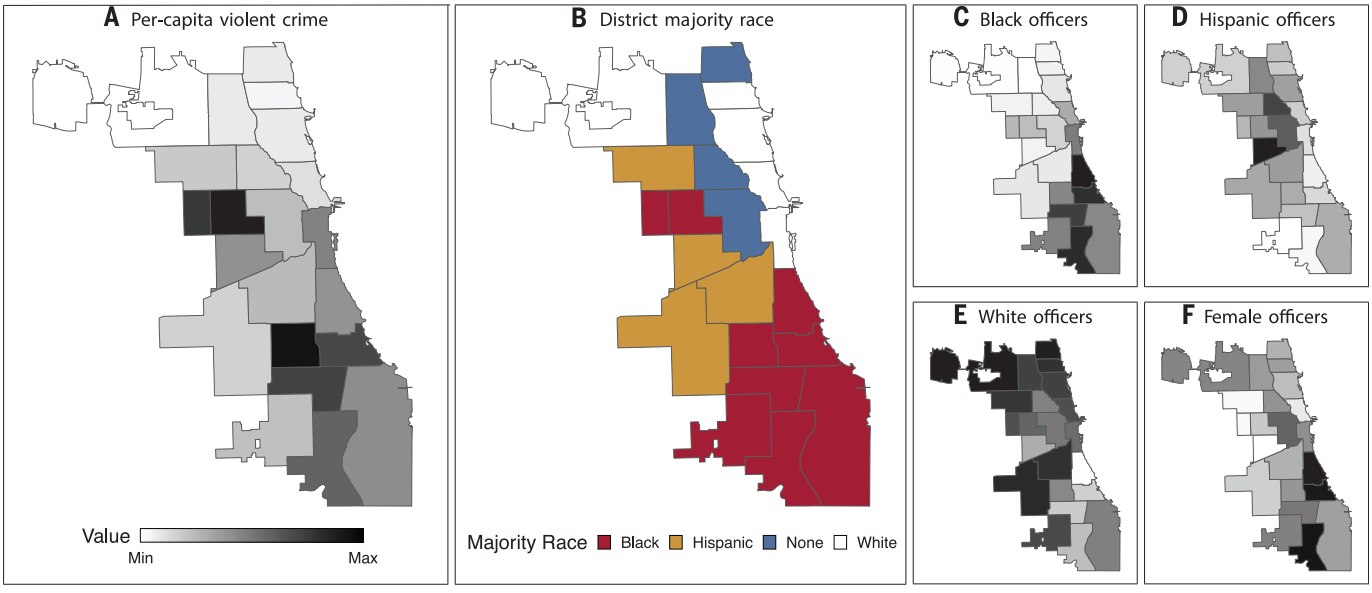


Fig. 2. Unit assignments of various officer groups. (**A** and **B**) Panels depict background information on CPD districts. (**C** to **F**) Panels show, for each district, the proportion of assigned officers belonging to a particular demographic group. Majority race of district residents is based on 2010 decennial Census data; all other plots use 2013–2016 CPD records.

with comparable working conditions over the course of 100 shifts, this group makes 15.16 fewer stops and 1.93 fewer arrests, and it uses force 0.10 fewer times than white counterparts on average—that is, compared to white officers given the same patrol assignment, in the same month, on the same day of the week, and at the same shift time (all $p_{adj} < 0.001$ after Benjamini-Hochberg multiple-testing correction for all cross-group comparisons and all categories and subcategories of enforcement). These gaps are large, representing 29, 21, and 32% of the average stop, arrest, and use-of-force volume for white officers citywide (see tables S1 to S3 for average enforcement activity by officer group. See tables S4 to S8 for full numeric results, including an additional analysis of Spanish language ability).

Figure 4 shows that these disparities are not uniform across situations but are driven by a reduced focus on Black civilians. For example, deploying Black officers instead of white yields 12.55 fewer stops of Black civilians per 100 shifts, a reduction equal to 39% of typical white-officer volume. By contrast, Black officers make only 1.31 fewer stops of white civilians per 100 shifts than their white counterparts (reduction equal to 17% of typical white-officer volume; all adjusted p-values < 0.001). The large differences in these scaled effects (39% versus 17%) suggest that they are not explained by the fact that police engage Black civilians in Chicago more often in general. Put another way, in 100 typical white-officer shifts, Black-civilian stops (32.45) are far more frequent than white-civilian stops (7.53), occurring with a baseline ratio of 4.31. By contrast, when deploying Black officers in lieu of white officers,

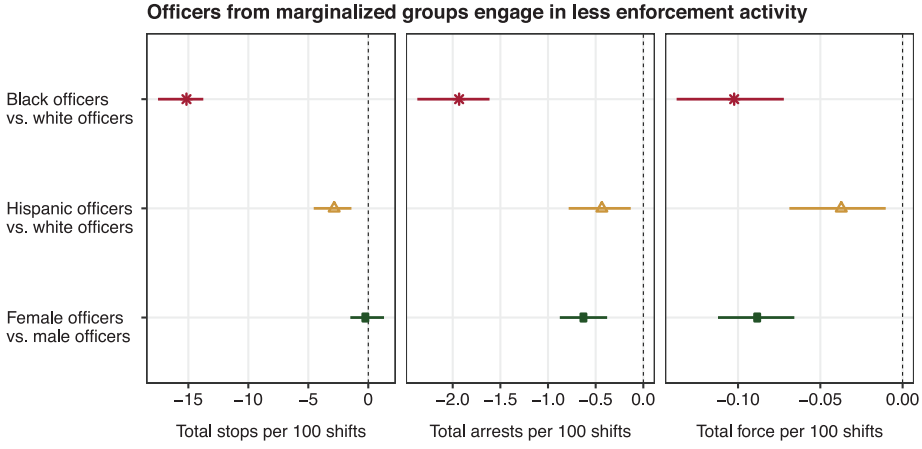


Fig. 3. Effects of deploying officers from marginalized groups on total enforcement activity. Average within-MDSB differences in rates of stops, arrests, and uses of force across officer groups. See tables S4 to S8 for numeric results.

the ratio of the reductions in stops of Black civilians (−12.55) to white civilians (−1.31) is twice as large: 9.60 (95% CI [8.00, 12.98]). Similarly, the ratio of the reductions on Black-civilian arrests (−1.46) to white-civilian arrests (−0.18) is 7.99 (95% CI [5.60, 15.01]). This is significantly larger than the ratio of typical white-officer enforcement volumes (5.90 for Black civilians, 1.17 for white civilians, ratio of 5.03). Black officers also deploy force against Black civilians 0.08 fewer times per 100 shifts than their white counterparts, and they use force that results in injury 0.03 fewer times per 100 shifts (reductions equal to 38 and 39% of typical white-officer volume, respectively; all adjusted p-values < 0.001). Reduced use of force against Black civilians accounts for 83% of the overall force disparity between white and Black officers.

Compared to white officers working in comparable places and times, Black officers also show reduced focus on enforcement activities that are more discretionary in nature. For example, Black officers make 5.72 fewer stops per 100 shifts for "suspicious behavior" (a reduction equal to 31% of average white-officer volume). The reduction resulting from deploying Black officers on drug arrests (−0.31 per 100 shifts) is also estimated to be larger than the effect on violent arrests (−0.23 per 100 shifts). Though the raw effects on drug and violent crime arrest counts are not statistically distinguishable from one another, comparing these effects to typical baseline enforcement

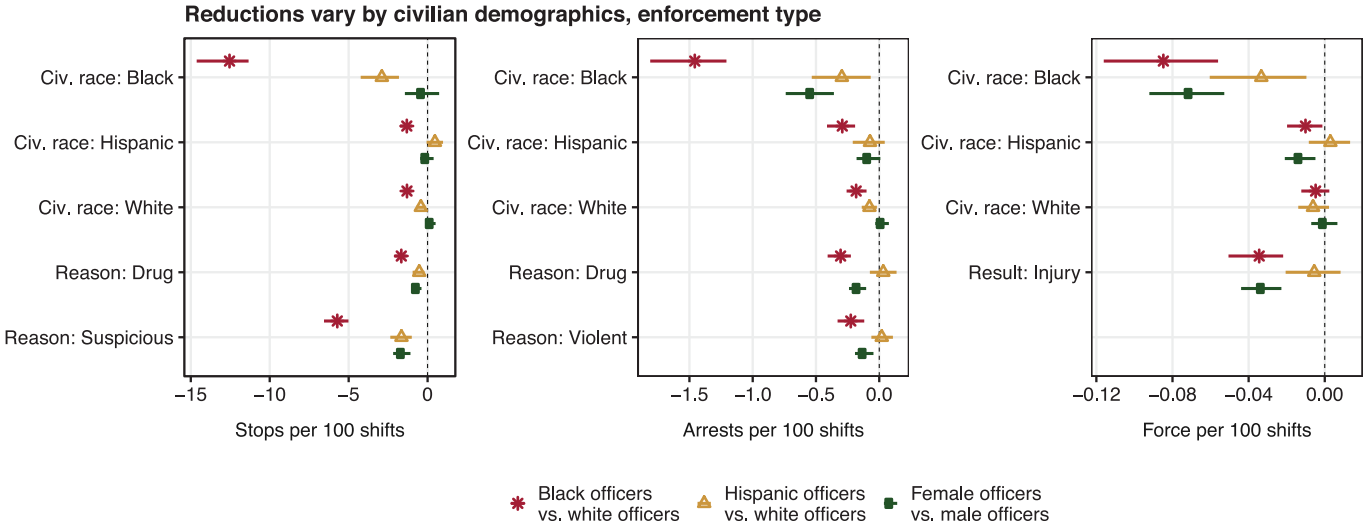


Fig. 4. Effects of deploying officers from marginalized groups, by enforcement subcategory. Average within-MDSB differences in rates of stops, arrests, and uses of force across officer groups, for selected subcategories of each enforcement type. See fig. S14 and tables S4 to S8 for complete subcategory results.

volume puts them in context. The ratio of these reductions, 1.36 (95% CI [0.87, 2.71])—i.e., slightly larger reductions in drug arrests versus violent-crime arrests—is larger than the baseline ratio of drug arrests (1.13 per 100 shifts) to violent arrests (2.16 per 100 shifts) typically made by white officers (ratio of 0.52). (See text S3.5, tables S4 to S8, and fig. S14 for detailed results, including additional enforcement subcategories. Text S3.8 and figs. S15 to S18 show that results obtained with a wide range of alternative estimators are almost identical.)

These patterns are largely in line with the hopes of proponents of racial diversification, who seek to reduce abusive policing and mass incarceration, especially in Black communities.

Like their Black colleagues, Hispanic officers facing the same working conditions conduct fewer stops, make fewer arrests, and use force less than white officers, though the gaps are more modest. Notably, disparities are primarily driven by less engagement with Black civilians; Hispanic officers exhibit nearly the same average volume of enforcement activity against Hispanic civilians as do white officers. Hispanic officers make 2.84 fewer stops per 100 shifts (a reduction equal to 6% of average white-officer volume, p_{adj} = 0.001); 0.44 fewer arrests per 100 shifts (p_{adj} = 0.012, 5%); and 0.04 fewer uses of force per 100 shifts (p_{adj} = 0.021, 12%). We caution that the descriptor "Hispanic" encompasses a range of cultures and national origins that our data do not allow us to parse and that may correspond to important heterogeneity in behavior. (For example, in tables S7 and S8, we show suggestive evidence for differences between Hispanic officers who can and cannot speak Spanish.) More fine-grained data on officers of Hispanic identity are needed to explore this finding.

We also find differences in female officers' behavior relative to male officers, though these are generally smaller in magnitude. Female officers make 0.61 fewer total arrests per 100 shifts (a reduction equal to 7% of average male officer arrests) and 0.54 fewer arrests of Black civilians per 100 shifts (reduction equal to 9% of average male volume, both $p_{adj} < 0.001$). Indeed, about 88% of this disparity in arrest rate is due to reduced arrests of Black civilians. We also find that female officers use force 0.09 fewer times overall (a reduction equal to 28% of average male volume) and 0.07 fewer times per 100 shifts against Black civilians (reduction equal to 31% of average male volume, both $p_{adj} < 0.001$), with the latter accounting for 81% of overall force reduction. (Figure S19 shows that within each racial and ethnic group, female officers use significantly less force than male counterparts.)

Figure 5 displays core results estimated separately in districts where different racial and ethnic resident groups represent majorities. The gap in activity between white and Black officers is most pronounced in majority-Black areas of the city—further evidence that reductions in stops, arrests, and uses of force by Black officers are driven by a reduced focus on Black civilians. Figure S20 illustrates how enforcement differences in these areas are particularly pronounced at night. We see much less heterogeneity across neighborhoods when comparing Hispanic and white officers. (For additional results on gender heterogeneity, see text S3.7 and fig. S19.)

Discussion

Violent and sometimes fatal encounters between white police officers and unarmed racial minorities continue to prompt widespread calls for law enforcement reforms. Protests against police brutality and racial bias remain ubiquitous, including recent unrest in the location of this study, Chicago. Prominent among the many proposed reforms is increasing the level of racial and gender diversity of police agencies. To evaluate the impact of this approach on police-civilian interactions, we leverage unusually rich data on police personnel and activity in Chicago, a jurisdiction that has already instituted diversity reforms.

We first show that minority officers receive vastly different patrol assignments. Without accounting for this disparity, there is no way to meaningfully characterize the differences in behavior across officer groups. In supplemental analyses (text S3.6 and fig. S13) we replicate our core analysis while iteratively imposing data restrictions common in previous analyses to show that common data constraints can lead to severely mistaken inferences, sometimes reversing substantive conclusions entirely. These disparities between analytic approaches suggest one explanation for the at-present mixed conclusions in studies on officer diversity: Data scarcity has imposed severe analytic constraints that can produce divergent, mistaken conclusions.

We account for these differences in working conditions by analyzing officers working in comparable places and times. We show that Black officers are less likely to stop, arrest, and use force against civilians, especially Black civilians, relative to white officers. These disparities are driven by reduced discretionary stops and arrests for petty crimes, including drug offenses, which have long been thought to fuel mass incarceration (*1*). By contrast, Black officers' violent crime enforcement is

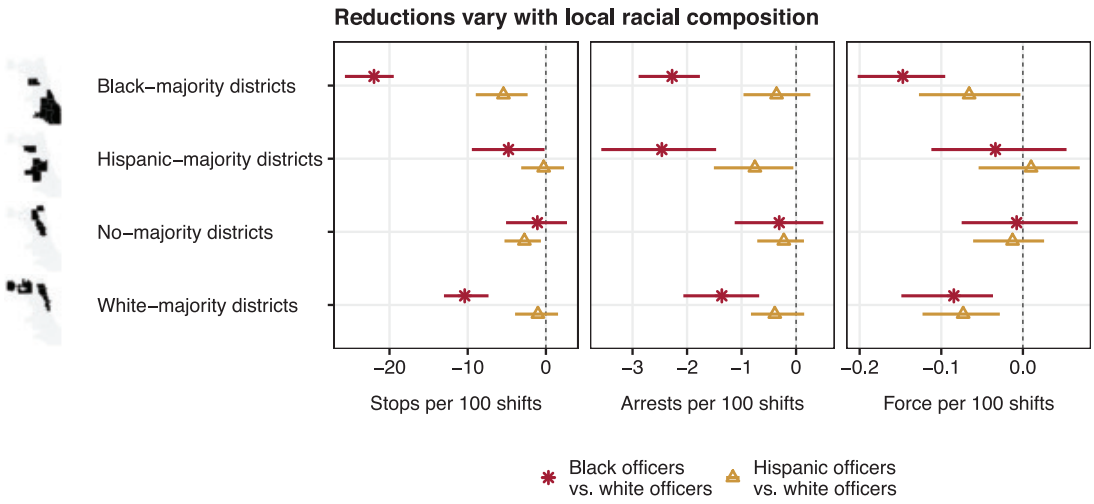


Fig. 5. Effects of deploying officers from marginalized groups, by local racial composition. Average within-MDSB differences in stops, arrests, and uses of force across officer groups, disaggregating CPD districts by majority-resident demographics.

only slightly lower than white officers'. Hispanic officers display lower levels of enforcement activity than whites overall, but their behavior toward Hispanic civilians is broadly comparable to that of white officers, a pattern that deserves further investigation with more fine-grained data on this ethnic group. We also find substantial differences in the behavior of female officers—both relative to male officers generally and within racial and ethnic groups—with the most substantial differences pertaining to use of force. The vast majority of gendered reductions stem from a reduced focus on arresting and using force against Black civilians.

Our results also reveal patterns requiring further study, especially with regard to causal mechanisms. One explanation for these disparities centers on racial bias, i.e., white officers are more likely than Black officers to harass Black civilians. Technically, it is also possible that Black officers respond more leniently when observing crimes in progress (*25*). Though we cannot fully disentangle these observationally equivalent explanations, our data show that these enforcement disparities are predominantly focused on relatively minor crimes, not violent offenses, suggesting little trade-off in terms of public safety. Arbitrating between these competing mechanisms will require objective information on civilian behavior.

Nevertheless, these results help evaluate the promise of proposed personnel reforms by showing what average behavior can be expected when deploying officers of a given demographic profile, relative to their counterparts, holding environmental factors fixed. If we were unable to discern disparities in behavior across these officer groups, diversity reforms would be unlikely to meaningfully alter the volume and character of policing. In fact, not only do we observe differences in enforcement patterns, we also find that these gaps remain nearly identical when adjusting for officer experience (see figs. S15 to S17), an important consideration when extrapolating from retrospective assessments to the future hiring of inexperienced officers. Despite these effects, and even in this highly diverse department, Black and Hispanic civilians in Chicago are engaged by police at rates disproportionate to their shares of the population (though such disparities can arise from a variety of sources, including deployment patterns, civilian behavior, or officer bias). Although our results show that diversity in law enforcement can narrow these gaps, it cannot, on its own, fully address the substantial racial disparities that characterize the American carceral system.

Our analysis uses data from a single city, allowing for an unusually detailed analysis at some cost in generalizability. At present, a patchwork of nonstandard record-keeping and disclosure practices across roughly 18,000 U.S. police agencies (*39*) has severely impeded broader policy evaluations. Our approach, patrol-assignment analyses, offers a useful and widely applicable template for other scholars to follow when testing whether our findings hold in other places and times. But these efforts will require collection of similar data elsewhere, likely necessitating open-records requests, litigation, or data-transparency reforms to compel the release of patrol records that have rarely been shared freely. Acquiring data in these ways can also help mitigate selection bias that can result from forming research partnerships with police agencies, an approach that may skew the literature by focusing on cooperative jurisdictions.

Taken together, these results strongly suggest that diversification can reshape police-civilian encounters. But extrapolation to future hiring hinges on whether recruits come from a comparable pool of potential employees and are deployed in comparable ways. Policing is evolving rapidly, and a complete understanding of the efficacy of reforms requires continued, in-depth research. As officers from marginalized communities increasingly join police forces, their presence will necessarily lead to shifts in deployment and department norms. In turn, shifting deployment patterns may reshuffle officers with particular dispositions to different locations. This could produce different results if, for example, the white officers who are most violent toward Black civilians are then removed from Black neighborhoods, which could shrink the gap in force rates relative to Black officers. If so, the cost-benefit calculus of diversification would be further complicated. The framework that we provide in this study provides a template for future scholars to reevaluate these effects as necessary.

The effects of diversification are likely neither simple nor monolithic. Officers are multidimensional, and crafting effective personnel reforms will likely require thinking beyond the coarse demographic categories typically used in diversity initiatives and consideration of how multiple attributes relate police to the civilians they serve.

REFERENCES AND NOTES

1. M. Alexander, *The New Jim Crow: Mass Incarceration in the Age of Colorblindness* (The New Press, 2010).
2. A. Lerman, V. Weaver, *Arresting Citizenship: The Democratic Consequences of American Crime Control* (Univ. of Chicago Press, 2014).
3. J. Forman Jr., *Locking Up Our Own: Crime and Punishment in Black America* (Farrar, Straus and Giroux, 2017).
4. U.S. Department of Justice, "Advancing Diversity in Law Enforcement" (2016).
5. A. Fantz, C. Tolan, "Want to reform the police? Hire more women," CNN, 2020.
6. Z. Qian, *Race Soc.* **5**, 33–47 (2002).
7. T. A. Laveist, A. Nuru-Jeter, *J. Health Soc. Behav.* **43**, 296–306 (2002).
8. E. M. Kearns, *Justice Q.* **34**, 1213–1245 (2017).
9. H. Tajfel, J. Turner, in *The Social Psychology of Intergroup Relations* W. G. Austin, S. Worchel, Eds. (Wadsworth, 1979), pp. 56–65.
10. C. Neil Macrae, G. V. Bodenhausen, *Annu. Rev. Psychol.* **51**, 93–120 (2000).
11. S. T. Fiske, *Curr. Directions Psychol. Sci.* **11**, 123–128 (2002).
12. J. D. Clinton, D. E. Lewis, *Polit. Anal.* **16**, 3–20 (2008).
13. D. E. Lewis, *Int. Public Manage. J.* **13**, 46–55 (2010).
14. E. Linos, N. Riesch, *Public Admin. Rev.* **80**, 92–103 (2020).

15. J. Q. Wilson, *Bureaucracy: What Government Agencies Do and Why They Do It*, J. Q. Wilson, Ed. (Basic Books 1989).
16. K. J. Meier, *Am. Polit. Sci. Rev.* **69**, 526–542 (1975).
17. V. Ray, *Am. Sociol. Rev.* **84**, 26–53 (2019).
18. S. H. Alim, G. Smitherman, *Articulate While Black: Barack Obama, Language, and Race in the U.S.* (Oxford Univ. Press, 2012).
19. J. Legewie, J. Fagan, "Group threat, police officer diversity and the deadly use of police force," *Columbia Public Law Research Paper* (2016).
20. J. J. Donohue III, S. D. Levitt, *J. Law Econ.* **44**, 367–394 (2001).
21. B. Smith, *Policy Stud.* **31**, 147–162 (2003).
22. A. Fagan, A. Braga, R. K. Brunson, A. Pattavina, *Fordham Urban Law J.* **43**, 539–614 (2016).
23. P. B. Hoffman, E. R. Hickey, *J. Crim. Justice* **33**, 145–151 (2005).
24. E. K. Weisburst, *AEA Papers and Proceedings* **109**, 152–156 (2019).
25. J. West, "Racial Bias in Police Investigations," Working Paper, 2018; https://people.ucsc.edu/~jwest1/articles/West_RacialBiasPolice.pdf.
26. J. McCrary, *Am. Econ. Rev.* **97**, 318–353 (2007).
27. A. R. Miller, C. Segal, *Rev. Econ. Stud.* **86**, 2220–2247 (2019).
28. A. Harvey, T. Mattia, "Reducing Racial Disparities in Crime Victimization," Working Paper, 2019; https://wp.nyu.edu/annaharvey/wp-content/uploads/sites/6417/2020/07/Victimization_Revised.pdf.
29. M. Garner, A. Harvey, H. Johnson, *Int. Rev. Law Econ.* **62**, 105881 (2020).
30. D. A. Sklansky, *J. Crim. Law Criminol.* **96**, 1225 (2005).
31. ACLU, "Stop and Frisk in Chicago," American Civil Liberties Union of Illinois, 2015.
32. J. Byrne, R. Pearson, J. Gorner, "Mayor Rahm Emanuel rejects President Donald Trump's call for Chicago police to use stop-and-frisk tactics," *Chicago Tribune*, 9 October 2018.
33. Police Accountability Task Force, "Recommendations for Reform: Restoring Trust Between the Chicago Police and the Communities They Serve," 2016.
34. E. McClelland, *Politico*, 29 September 2018.
35. WTTW, "Power, Politics, & Pride: Afro-American Patrolmen's League," https://interactive.wttw.com/dusable-to-obama/afro-american-patrolmens-league [accessed 28 January 2021].
36. M. M. Adams, "The Patrolmen's Revolt: Chicago Police and the Labor and Urban Crises of the Late Twentieth Century," thesis, University of California, Berkeley (2012).
37. B. A. Ba, "Going the Extra Mile: The Cost of Complaint Filing, Accountability, and Law Enforcement Outcomes in Chicago," Working paper, 2019.
38. N. Rim, R. G. Rivera, B. A. Ba, In-group Bias and the Police: Evidence from Award Nominations (2019). Faculty Scholarship at Penn Law. 2132. https://scholarship.law.upenn.edu/faculty_scholarship/2132.
39. P. Atiba Goff, K. Barsamian Kahn, *Soc. Issues Policy Rev.* **6**, 177–210 (2012).
40. B. Ba, D. Knox, J. Mummolo, R. Rivera, The Role of Officer Race and Gender in Police-Civilian Interactions in Chicago [Source Code] (2021). https://doi.org/10.24433/CO.4519202.v1.

ACKNOWLEDGMENTS

Data for this project were gathered in collaboration with the Invisible Institute, with the help of S. Stecklow and E. Herman. We thank R. Ryley and the Lucy Parsons Lab for generously sharing data. We thank R. Austin, P. Bayer, S. Black, T. Burch, B. Canes-Wrone, T. Clark, J. Grimmer, A. Harris, D. Kronick, A. Harvey, H. Jefferson, G. Lenz, B. Magaloni, T. Mendelberg, V. Weaver, A. White, and attendees of the 2019 Princeton Policing Conference for helpful feedback. This project was approved by institutional review boards at Columbia University; Princeton University; University of California, Irvine; and University of Pennsylvania. **Funding: Author contributions:** All authors contributed equally to this article. **Competing interests:** D.K. and J.M. are paid consultants for the U.S. Department of Justice Civil Rights Division and the American Civil Liberties Union. R.R. worked for the Invisible Institute and the plaintiffs in *Campbell v. City of Chicago* between 2017 and 2018. **Data and materials availability:** All data, interactive replication materials, and guaranteed reproducible results are publicly available as a Code Ocean capsule (*40*). We encourage readers to probe and extend our analyses.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/371/6530/696/suppl/DC1
Supplementary Text
Figs. S1 to S20
Tables S1 to S8
References (*41*–*46*)

17 July 2020; accepted 7 December 2020
10.1126/science.abd8694

NUCLEOTIDE SYNTHESIS

Catalytic asymmetric and stereodivergent oligonucleotide synthesis

Aaron L. Featherston[1]*†, Yongseok Kwon[1]*‡, Matthew M. Pompeo[1]*, Oliver D. Engl[2], David K. Leahy[2]§¶, Scott J. Miller[1]¶

We report the catalytic stereocontrolled synthesis of dinucleotides. We have demonstrated, for the first time to our knowledge, that chiral phosphoric acid (CPA) catalysts control the formation of stereogenic phosphorous centers during phosphoramidite transfer. Unprecedented levels of diastereodivergence have also been demonstrated, enabling access to either phosphite diastereomer. Two different CPA scaffolds have proven to be essential for achieving stereodivergence: peptide-embedded phosphothreonine-derived CPAs, which reinforce and amplify the inherent substrate preference, and C2-symmetric BINOL-derived CPAs, which completely overturn this stereochemical preference. The presently reported catalytic method does not require stoichiometric activators or chiral auxiliaries and enables asymmetric catalysis with readily available phosphoramidites. The method was applied to the stereocontrolled synthesis of diastereomeric dinucleotides as well as cyclic dinucleotides, which are of broad interest in immuno-oncology as agonists of the stimulator of interferon genes (STING) pathway.

Synthetically modified nucleotides that are designed to mimic and interact with their physiological counterparts continue to grow in importance (*1*). Examples include cyclic dinucleotides (CDNs), a class of macrocyclic natural products with profound biological significance. The heterodimer 2′,3′-cyclic guanosine monophosphate–adenosine monophosphate (cGAMP) (Fig. 1A) elicits an immune response through activation of the stimulator of interferon genes (STING) protein (*2*–*5*). This activation pathway is currently being evaluated in multiple clinical studies for oncology therapeutics (*6*). Chemical modification of the native phosphodiester linkages is a powerful strategy to prepare small-molecule therapeutics with increased metabolic stability and therapeutic potential (*7*). One effective modification is the permutation of the phosphodiester linkage to a phosphorothioate (PS). However, this alteration adds chemical complexity because it renders the central phosphorus atoms stereogenic. Thus, four possible diastereomers of a PS-derived CDN exist, each with markedly different biological activities and physical properties (*6*, *8*–*10*). Most synthetic methods for the preparation of these CDN derivatives rely on nonselective syntheses requiring tedious reversed-phase high-performance liquid chromatography (HPLC) purifications to isolate diastereomerically pure analogs in small quantities (*11*). To date, the only stereoselective synthesis of a modified 3′,3′-CDN uses chiral auxiliaries, which despite impressive levels of stereocontrol, are deployed as components of stoichiometric reagents (*12*). Given the critical importance of these new stereogenic-at-phosphorus compounds, efforts to develop more robust and selective processes to access stereopure cyclic nucleotides and antisense oligonucleotides are of broad general interest.

The most general method for the synthesis of PS-modified oligonucleotides uses commercially available P(III)-phosphoramidites and a stoichiometric activator to afford a stochastic mixture of 2^n diastereomers (*13*). The use of chiral auxiliaries thus remains the only viable option to establish reliable stereocontrol (Fig. 1B). In a series of pioneering studies, Wada *et al.* demonstrated a P(III)-stereogenic chiral auxiliary that enabled access to stereochemically pure phosphodiester analogs (*14*, *15*). More recently, Baran *et al.* unveiled a powerful P(V)-based approach used for the stereocontrolled synthesis of PS compounds (*12*, *16*). However, by design, these strategies rely on multistep syntheses to prepare and install the auxiliaries; a stoichiometric quantity of by-product is also intrinsic to the approach, because the chiral auxiliary is eventually cleaved from the target compound after the stereospecific installation of the P-stereogenic functionality.

Asymmetric catalysis can potentially address the shortcomings of chiral auxiliary-based approaches. DiRocco *et al.* disclosed a multifunctional catalyst to assemble prodrugs at the P(V) oxidation state through a dynamic kinetic asymmetric transformation (DYKAT) (*17*), expanding greatly on the general principle of catalytic enantioselective phosphate transfer (*18*, *19*). However, application of this

[1]Department of Chemistry, Yale University, New Haven, CT 06520, USA. [2]Process Chemistry Development, Takeda Pharmaceuticals International Co., Cambridge, MA 02139, USA. *These authors contributed equally to this work. †Present address: Process Research & Development, AbbVie Inc., Chicago, IL 60064, USA. ‡Present address: Department of Chemistry, Sogang University, Seoul 121-742, Korea. §Present address: Biohaven Pharmaceuticals, New Haven, CT, 06510, USA.
¶Corresponding author. Email: scott.miller@yale.edu (S.J.M.); david.leahy@biohavenpharma.com (D.K.L.)

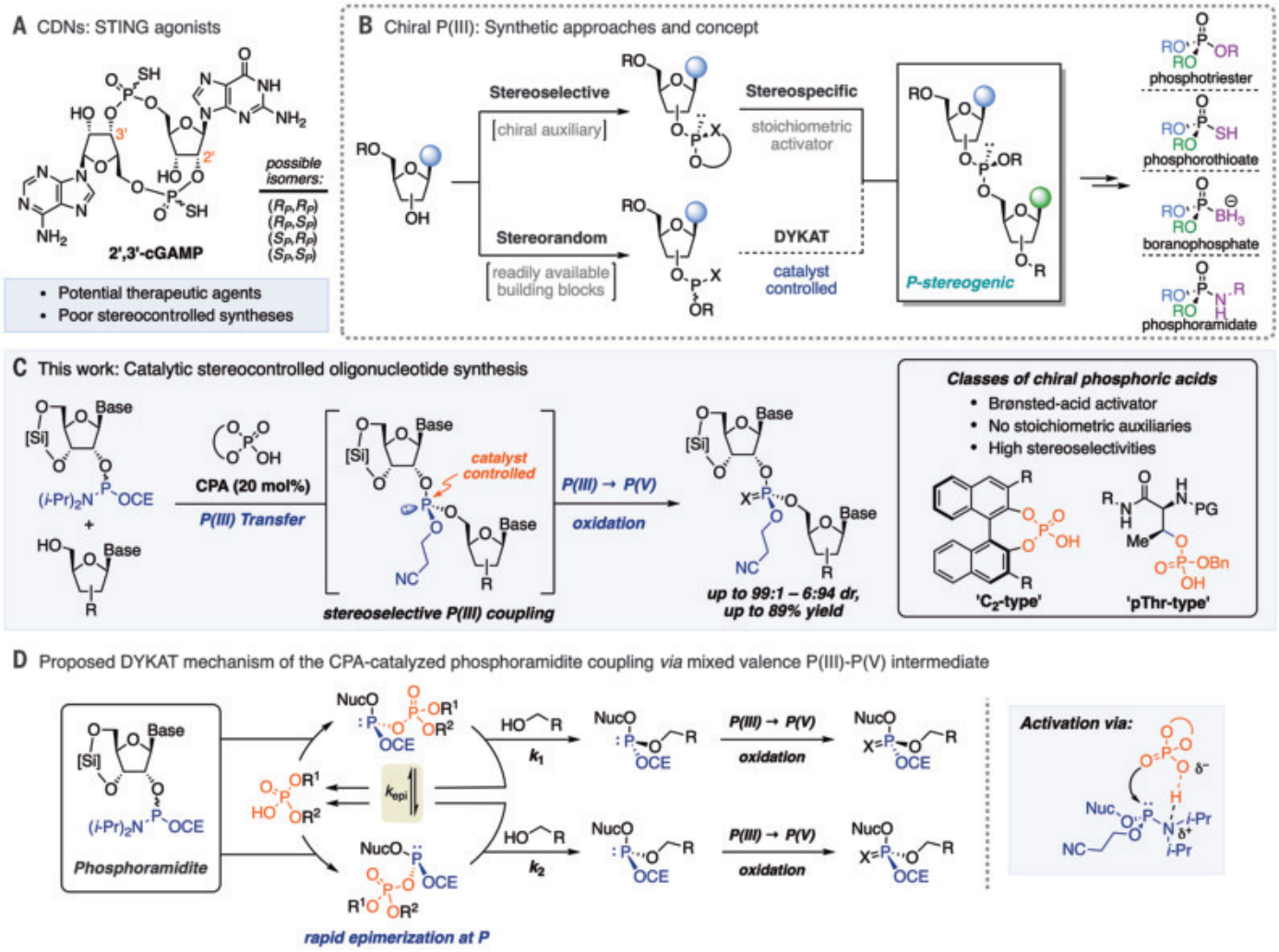


Fig. 1. Background and reaction development. (**A**) Structure of PS CDN. (**B**) Strategies toward the stereoselective synthesis of P(III) compounds used in oligonucleotide synthesis. (**C**) Brønsted acid–catalyzed dynamic asymmetric transformation of phosphoramidites for the synthesis of stereogenic P(III) compounds. (**D**) Proposed mechanism for CPA-catalyzed activation of phosphoramidite and epimerization of P-stereogenic center. Nuc, nucleoside; [Si], 1,1,3,3-tetraisopropyldisiloxane; OCE, 2-cyanoethoxy.

catalytic modality to the formation of stereogenic PS compounds has not yet been disclosed, nor was catalyst-controlled diastereodivergence demonstrated, which is critical in enabling a fully stereocontrolled synthesis of all possible oligonucleotide stereoisomers. Moreover, no catalytic asymmetric method has yet been reported for selective P(III)-based transfers within the context of the widely used phosphoramidite method for nucleotide synthesis.

Classically, for the synthesis of PS-modified CDNs, the first thiophosphate linkage is often formed from a phosphoramidite coupling from readily available nucleoside fragments. Brønsted acid activators of the P(III) phosphoramidite, which are often used in stoichiometric or superstoichiometric quantities, produce an interconverting mixture of chiral P(III) intermediates, followed by stereorandom nucleophilic displacement to afford a chiral P(III) phosphite intermediate (*20*, *21*). Subsequent oxidative sulfurization affords the desired stereogenic P(V) species as a mixture of diastereomers. Catalytic stereoselective phosphoramidite transfer would represent an important step forward if it were able to harness readily accessible phosphoramidites that are staples of the global oligonucleotide synthesis enterprise. However, catalytic concepts for asymmetric transfer of P(III) phosphoramidites are in their infancy, having been applied only to a small set of enantioselective and site-selective reactions (*22*, *23*). Analogous diastereoselective methods to produce P-stereogenic compounds at the P(III) oxidation state are also not yet available. Accordingly, we present herein a catalytic synthesis of P(III)-stereogenic compounds based on a previously unknown catalytic approach, the use of chiral phosphoric acid (CPA)–mediated couplings to access stereodefined dinucleotides in a stereocontrolled manner (Fig. 1C). Our mechanistic concept is predicated on a hypothesized mixed-valence P(III)-P(V) substrate-catalyst complex (Fig. 1D) that undergoes rapid, pre-stereochemistry–determining epimerization at P(III) and is exemplified with the synthetic and mechanistic studies described below.

Results and discussion

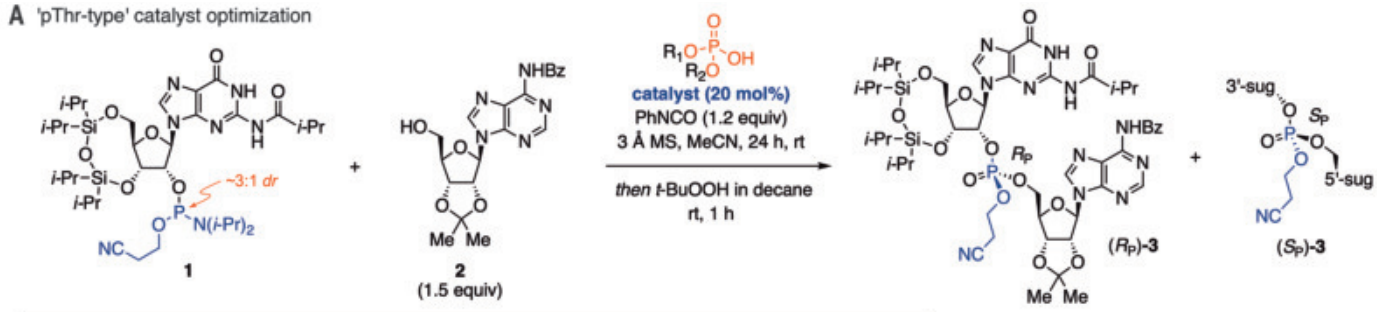
The efficient coupling of nucleoside P(III) phosphoramidites generally requires Brønsted acid activators such as 1*H*-tetrazole or 4,5-dicyanoimidazole (*13*, *24*, *25*). However, an excess of the acid is required due to catalyst sequestration as the dialkylammonium salt resulting from the formation of stoichiometric dialkylamine. Recent advances enable the use of these activators in substoichiometric quantities (*13*), yet none has addressed the challenge of stereocontrol. Perhaps the sole example of catalytic asymmetric phosphitylation used tetrazoylalanine-based catalysts in

the desymmetrization of D-*myo*-inositol through enantioselective phosphoramidite transfer (*22*). In these reactions, the chiral tetrazole serves as a Brønsted acid in conjunction with 10-Å molecular sieves to scavenge the liberated diethylamine, thereby facilitating catalyst turnover. This result supported the notion that chiral Brønsted acids could serve as effective catalysts for stereocontrolled oligonucleotide synthesis.

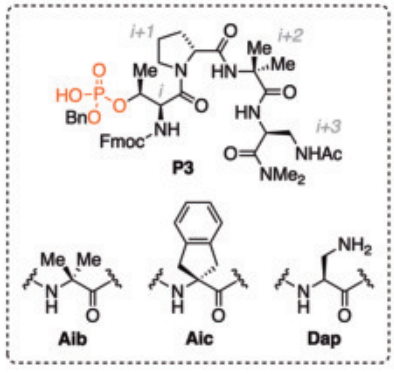
Over the past decade, C2-symmetric CPAs have emerged as powerful Brønsted acid catalysts for myriad asymmetric transformations (*26*). Recent efforts by our group have focused on the development of peptide-embedded phosphothreonine (pThr) as a new class of CPA catalyst (*27*–*29*). To assess the ability of these CPAs to affect an asymmetric phosphitylation, we began with the 2′ to 5′ coupling of guanosine phosphoramidite **1** [~3:1 diastereomeric ratio (dr)] and adenosine derivative **2** (Fig. 2A). Indeed, simple phosphoric acids catalyze the coupling of **1** and **2** to provide a P(III) dinucleotide intermediate, which then undergoes a stereoretentive oxidation with *t*-BuOOH to afford the desired stereogenic phosphate ester **3**. Phenyl isocyanate was found to be an effective scavenger of the liberated diisopropylamine by irreversible formation of a phenyl urea adduct, thereby enabling catalyst turnover (*30*), whereas the use of 3-Å molecular sieves prevented any acid-catalyzed hydrolysis of phosphoramidite **1** (*22*).

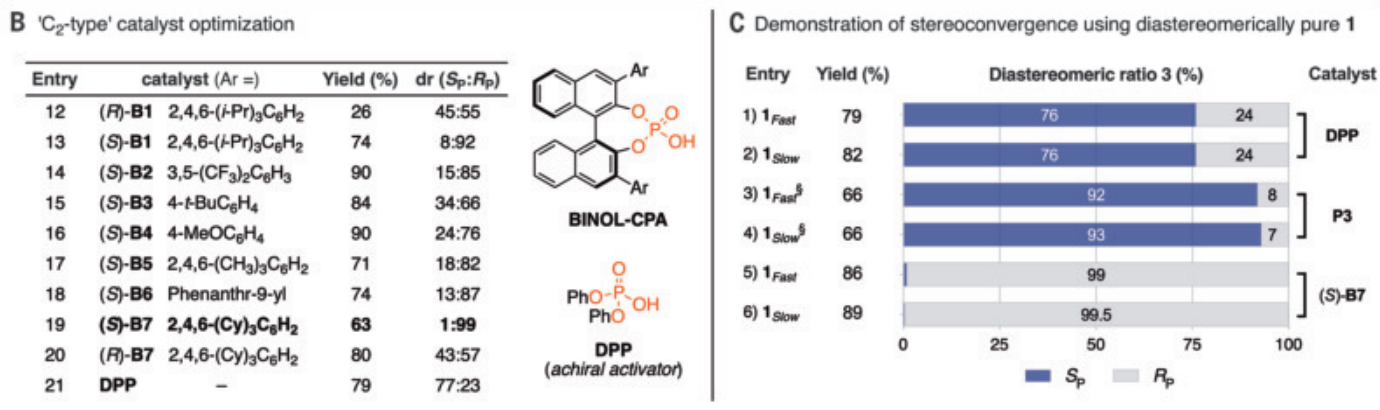
Catalyst optimization

To assess the feasibility of stereocontrol, we evaluated the two families of CPAs and compared the results with those obtained with diphenylphosphoric acid (**DPP**), an achiral catalyst used to benchmark the intrinsic selectivity innate to the chiral substrates. Evaluation of pThr-derived catalysts quickly revealed that catalysts biased toward the β-turn structure with a DPro and Aib motif in the *i*+1 and *i*+2 positions, respectively, emerged



Entry	Catalyst					Temp. (°C)	Conc. [M]	Yield (%)	dr (S_P:R_P)
		i	*i*+1	*i*+2	*i*+3				
1	**P1**	Fmoc-pThr(Bn)	DPro	Aib	Dap(Ac)-NMe_2	rt	0.10	78	70:30
2	**P2**	Fmoc-pThr(Bn)	DPro	Aic	Dap(Ac)-NMe_2	rt	0.10	92	68:32
3	**P3**	Fmoc-*allo*-pThr(Bn)	DPro	Aib	Dap(Ac)-NMe_2	rt	0.10	77	85:15
4	**P3**	Fmoc-*allo*-pThr(Bn)	DPro	Aib	Dap(Ac)-NMe_2	rt	0.04	81	91:9
5	**P4**	Fmoc-*allo*-pThr(Bn)	DPro	Aib	Dap(Bz)-NMe_2	rt	0.04	65	89:11
6	**P5**	Fmoc-*allo*-pThr(Bn)	DPip	Aib	Dap(Ac)-NMe_2	rt	0.04	89	87:13
7	**P6**	Fmoc-*allo*-pThr(Bn)	DHyp(*t*-Bu)	Aib	Dap(Ac)-NMe_2	rt	0.04	82	87:13
8	**P7**	Fmoc-*allo*-pThr(Bn)	DPro	Aib	Dap(Ac)-NEt_2	rt	0.04	84	88:12
9	**P8**	Ac-Fmoc-*allo*-pThr(Bn)	DPro	Aib	Dap(Ac)-NMe_2	rt	0.04	66	84:16
10	**P3**	**Fmoc-*allo*-pThr(Bn)**	**DPro**	**Aib**	**Dap(Ac)-NMe_2**	**4**	**0.04**	**79**	**94:6**
11†	**P3**	Fmoc-*allo*-pThr(Bn)	DPro	Aib	Dap(Ac)-NMe_2	4	0.04	67	88:12





Entry	catalyst (Ar =)		Yield (%)	dr (S_P:R_P)
12	(*R*)-**B1**	2,4,6-(*i*-Pr)$_3$$C_6H_2$	26	45:55
13	(*S*)-**B1**	2,4,6-(*i*-Pr)$_3$$C_6H_2$	74	8:92
14	(*S*)-**B2**	3,5-$(CF_3)_2C_6H_3$	90	15:85
15	(*S*)-**B3**	4-*t*-BuC_6H_4	84	34:66
16	(*S*)-**B4**	4-MeOC_6H_4	90	24:76
17	(*S*)-**B5**	2,4,6-$(CH_3)_3C_6H_2$	71	18:82
18	(*S*)-**B6**	Phenanthr-9-yl	74	13:87
19	**(*S*)-B7**	**2,4,6-$(Cy)_3C_6H_2$**	**63**	**1:99**
20	(*R*)-**B7**	2,4,6-$(Cy)_3C_6H_2$	80	43:57
21	**DPP**	–	79	77:23

Fig. 2. Catalyst and reaction optimization studies. (**A**) Optimization of pThr-catalyst favoring (S_P)-**3**. (**B**) Optimization of BINOL CPA favoring (R_P)-**3** diastereomer. (**C**) Diastereomeric ratios of **3** starting from stereopure phosphoramidite **1**. "Fast" and "slow" refer to the elution rates of phosphoramidites during purification using reversed-phase chromatography. Reactions were performed on 20 μmol scale at room temperature unless otherwise noted. Yields and diastereomeric ratios are of isolated material after purification. Yield refers to combined yield of both diastereomers; dr determined by ^{1}H-NMR or HPLC analysis. †Yield and dr were determined by ^{31}P analysis of the crude reaction mixture on 50 μmol scale. §Reactions performed at 4°C. MS, molecular sieves; Fmoc, fluorenylmethoxycarbonyl; Aib, α-aminoisobutyric acid; Dap, 2,3-diaminopropionic acid; Ac, acetyl.

as selective catalysts (Fig. 2A, entries 1 to 11) (*31*, *32*). In particular, catalyst **P3** (entry 3; 20 mol %) afforded the product in 77% yield and 85:15 dr, amplifying the S_P diastereomer (relative to the 77:23 ratio obtained with **DPP**; entry 21). Two notable features of this catalyst are (i) the diaminopropionic acid (Dap) *i*+3 residue, which often captures additional noncovalent interactions between catalysts and substrates (*28*), and (ii) the *allo* stereochemistry of the threonine residue (*29*). Further modifications to the catalyst architecture did not lead to a profound increase in selectivity (**P4** to **P8**). However, several other variables influence the stereochemical outcome. By reducing the reaction concentration to 0.04 M in acetonitrile, the selectivity was improved to 91:9 dr (entry 4); lowering the reaction temperature to 4°C afforded the product in 79% isolated yield and 94:6 dr (entry 10). Scaling up to 0.05 mmol scale under a normalized set of reaction conditions (entry 11), the product was obtained in 67% yield and 88:12 dr.

Given the substantial impact of P-stereogenicity on the efficacy of oligonucleotide-based therapeutics (*6*, *8*–*10*), we sought a complementary, stereodivergent phosphitylation catalyst (*33*). The CPA chemotype based on the C2-symmetric BINOL scaffold provided this outcome (*34*, *35*). Initially, (*R*)-TRIP (**B1**) was evaluated under the standard reaction conditions, and dinucleotide **3** was obtained in 26% yield and with a modest but diastereomerically reversed 45:55 dr (Fig. 2B, entry 12), favoring slightly the alternative R_P diastereomer. When using (*S*)-**B1** as the catalyst, a marked improvement in yield and selectivity was observed (74% yield and 8:92 dr; entry 13), substantially enhancing the selectivity for the opposite diastereomer with respect to that observed with the pThr-based catalysts. A survey of BINOL-CPAs revealed the tricyclohexyl catalyst (*S*)-**B7** to be superior (entries 14 to 19), affording the (R_P)-configured product in 63% yield and 1:99 dr. The enantiomeric catalyst (*R*)-**B7** was much less selective (43:57 dr) but still favored the same R_P diastereomer. It is thus demonstrated that the two distinct catalyst scaffolds, the pThr-based CPA framework and the BINOL-derived CPA architecture, enable catalytic asymmetric phosphoramidite couplings that exhibit notable degrees of stereodivergence.

Mechanistic hypothesis

One hallmark of these reactions is the stereoconvergent transformation of a diastereomeric mixture of phosphoramidite stereoisomers (**1**) to a highly diastereoenriched phosphite product, which is then oxidized stereospecifically to the corresponding stereogenic phosphate [(R_P)-**3** or (S_P)-**3**)]. In analogy to the widely accepted mechanism for 1*H*-tetrazole (*13*, *24*), we envisioned that the phosphoric acid catalyst acts as both an acid and nucleophilic catalyst after proton transfer to displace diisopropylamine (Fig. 1D). Critically, the resulting mixed-valence P(III)–P(V) intermediate could undergo rapid epimerization through displacement with an additional molecule of catalyst or, alternatively, through intramolecular inversion with the adjacent P=O. The stereochemical outcome of the coupling reactions thus tracks with the relative magnitude of the displacement rate constants k_1 and k_2 in a typical Curtin-Hammett type regime characteristic of a DYKAT mechanism (*36*–*38*).

To validate our assumptions and to probe the mechanistic hypothesis, we separated the phosphoramidite diastereomers (**1**) by reversed-phase chromatography (Fig. 2C). The resulting diastereomerically pure phosphoramidites ($\mathbf{1}_{Fast}$ and $\mathbf{1}_{Slow}$) were then subjected to the

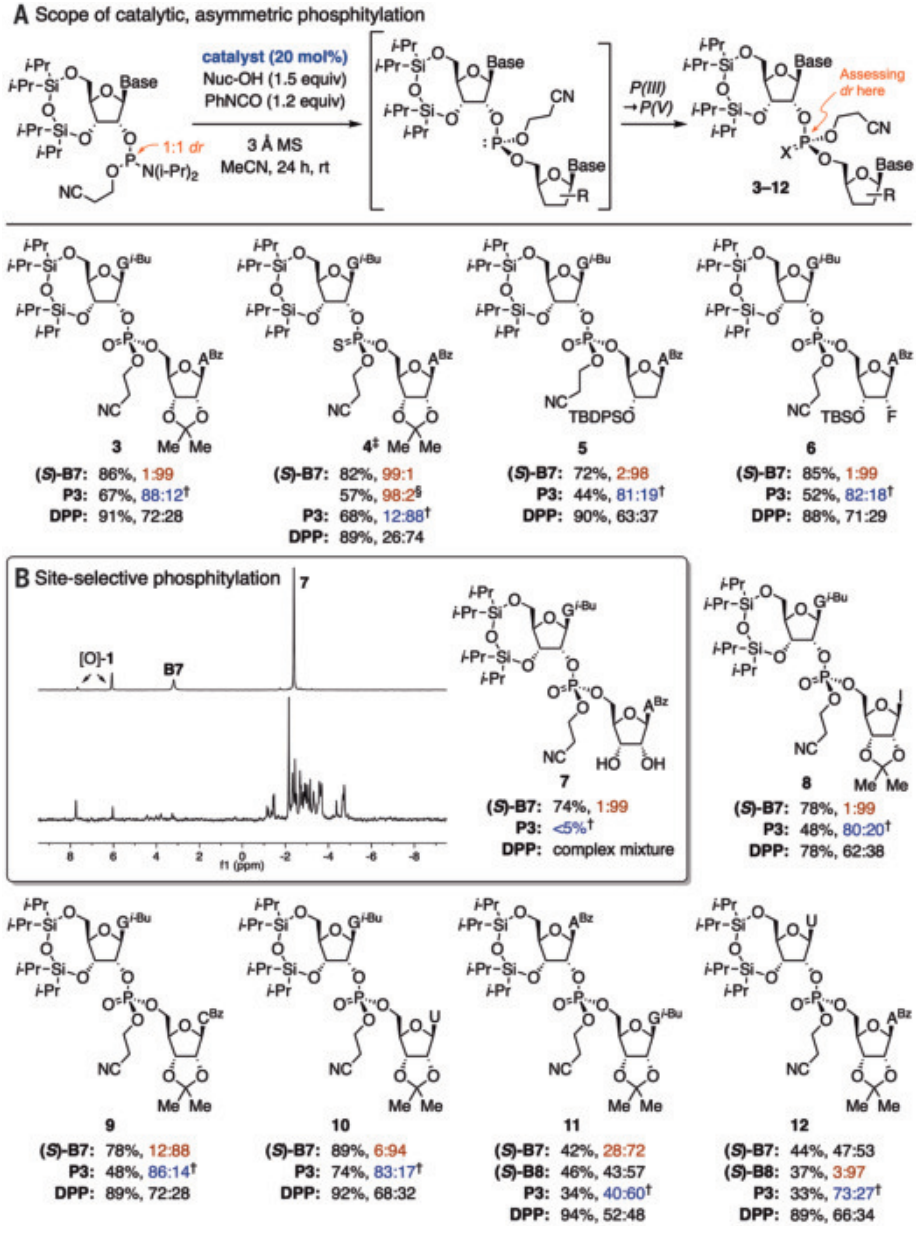
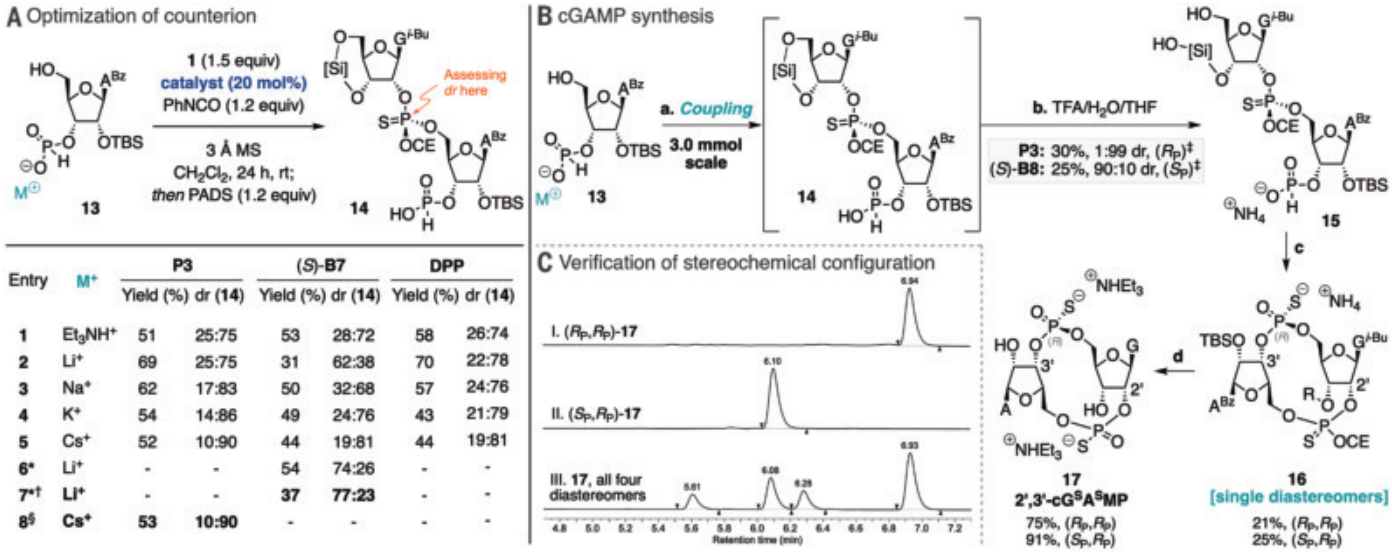


Fig. 3. Substrate scope. (**A**) Standard conditions: phosphoramidite (0.05 mmol), 20 mol % catalyst, nucleophile [1.5 equivalents (equiv)], PhNCO (1.2 equiv), 3-Å MS, MeCN, room temperature (RT), 24 hours then *t*-BuOOH, 1 hour. Yield and diastereoselectivity were determined by quantitative ^{31}P-NMR analysis of the crude reaction mixture with internal standard (triphenyl phosphate) and are the average of two trials; dr refers to the S_P:R_P ratio. †Reactions were performed at 4°C. ‡PADS (1.2 equiv) was used instead of *t*-BuOOH. §Reaction was performed on 0.20 mmol scale with 5 mol % (*S*)-**B7** at 0.40 M in MeCN. (**B**) ^{31}P-NMR analysis of crude dinucleotide **7** using (*S*)-**B7** (top) and **DPP** (bottom). $G^{i\text{-}Bu}$, N2-*iso*butyrylguanine; A^{Bz}, N6-benzoyladenine; I, inosine; C^{Bz}, N4-benzoylcytosine; U, uracil.



Entry	M+	P3 Yield (%)	P3 dr (14)	(S)-B7 Yield (%)	(S)-B7 dr (14)	DPP Yield (%)	DPP dr (14)
1	Et_3NH^+	51	25:75	53	28:72	58	26:74
2	Li^+	69	25:75	31	62:38	70	22:78
3	Na^+	62	17:83	50	32:68	57	24:76
4	K^+	54	14:86	49	24:76	43	21:79
5	Cs^+	52	10:90	44	19:81	44	19:81
6*	Li^+	-	-	54	74:26	-	-
7*†	Li^+	-	-	37	77:23	-	-
8§	Cs^+	53	10:90	-	-	-	-

Fig. 4. Gram-scale CDN synthesis. (**A**) Optimization of H-phosphonate counterion. Standard conditions: **1** (0.03 mmol), 20 mol % catalyst, nucleophile (1.0 equiv), PhNCO (1.2 equiv), 3-Å MS, CH_2Cl_2, RT, 24 hours then PADS (1.2 equiv), 16 hours. Yield and diastereoselectivity were determined by quantitative ^{31}P-NMR analysis of the crude reaction mixture with triphenyl phosphate as internal standard; dr refers to the S_P:R_P ratio. *Reaction performed with 1.0 equiv **1** in MeCN/DMF (9:1 v/v) using (*S*)-**B8**. †Reaction performed at 4°C. §Reaction performed with 1.0 equiv **1**. (**B**) Reaction conditions: (a) **1** (1.0 equiv), **13** (1.0 equiv), catalyst (20 mol %), PhNCO (1.2 equiv), 3-Å MS for 7 to 48 hours and then PADS (1.2 equiv) for 16 hours; (b) TFA, H_2O, THF, 4°C, 5 hours; (c) NMI (20 mol %), 2,6-lutidine (3.6 equiv), diphenyl phosphoryl chloride (2.0 equiv), 0°C, 1 hour then Beaucage reagent (3.0 equiv), RT, 30 min; (d) (*i*) 33% $MeNH_2$ in EtOH, 5 hours; (*ii*) Et_3N•HF (5.00 equiv), pyridine, Et_3N, 50°C, 18 hours. ‡**13**•Cs was used with **P3** and **13**•Li was used with (*S*)-**B8**. (**C**) HPLC analysis (λ = 254 nm) of **17** as a stochastic mixture of diastereomers (bottom trace), (S_P,R_P)-**17** (middle trace), and (R_P,R_P)-**17** (top trace). A, adenine; G, guanine; A^{Bz}, N6-benzoyladenine; $G^{i\text{-}Bu}$, N2-*iso*butyrylguanine.

optimized reaction conditions for each catalyst: **P3**, (*S*)-**B7**, and **DPP**. The stereochemical outcome of dinucleotide ***3*** was found to be independent of the starting dr of **1**, because the catalysts converted either diastereomer of **1** to either (R_P)-***3*** or (S_P)-***3*** with the same diastereoselectivity.

Substrate scope

With a pair of stereodivergent catalysts in hand, we explored the scope of the coupling with a selection of representative substrates (Fig. 3). By controlling the stereochemistry at P(III), we can access a variety of P(V) products. For example, treatment of the intermediate phosphite with *t*-BuOOH affords the phosphate ***3***, whereas sulfurization using phenylacetyl disulfide (PADS) affords the PS ***4*** (*39*). The loading of (*S*)-**B7** could be reduced to 5 mol % to produce (S_P)-***4*** in 57% yield at 98:2 dr (*40*) by increasing the reaction concentration to 0.40 M. Because methods for the synthesis of selenophosphate (*41*), boranophosphate (*42*), phosphoramidate (*43*), and methyl phosphonate (*44*) from the P(III) oxidation state have all been reported, this new catalytic protocol enables access to a wide range of functionalized P(V) products (Fig. 1B).

In addition to isopropylidene adenosine (**3**), deoxyadenosine (**5**) and C2′-fluoro-deoxyadenosine (**6**) were also well tolerated. In each case, analogous diastereodivergence was observed with enhancement of intrinsic selectivity with catalyst **P3** and near-total reversal of selectivity with catalyst (*S*)-**B7**. A particularly notable result was obtained when 2′,3′-unprotected adenosine was used as the nucleophilic coupling partner. In this case, a highly site-selective and diastereoselective phosphorylation was achieved with (*S*)-**B7**, selecting for a single product (**7**) over five possible alternative stereoisomers or regioisomers of monocoupled product. By contrast, a complex mixture of products was obtained when **DPP** was used as the catalyst, as shown in the crude ^{31}P–nuclear magnetic resonance (NMR) spectra (Fig. 3B). Other ribonucleosides could also be coupled to guanosine **1** in good yields and selectivity (**8** to **10**). Selectivities were lower when the roles of the coupling partners were reversed; when adenosine-derived phosphoramidite was coupled to the 5′-OH of guanosine acetonide, both catalysts favored slightly the same (unassigned) diastereomer of dinucleotide **11**. Furthermore, the case of uridine illustrates the potential for catalyst refinement in situations in which the selectivity for new substrates is modest. For example, in the case of substrate **12**, BINOL-derived catalyst (*S*)-**B7** delivered the product with an unimpressive dr of 47:53, yet an alternative CPA (*S*)-**B8** (Ar = 9-anthryl; see the supplementary materials) delivered the product with an excellent 3:97 dr. Peptide-derived catalyst **P3** did slightly enhance the selectivity for the antipodal diastereomer in this case but with lower selectivity than observed with the model substrate (73:27 dr).

Taken together, our scope studies reveal that the diastereodivergent catalysis that we observed can be applied to other nucleobase pairs. The efficacy of the catalysts optimized for the coupling of guanosine phosphoramidite **1** and adenosine **2** appears to be more strongly dependent on the structure of the electrophile than the nucleophile. Selectivities remain high when coupling **1** to an array of different nucleosides (**3** to **10**) yet decrease when coupling other phosphoramidites (**11** and **12**). Nonetheless, whereas neither universal catalysts nor a complete set of catalysts for all possible couplings is yet in hand, the concepts we have demonstrated seem generalizable for this very ambitious longer-term goal.

CDN synthesis

To address the reported synthetic inefficiencies associated with both nonselective nucleoside couplings and (super)stoichiometric activators and auxiliaries, we then prepared stereodefined CDNs in a catalyst-controlled manner (*2*–*5*, *11*). We envisioned a concise asymmetric approach using this newly developed stereoselective phosphoramidite coupling between **1** and **13**, followed by an H-phosphonate macrocyclization to access PS **16** with the desired phosphorus stereochemistry (Fig. 4) (*45*). Initially, coupling of **1** and **13**•Et_3N under the optimized catalytic conditions

revealed a disappointing level of selectivity (Fig. 4A, entry 1). Control experiments in the absence of chiral catalyst revealed a fast, nonselective background reaction, likely a result of autocatalysis promoted by the triethylammonium cation ($pKa_{(DMSO)}$ = 9.0), which outcompetes the catalyzed process (see the supplementary materials) (*46*).

To circumvent this autocatalysis, we prepared a selection of alkali-metal phosphonate salts through ion exchange and evaluated their efficacy. When the reaction was performed with **13**•Na in the absence of catalyst, the product was obtained in 23% yield after 24 hours, compared with 46% with **13**•Et_3N, suggesting a suppressed background reaction (see the supplementary materials). With our previously optimized catalysts, we found that the selectivity was dependent on both the counterion and the catalyst structure. Using pThr catalyst **P3**, we found that the Cs^+ phosphonate salt afforded the highest selectivity for the R_P diastereomer of **14** (10:90 dr) (*40*). Furthermore, catalyst (*S*)-**B8** emerged from a catalyst screen as the most selective for the S_P diastereomer using the Li^+ salt of **13** (74:26 dr). With further refinement, the use of MeCN/DMF at 4°C led to increased selectivity (77:23 dr) because of the increased solubility of the Li^+ salt under the reaction conditions. It is noteworthy that with these highly functionalized zwitterionic substrates, the peptidic catalyst **P3** actually outperformed the BINOL-derived catalyst (*S*)-**B8**, in contrast to the less functionalized substrates presented in Fig. 3. Thus, catalyst tuning and counterion optimization enabled the preparation of each diastereomer of dinucleotide **14** with high diastereoselectivity.

To complete the syntheses of stereopure CDNs (*47*), we performed the couplings of **1** and **13**•Cs with **P3** on a 3.0 mmol scale to isolate >2 g of crude (R_P)-**14** after a simple crystallization. Site-selective hydrolysis of the disilane at C5′ with aqueous trifluoroacetic acid (TFA) then afforded the macrocyclization precursor (R_P)-**15** in 30% yield and 1:99 dr over two steps after flash chromatography. When subjected to diphenyl phosphoryl chloride with 2,6-lutidine as the base and 20 mol % *N*-methylimidazole (NMI) as a nucleophilic catalyst in dichloromethane, (R_P,R_P)-**16** could be obtained with high diastereoselectivity (>90:10 dr) and isolated in pure form in 21% yield (*48*, *49*). After global deprotection, the final CDN 2′,3′-cG^{S}A^{S}MP [(R_P,R_P)-**17**] was isolated as a single diastereomer through crystallization from acetone in 75% yield and high purity over two steps. The (2′-5′)S_P,(3′-5′)R_P diastereomer could also be obtained in high selectivity through a similar sequence by using (*S*)-**B8** as a catalyst with **13**•Li (Fig. 4C). The relative stereochemical configuration at phosphorus of (R_P,R_P)-**17** and (S_P,R_P)-**17** was determined by measuring the HPLC elution order and comparison with the work of Lim and Kim (*50*). Thus, peptide-derived catalyst **P3** selects for the (R_P)-configured PS, whereas BINOL-derived catalysts (*S*)-**B7** and (*S*)-**B8** select for the (S_P)-configured PS. The relative configurations of the 2′,5′-linked dinucleotide substrates **3** to **10** and **12** were thus assigned by analogy (Fig. 3).

Conclusion

CPAs based on both peptidic frameworks and C2-symmetric BINOL-derived scaffolds catalyze phosporamidite coupling reactions with substantial control of stereochemistry at phosphorus (III). Their effectiveness, and indeed their complementarity, are revealed in their capacity to dictate stereodivergence in coupling reactions with readily available nucleotide reaction partners. These catalysts were applied to the stereocontrolled synthesis of CDNs, which not only demonstrates streamlined access to these compounds, but also reveals a capacity for these catalytic reactions to be optimized in the presence of highly functionalized salt forms of nucleotide substrates. Overall, each class of catalyst, and the CPA framework itself, unambiguously provides a new catalytic approach to the synthesis of P-stereogenic nucleotide derivatives, which is of broad fundamental significance and utility.

REFERENCES AND NOTES

1. L. P. Jordheim, D. Durantel, F. Zoulim, C. Dumontet, *Nat. Rev. Drug Discov.* **12**, 447–464 (2013).
2. P. Clivio, S. Coantic-Castex, D. Guillaume, *Chem. Rev.* **113**, 7354–7401 (2013).
3. H. Ishikawa, G. N. Barber, *Nature* **455**, 674–678 (2008).
4. D. L. Burdette *et al*., *Nature* **478**, 515–518 (2011).
5. L. Sun, J. Wu, F. Du, X. Chen, Z. J. Chen, *Science* **339**, 786–791 (2013).
6. H. Zhang, Q. D. You, X. L. Xu, *J. Med. Chem.* **63**, 3785–3816 (2020).
7. L. Li *et al*., *Nat. Chem. Biol.* **10**, 1043–1048 (2014).
8. A. V. Lebedev, E. Wickstrom, *Perspect. Drug Discov. Des.* **4**, 17–40 (1996).
9. B. Roy, A. Depaix, C. Périgaud, S. Peyrottes, *Chem. Rev.* **116**, 7854–7897 (2016).
10. N. Iwamoto *et al*., *Nat. Biotechnol.* **35**, 845–851 (2017).
11. T. Lioux *et al*., *J. Med. Chem.* **59**, 10253–10267 (2016).
12. K. W. Knouse *et al*., *Science* **361**, 1234–1238 (2018).
13. Y. Hayakawa, M. Kataoka, *J. Am. Chem. Soc.* **119**, 11758–11762 (1997).
14. N. Oka, T. Wada, K. Saigo, *J. Am. Chem. Soc.* **124**, 4962–4963 (2002).
15. N. Oka, T. Wada, K. Saigo, *J. Am. Chem. Soc.* **125**, 8307–8317 (2003).
16. D. Xu *et al*., *J. Am. Chem. Soc.* **142**, 5785–5792 (2020).
17. D. A. DiRocco *et al*., *Science* **356**, 426–430 (2017).
18. B. R. Sculimbrene, S. J. Miller, *J. Am. Chem. Soc.* **123**, 10125–10126 (2001).
19. C. M. Longo, Y. Wei, M. F. Roberts, S. J. Miller, *Angew. Chem. Int. Ed.* **48**, 4158–4161 (2009).
20. B. H. Dahl, J. Nielsen, O. Dahl, *Nucleic Acids Res.* **15**, 1729–1743 (1987).
21. M. A. Russell, A. P. Laws, J. H. Atherton, M. I. Page, *Org. Biomol. Chem.* **6**, 3270–3275 (2008).
22. P. A. Jordan, K. J. Kayser-Bricker, S. J. Miller, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 20620–20624 (2010).
23. P. A. Jordan, S. J. Miller, *Angew. Chem. Int. Ed.* **51**, 2907–2911 (2012).
24. Y. Hayakawa *et al*., *J. Am. Chem. Soc.* **123**, 8165–8176 (2001).
25. C. Vargeese *et al*., *Nucleic Acids Res.* **26**, 1046–1050 (1998).
26. D. Parmar, E. Sugiono, S. Raja, M. Rueping, *Chem. Rev.* **114**, 9047–9153 (2014).
27. C. R. Shugrue, S. J. Miller, *Angew. Chem. Int. Ed.* **54**, 11173–11176 (2015).
28. A. L. Featherston, C. R. Shugrue, B. Q. Mercado, S. J. Miller, *ACS Catal.* **9**, 242–252 (2019).
29. Y. Kwon *et al*., *J. Am. Chem. Soc.* **141**, 6698–6705 (2019).
30. P. B. Brady, E. M. Morris, O. S. Fenton, B. R. Sculimbrene, *Tetrahedron Lett.* **50**, 975–978 (2009).
31. A. J. Metrano *et al*., *Chem. Rev.* **120**, 11479–11615 (2020).
32. T. S. Haque, J. C. Little, S. H. Gellman, *J. Am. Chem. Soc.* **118**, 6975–6985 (1996).
33. S. Krautwald, E. M. Carreira, *J. Am. Chem. Soc.* **139**, 5627–5639 (2017).
34. T. Akiyama, J. Itoh, K. Yokota, K. Fuchibe, *Angew. Chem. Int. Ed.* **43**, 1566–1568 (2004).
35. D. Uraguchi, M. Terada, *J. Am. Chem. Soc.* **126**, 5356–5357 (2004).
36. J. Steinreiber, K. Faber, H. Griengl, *Chemistry* **14**, 8060–8072 (2008).
37. W. B. Farnham, R. A. Lewis, R. K. Murray Jr., K. Mislow, *J. Am. Chem. Soc.* **92**, 5808–5809 (1970).
38. M. Dutartre, J. Bayardon, S. Jugé, *Chem. Soc. Rev.* **45**, 5771–5794 (2016).
39. J. L. Scotson, B. I. Andrews, A. P. Laws, M. I. Page, *Org. Biomol. Chem.* **14**, 8301–8308 (2016).
40. According to the Cahn-Ingold-Prelog (CIP) priority rules, phosphate and its corresponding thiophosphate analog have opposite *R*/*S* stereochemical assignments for the same absolute stereochemical configuration. That is to say, the (R_P)-phosphate has the same absolute stereochemical assignment as the (S_P)-thiophosphate.
41. W. J. Stec, G. Zon, W. Egan, *J. Am. Chem. Soc.* **106**, 6077–6079 (1984).
42. R. I. Hara *et al*., *J. Org. Chem.* **84**, 7971–7983 (2019).
43. J. Nielsen, M. H. Caruthers, *J. Am. Chem. Soc.* **110**, 6275–6276 (1988).
44. F. Seela, U. Kretschmer, *J. Org. Chem.* **56**, 3861–3869 (1991).
45. B. L. Gaffney, E. Veliath, J. Zhao, R. A. Jones, *Org. Lett.* **12**, 3269–3271 (2010).
46. B. L. Gaffney, R. A. Jones, *Curr. Protoc. Nucleic Acid Chem.* Chapter 14: Unit 14.8.1–14.8.7 (2012).
47. M. Yoshikawa *et al*., U.S. Patent WO2018100558 (2018).
48. M. Sobkowski, J. Stawinski, A. Kraszewski, *Nucleosides Nucleotides Nucleic Acids* **29**, 628–645 (2010).
49. N. Powles, J. Atherton, M. I. Page, *Org. Biomol. Chem.* **10**, 5940–5947 (2012).
50. J. Lim, H.-Y. Kim, *ACS Omega* **5**, 14173–14179 (2020).

ACKNOWLEDGMENTS

We thank N. Ichiishi (Takeda), K. Murai (Takeda), and P. Rapp for helpful discussions; E. Paulson for assistance with NMR spectroscopy; F. Menges for assistance with mass spectroscopy; B. Mercado for x-ray analysis; and J. Lim (Merck) and H.-Y. Kim (Merck) for providing additional information regarding HPLC methods to assist in the stereochemical assignment of 2′,3′-cG^{S}A^{S}MP. **Funding:** This work was supported by Takeda Pharmaceuticals International Co. and the National Institute of General Medical Sciences (NIH grant R35 GM132092 to S.J.M.). **Author contributions:** All authors contributed to the conceptualization of the project, the planning of experiments, and the interpretation of results and made intellectual contributions. D.K.L. and S.J.M. directed the research. Y.K., A.L.F., and M.M.P. conducted the experiments. All authors contributed to writing the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Crystallographic data are available free of charge from the Cambridge Crystallographic Data Centre under CCDC deposition no. 2040367. All experimental data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/371/6530/702/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S28
Tables S1 to S10
NMR Spectra
References (*51*–*74*)

26 October 2020; accepted 13 January 2021
10.1126/science.abf4359

CORONAVIRUS

Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK

Louis du Plessis[1]*, John T. McCrone[2]*, Alexander E. Zarebski[1]*, Verity Hill[2]*, Christopher Ruis[3,4]*, Bernardo Gutierrez[1,5], Jayna Raghwani[1], Jordan Ashworth[2], Rachel Colquhoun[2], Thomas R. Connor[6,7], Nuno R. Faria[1,8], Ben Jackson[2], Nicholas J. Loman[9], Áine O'Toole[2], Samuel M. Nicholls[9], Kris V. Parag[8], Emily Scher[2], Tetyana I. Vasylyeva[1], Erik M. Volz[8], Alexander Watts[10,11], Isaac I. Bogoch[12,13], Kamran Khan[10,11,12], COVID-19 Genomics UK (COG-UK) Consortium†, David M. Aanensen[14,15], Moritz U. G. Kraemer[1]‡, Andrew Rambaut[2]‡§, Oliver G. Pybus[1,16]‡§

The United Kingdom's COVID-19 epidemic during early 2020 was one of world's largest and was unusually well represented by virus genomic sampling. We determined the fine-scale genetic lineage structure of this epidemic through analysis of 50,887 severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genomes, including 26,181 from the UK sampled throughout the country's first wave of infection. Using large-scale phylogenetic analyses combined with epidemiological and travel data, we quantified the size, spatiotemporal origins, and persistence of genetically distinct UK transmission lineages. Rapid fluctuations in virus importation rates resulted in >1000 lineages; those introduced prior to national lockdown tended to be larger and more dispersed. Lineage importation and regional lineage diversity declined after lockdown, whereas lineage elimination was size-dependent. We discuss the implications of our genetic perspective on transmission dynamics for COVID-19 epidemiology and control.

Infectious disease epidemics are composed of chains of transmission, yet surprisingly little is known about how co-circulating transmission lineages vary in size, spatial distribution, and persistence, or how key properties such as epidemic size and duration arise from their combined action. Although individual-level contact-tracing investigations can reconstruct the structure of small-scale transmission clusters [e.g., (*1*–*3*)], they cannot be extended practically to large national epidemics. However, recent studies of Ebola, Zika, influenza, and other viruses have demonstrated that virus emergence and spread can instead be tracked using large-scale pathogen genome sequencing [e.g., (*4*–*7*)]. Such studies show that regional epidemics can be highly dynamic at the genetic level, with recurrent importation and extinction of transmission chains within a given location. In addition to measuring genetic diversity, understanding pathogen lineage dynamics can help researchers to target interventions effectively [e.g., (*8*, *9*)], track variants with potentially different phenotypes [e.g., (*10*, *11*)], and improve the interpretation of incidence data [e.g., (*12*, *13*)].

The rate and scale of virus genome sequencing worldwide during the COVID-19 pandemic has been unprecedented, with >100,000 severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genomes shared online by 1 October 2020 (*14*). About half of these represent infections in the United Kingdom and were generated by the national COVID-19 Genomics UK (COG-UK) consortium (*15*). The UK experienced one of the largest epidemics worldwide during the first half of 2020. Numbers of positive SARS-CoV-2 tests rose in March and peaked in April; by 26 June, there had been 40,453 nationally notified COVID-19 deaths in the UK [deaths occurring ≤28 days after first positive test (*16*)]. Here, we combine this large genomic dataset with epidemiological and travel data to provide a full characterization of the genetic structure and lineage dynamics of the UK epidemic.

Our study encompasses the initial epidemic wave of COVID-19 in the UK and comprises all SARS-CoV-2 genomes available before 26 June 2020 (50,887 genomes, of which 26,181 were from the UK; Fig. 1A) (*17*). The data represent genomes from 9.29% of confirmed UK COVID-19 cases by 26 June (*16*). Further, using an estimate of the actual size of the UK epidemic (*18*), we infer that virus genomes were generated for 0.66% [95% confidence interval (CI), 0.46 to 0.95%] of all UK infections by 5 May (Fig. 1B).

Genetic structure and lineage dynamics of the UK epidemic from January to June

We first sought to identify and enumerate all independently introduced, genetically distinct chains of infection within the UK. We developed a large-scale molecular clock phylogenetic pipeline to identify "UK transmission lineages" that (i) contain two or more UK genomes and (ii) descend from an ancestral lineage inferred to exist outside of the UK (Fig. 2, A and B). Sources of statistical uncertainty in lineage assignation were taken into account (*17*). We identified a total of 1179 [95% highest posterior density (HPD), 1143 to 1286] UK transmission lineages. Although each is intended to capture a chain of local transmission arising from a single importation event, some UK transmission lineages will be unobserved or aggregated as a result of limited SARS-CoV-2 genetic diversity (*19*) or incomplete or uneven genome sampling (*20*, *21*). Therefore we expect this number to be an

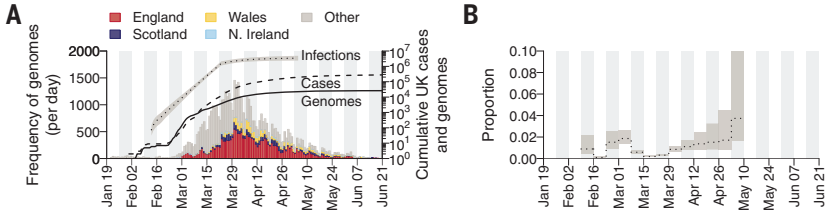


Fig. 1. Genomic sequence data. (**A**) Collection dates of the 50,887 genomes analyzed here (left axis). Genomes are colored by sampling location (red, England; dark blue, Scotland; yellow, Wales; light blue, Northern Ireland; gray, elsewhere). The solid line shows the cumulative number of UK virus genomes (right axis). The dashed and dotted lines show, respectively, the cumulative number of laboratory-confirmed UK cases (by specimen date) and the estimated number of UK infections (*18*); gray shading denotes the 95% CI. As a result of retrospective screening, the cumulative number of genomes early in the epidemic exceeds that of confirmed cases. (**B**) Proportion of weekly estimated UK infections (*18*) included in our genome sequence dataset.

[1]Department of Zoology, University of Oxford, Oxford, UK. [2]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK. [3]Molecular Immunity Unit, Department of Medicine, University of Cambridge, Cambridge, UK. [4]Department of Veterinary Medicine, University of Cambridge, Cambridge, UK. [5]School of Biological and Environmental Sciences, Universidad San Francisco de Quito, Quito, Ecuador. [6]School of Biosciences, Cardiff University, Cardiff, UK. [7]Pathogen Genomics Unit, Public Health Wales NHS Trust, Cardiff, UK. [8]MRC Centre for Global Infectious Disease Analysis, J-IDEA, Imperial College London, London, UK. [9]Institute of Microbiology and Infection, University of Birmingham, Birmingham, UK. [10]Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Canada. [11]BlueDot, Toronto, Canada. [12]Department of Medicine, University of Toronto, Toronto, Canada. [13]Divisions of General Internal Medicine and Infectious Diseases, University Health Network, Toronto, Canada. [14]Centre for Genomic Pathogen Surveillance, Wellcome Genome Campus, Hinxton, UK. [15]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [16]Department of Pathobiology and Population Sciences, Royal Veterinary College London, London, UK.
*These authors contributed equally to this work. †See supplementary materials for list of consortium members and affiliations. ‡These authors contributed equally to this work.
§Corresponding author. Email: a.rambaut@ed.ac.uk (A.R.); oliver.pybus@zoo.ox.ac.uk (O.G.P.)

underestimate (*17*). In our phylogenetic analysis, 1650 (95% HPD, 1611 to 1783) UK genomes could not be allocated to a UK transmission lineage (singletons). Had more genomes been sequenced, it is likely that many of these singletons would have been assigned to a UK transmission lineage. Further, many singleton importations are likely to be unobserved.

Most transmission lineages are small, and 72.4% (95% HPD, 69.3 to 72.9%) contain <10 genomes (Fig. 2C). However, the lineage size distribution is strongly skewed and follows a power-law distribution (Fig. 2C, inset), such that the eight largest UK transmission lineages contain >25% of all sampled UK genomes (Fig. 2D; figs. S2 to S5 show further visualizations). Although the two largest transmission lineages are estimated to comprise >1500 UK genomes each, there is phylogenetic uncertainty in their sizes (95% HPDs, 1280 to 2133 and 1342 to 2011 genomes, respectively). Because our dataset constitutes only a small fraction of all UK infections, these observed lineage sizes will underestimate true lineage size. However, the true distribution of relative lineage sizes will closely match our observation, and its power-law shape indicates that almost all unobserved lineages will be small. All eight largest lineages were first detected before the UK national lockdown was announced on 23 March and, as expected, larger lineages were observed for longer (Pearson's r = 0.82; 95% CI, 0.8 to 0.83; fig. S7). The sampling frequency of lineages of varying sizes differed over time (Fig. 3A and figs. S8 and S9); whereas UK transmission lineages containing >100 genomes consistently accounted for >40% of weekly sampled genomes, the proportion of small transmission lineages (≤10 genomes) and singletons decreased over the course of the epidemic (Fig. 3A).

The detection of UK transmission lineages in our data changed markedly through time. In early March, the epidemic was characterized by lineages first observed within the previous week (Fig. 3B). The per-genome rate of appearance of new lineages was initially high, then declined throughout March and April (Fig. 3C), such that by 1 May, 96.2% of sampled genomes belonged to transmission lineages that were first observed >7 days previously. By 1 June, a growing number of lineages (>73%) had not been detected by genomic sampling for >4 weeks, which suggests that they were rare or had gone extinct; this result is robust to the sampling rate (Fig. 1, A and B, and Fig. 3C). Together, these results indicate that the UK's first epidemic wave resulted from the concurrent growth of many hundreds of independently introduced transmission lineages, and that the introduction of nonpharmaceutical interventions (NPIs) was followed by the apparent extinction of lineages in a size-dependent manner.

Transmission lineage diversity and geographic range

We also characterized the spatial distribution of UK transmission lineages using available data on 107 virus genome sampling locations, which correspond broadly to UK counties or metropolitan regions (data S1). Although genomes were not collected randomly [some lineages and regions will be overrepresented because of targeted investigation of local outbreaks; e.g., (*22*)], the number of UK lineages detected in each region correlates with the number of genomes sequenced (Fig. 4A; Pearson's r = 0.96; 95% CI, 0.95 to 0.98) and the number of reported cases (fig. S10; Pearson's r = 0.53; 95% CI, 0.35 to 0.67; see also data S2) in each region. Further, larger lineages were observed in more locations; every 100 additional genomes in a lineage increases its observed range by six or seven regions (Fig. 4B; Pearson's r = 0.8; 95% CI, 0.78 to 0.82). Thus, bigger regional epidemics comprised a greater diversity of

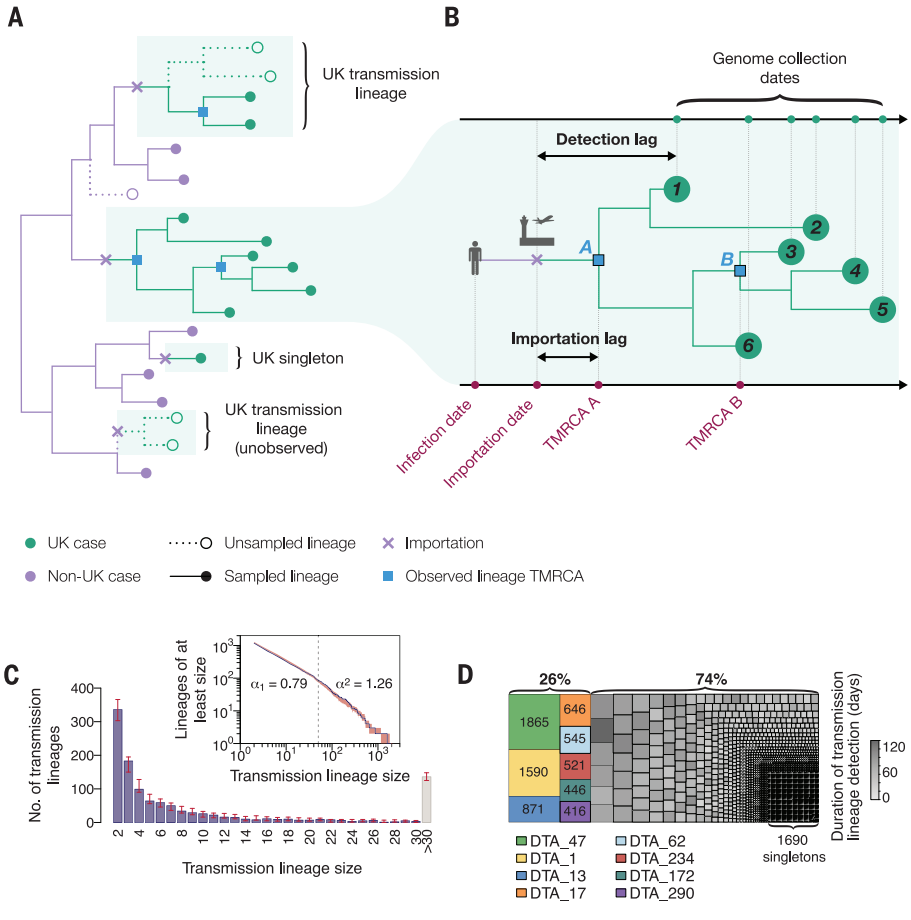


Fig. 2. Structure of UK transmission lineages detected through genome sampling. (**A**) Figurative illustration of the international context of UK transmission lineages. Note that only half of the cases in the top UK transmission lineage are observed, and the bottom UK transmission lineage is unobserved. To be detected, a UK transmission lineage must contain two or more sampled genomes; singletons are not classified here as UK transmission lineages. (**B**) Detailed view of one of the UK transmission lineages from (A), used to illustrate the terms TMRCA, detection lag, and importation lag. The lineage TMRCA is sample-dependent; for example, TMRCA A is observed if genomes 1 to 6 are sampled, and TMRCA B is observed if only genomes 3 to 5 are sampled. (**C**) Distribution of UK transmission lineage sizes. Blue bars show the number of transmission lineages of each size; error bars are 95% HPDs of these sizes across the posterior tree distribution. The inset shows the corresponding cumulative frequency distribution of lineage size (blue line) on double logarithmic axes; red shading denotes the 95% HPD of this distribution across the posterior tree distribution. Values to either side of the vertical dashed line show coefficients of power-law distributions [$P(X \geq x) \sim x^{-\alpha}$] fitted to lineages containing ≤50 (α_1) and >50 (α_2) virus genomes, respectively. (**D**) Partition of 26,181 UK genomes into UK transmission lineages and singletons, colored by (i) lineage, for the eight largest lineages, or (ii) duration of lineage detection (time between the lineage's oldest and most recent genomes) for the remainder. The sizes of the eight largest lineages are also shown.

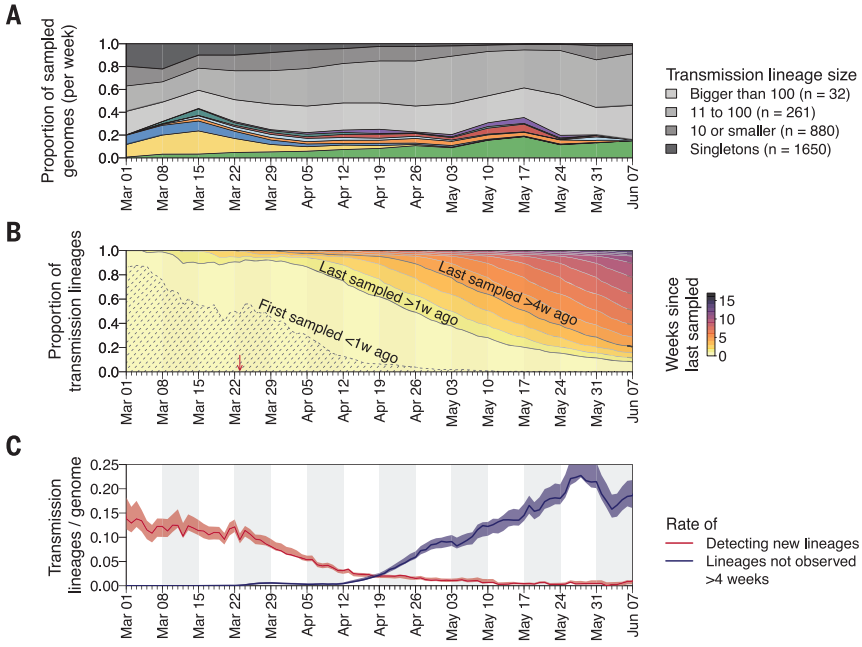


Fig. 3. Dynamics of UK transmission lineages. (**A**) Lineage size breakdown of UK genomes collected each week. Colors of the eight largest lineages are as depicted in Fig. 2D. (**B**) Trends through time in the detection of UK transmission lineages. For each day, all lineages detected up to that day are colored by the time since the transmission lineage was last sampled. Isoclines correspond to weeks. Shaded area denotes transmission lineages that were first sampled less than 1 week ago. The red arrow indicates the start of the UK lockdown. (**C**) The daily rate of detecting new transmission lineages (red line) and the rate at which lineages have not been observed for >4 weeks (blue line); shading denotes the 95% HPD across the posterior distribution of trees.

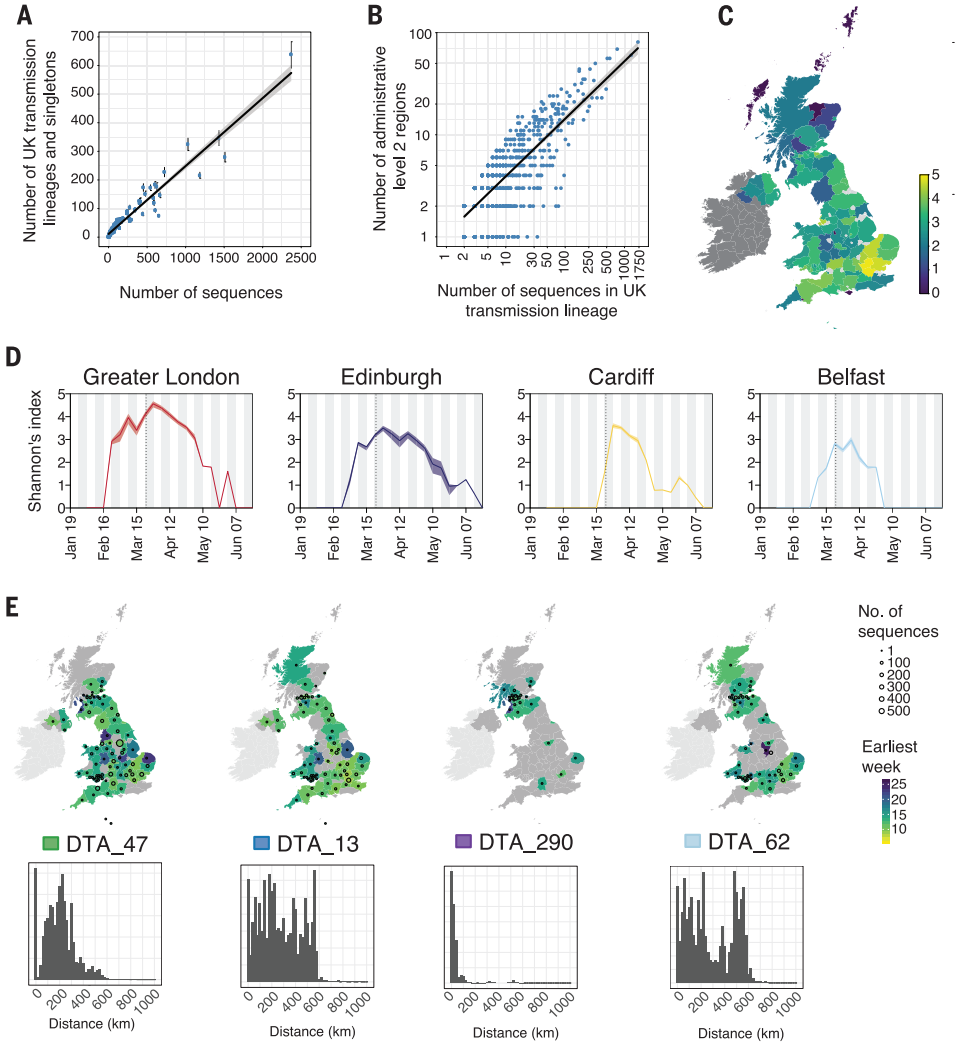


Fig. 4. Spatial distribution of UK transmission lineages. (**A**) Correlation between the number of transmission lineages detected in each region (points, median values; bars, 95% HPD intervals) and the number of UK virus genomes from each region (Pearson's r = 0.96; 95% CI, 0.95 to 0.98). (**B**) Correlation between the spatial range of each transmission lineage and the number of virus genomes it contains (Pearson's r = 0.8; 95% CI, 0.78 to 0.82). (**C**) Map showing Shannon's index (SI) for each region, calculated across the study period (2 February to 26 June). Yellow colors indicate higher SI values; darker colors, lower values. (**D**) SI through time for the UK national capital cities. The dotted lines indicate the start of the UK national lockdown. (**E**) Illustration of the diverse spatial range distributions of UK transmission lineages. Colors represent the week of the first detected genome in the transmission lineage in each location. Circles show the number of sampled genomes per location. Histograms (bottom row) show the distribution of geographic distances for all sequence pairs within each lineage (see data S4 and fig. S12 for further details). Colored boxes next to lineage names are as depicted in Fig. 2D.

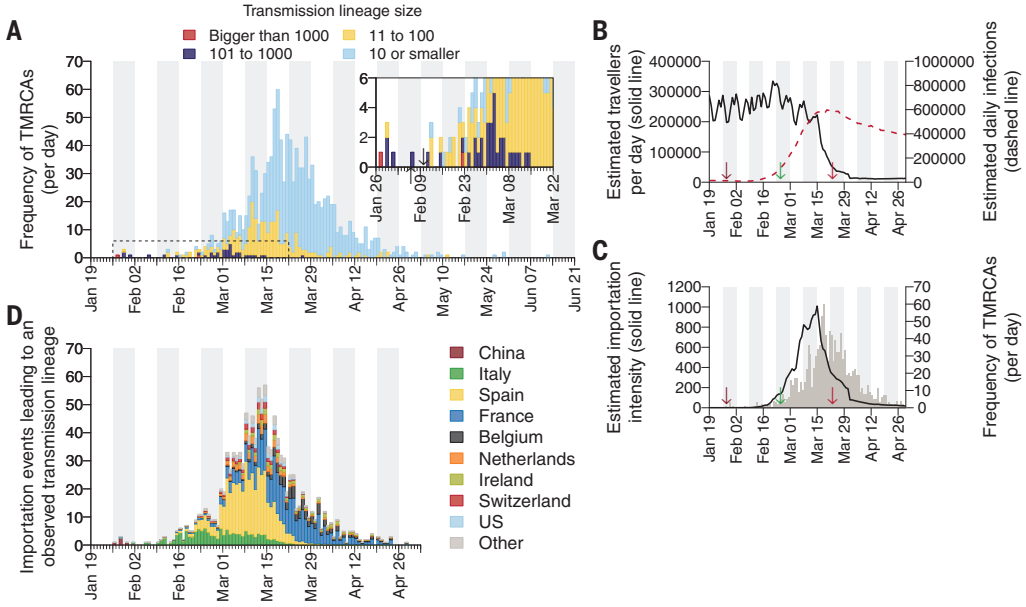


Fig. 5. Dynamics of UK transmission lineage importation. (**A**) Histogram of lineage TMRCAs, colored by lineage size. The inset is an expanded view of the days prior to UK lockdown; the upward arrow indicates the collection date of the UK's first laboratory-confirmed case, and the downward arrow shows the collection date of the earliest UK virus genome in our dataset. (**B**) Estimated number of inbound travelers to the UK per day (black line) and estimated number of infectious cases worldwide (dashed red line). Arrows show, from left to right, the dates of the first self-isolation advice for returning travelers from China, the same for Italy, and the start of the UK national lockdown. (**C**) Estimated importation intensity (EII) curve (black) and histogram of lineage TMRCAs (gray). (**D**) Estimated histogram of virus lineage importation events per day, obtained from our lag model. Colors show the proportion attributable each day to inbound travel from various countries (see table S4 and figs. S19 and S20). This assignment is statistical (i.e., we cannot ascribe a specific source location to any given lineage).

transmission lineages, and larger lineages were more geographically widespread. These observations indicate substantial dissemination of a subset of lineages across the UK and suggest that many regions experienced a series of introductions of new lineages from elsewhere, potentially hindering the impact of local interventions.

We quantified the substantial variation among regions in the diversity of transmission lineages present using Shannon's index (SI; this value increases as both the number of lineages and the evenness of their frequencies increase; Fig. 4C and data S3). We observed the highest SIs in Hertfordshire (4.77), Greater London (4.62), and Essex (4.49); these locations are characterized by frequent commuter travel to or within London and proximity to major international airports (*23*). Locations with the three lowest nonzero SIs were in Scotland (Stirling = 0.96, Aberdeenshire = 1.04, Inverclyde = 1.32; Fig. 4C). We speculate that regional differences in transmission lineage diversity may be related to the level of connectedness to other regions.

To illustrate temporal trends in transmission lineage diversity, we plotted SI through time for each of the UK's national capital cities (Fig. 4D). Lineage diversities in each peaked in late March and declined after the UK national lockdown, congruent with Fig. 3, C and D. Greater London's epidemic was the most diverse and was characterized by an early, rapid rise in SI (Fig. 4D), consistent with epidemiological trends there (*16*, *24*). Belfast's lineage diversity was notably lower (data S4 shows other locations).

We observe variation in the spatial range of individual UK transmission lineages. Although some lineages are widespread, most are more localized and the range size distribution is right-skewed (fig. S11), congruent with an observed abundance of small lineages (Figs. 2C and 4B) and biogeographic theory [e.g., (*25*)]. For example, lineage DTA_13 is geographically dispersed (>50% of sequence pairs sampled >234 km apart), whereas DTA_290 is strongly local (95% of sequence pairs sampled <100 km apart) and DTA_62 has multiple foci of sampled genomes (Fig. 4E and fig. S12). The national distribution of cases therefore arose from the aggregation of multiple heterogeneous lineage-specific patterns.

Dynamics of international introduction of transmission lineages

The process by which transmission lineages are introduced to an area is an important aspect of early epidemic growth [e.g., (*26*)]. To investigate this at a national scale, we estimated the rate and source of SARS-CoV-2 importations into the UK. Because standard phylogeographic approaches were precluded by strong biases in genome sampling among countries (*20*), we developed a new approach that combines virus phylogenetics with epidemiological and travel data. First, we estimated the TMRCA (time of the most recent common ancestor) of each UK transmission lineage (*17*). The TMRCAs of most UK lineages are dated to March and early April [median = 21 March; interquartile range (IQR) = 14 to 29 March]. UK lineages with earlier TMRCAs tend to be larger and longer-lived than those whose TMRCAs postdate the national lockdown (Fig. 5A and fig. S15).

Because of incomplete sampling, TMRCAs best represent the date of the first inferred transmission event in a lineage, not its importation date (Fig. 2B). To infer the latter and to quantify the delay between importation and onward within-UK transmission, we generated daily estimates of the number of travelers arriving in the UK and of global SARS-CoV-2 infections (*17*) worldwide. Before March, the UK received ~1.75 million inbound travelers per week (school holidays explain the end-February ~10% increase; Fig. 5B). International arrivals fell by ~95% during March, and this reduction was maintained through April. Elsewhere, estimated numbers of infectious cases peaked in late March (Fig. 5B). We combined these two trends to generate an estimated importation intensity (EII), a daily empirical measure of the intensity of SARS-CoV-2 importation into the UK (*17*). Because both travel volumes and epidemic incidence fluctuate rapidly over orders of magnitude, the EII is robust to other sources of variation in the relative importation risk among countries (*17*). The EII peaked in mid-March, when high UK inbound travel volumes coincided with growing numbers of infectious cases elsewhere (Fig. 5, B and C).

Crucially, the EII's temporal profile closely matches, but precedes, that of the TMRCAs of UK transmission lineages (Fig. 5, A and C). The difference between the two represents the "importation lag," the time elapsed between lineage importation and the first detected local transmission event (Fig. 2B). Using a statistical model (*17*), we estimate importation lag to be on average 8.22 ± 5.21 days (IQR = 3.35 to 15.18) across all transmission lineages. Further, importation lag is strongly size-dependent; average lag is ~10 days for lineages comprising ≤10 genomes and <1 day

for lineages of >100 genomes (table S2). This size dependency likely arises because the earliest transmission event in a lineage is more likely to be captured if it contains many genomes (Fig. 2B) (*17*). We use this model to impute an importation date for each UK transmission lineage (Fig. 5D). Importation was unexpectedly dynamic, rising and falling substantially over only 4 weeks; hence, 80% of importations (that gave rise to detectable UK transmission lineages) occurred between 27 February and 30 March. The delay between the inferred date of importation and the first genomic detection of each lineage was 14.13 ± 5.61 days on average (IQR = 10 to 18) and declined through time (tables S2 and S3).

To investigate country-specific contributions to virus importation, we generated separate EII curves for each country (fig. S17). Using these values, we estimated the numbers of inferred importations each day attributable to inbound travel from each source location. This assignment is statistical and does not take the effects of superspreading events into account. As with the rate of importation (Fig. 5A), the relative contributions of arrivals from different countries were dynamic (Fig. 5D). Dominant source locations shifted rapidly in February and March, and the diversity of source locations increased in mid-March (fig. S17). The earliest importations were most likely from China or elsewhere in Asia but were rare relative to those from Europe. Over our study period, we infer that ~33% of UK transmission lineages stemmed from arrivals from Spain, 29% from France, 12% from Italy, and 26% from elsewhere (fig. S20 and table S4). These large-scale trends were not apparent from individual-level travel histories; routine collection of such data ceased on 12 March (*27*).

Conclusions

The exceptional size of our genomic survey provides insight into the micro-epidemiological patterns that underlie the features of a large, national COVID-19 epidemic, allowing us to quantify the abundance, size distribution, and spatial range of transmission lineages. Before the lockdown, high travel volumes and few restrictions on international arrivals (Fig. 5B and table S5) led to the establishment and co-circulation of >1000 identifiable UK transmission lineages (Fig. 5A), jointly contributing to accelerated epidemic growth that quickly exceeded national contact-tracing capacity (*27*). The relative contributions of importation and local transmission to initial epidemic dynamics under such circumstances warrant further investigation. We expect that similar trends occurred in other countries with comparably large epidemics and high international travel volumes; virus genomic studies from regions with smaller or controlled COVID-19 epidemics have reported high importation rates followed by more transient lineage persistence [e.g., (*28*–*30*)].

Earlier lineages were larger, more dispersed, and harder to eliminate, highlighting the importance of rapid or preemptive interventions in reducing transmission [e.g., (*31*–*33*)]. The high heterogeneity in SARS-CoV-2 transmission at the individual level (*34*–*36*) appears to extend to whole transmission lineages, such that >75% of sampled viruses belong to the top 20% of lineages ranked by size. Although the national lockdown coincided with limited importation and reduced regional lineage diversity, its impact on lineage extinction was size-dependent (Fig. 3, B and C). The overdispersed nature of SARS-CoV-2 transmission likely exacerbated this effect (*37*), thereby favoring, as the epidemic reproduction number (R_t) declined, greater survival of larger widespread lineages and faster local elimination of lineages in low-prevalence regions. The degree to which the surviving lineages contributed to the UK's ongoing second epidemic, including the effect of specific mutations on lineage growth rates [e.g., (*11*)], is currently under investigation. The transmission structure and dynamics measured here provide a new context in which future public health actions at regional, national, and international scales should be planned and evaluated.

REFERENCES AND NOTES

1. Centers for Disease Control and Prevention, *MMWR Morb. Mortal. Wkly. Rep.* **52**, 405–411 (2003).
2. O. Faye *et al*., *Lancet Infect. Dis.* **15**, 320–326 (2015).
3. K. H. Kim, T. E. Tandi, J. W. Choi, J. M. Moon, M. S. Kim, *J. Hosp. Infect.* **95**, 207–213 (2017).
4. J. Bahl *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **108**, 19359–19364 (2011).
5. G. J. Baillie *et al*., *J. Virol.* **86**, 11–18 (2012).
6. G. Dudas *et al*., *Nature* **544**, 309–315 (2017).
7. N. D. Grubaugh *et al*., *Nature* **546**, 401–405 (2017).
8. A. F. Y. Poon *et al*., *Lancet HIV* **3**, e231–e238 (2016).
9. J. Thomas *et al*., *N. Engl. J. Med.* **382**, 632–643 (2020).
10. M. A. Beale *et al*., *Nat. Commun.* **10**, 3255 (2019).
11. E. M. Volz *et al*., *Cell* **184**, 1–12 (2021).
12. L. M. Li, N. C. Grassly, C. Fraser, *Mol. Biol. Evol.* **34**, 2982–2995 (2017).
13. N. D. Grubaugh *et al*., *Nat. Microbiol.* **4**, 10–19 (2019).
14. Y. Shu, J. McCauley, *Euro Surveill.* **22**, 30494 (2017).
15. COVID-19 Genomics UK (COG-UK) Consortium, *Lancet Microbe* **1**, e99–e100 (2020).
16. U.K. Government, Coronavirus (COVID-19) in the UK; https://coronavirus.data.gov.uk/cases.
17. See supplementary materials.
18. S. Flaxman *et al*., *Nature* **584**, 257–261 (2020).
19. C. J. Villabona-Arenas, W. P. Hanage, D. C. Tully, *Nat. Microbiol.* **5**, 876–877 (2020).
20. M. Worobey *et al*., *Science* **370**, 564–570 (2020).
21. S. A. Nadeau, T. G. Vaughan, J. Sciré, J. S. Huisman, T. Stadler, medRxiv 20127738 [preprint]. 12 June 2020.
22. Welsh Government, Genomic analysis of Covid-19 lineages in Wales; https://gov.wales/genomic-analysis-covid-19-lineages-wales.
23. Greater London Authority Intelligence and Analysis Unit, "Census Information Scheme: Commuting in London" (2014); https://londondatastore-upload.s3.amazonaws.com/Zho%3Dttw-flows.pdf.
24. C. Angus, CoVid Plots and Analysis. University of Sheffield (2020); https://doi.org/10.15131/shef.data.12328226.
25. K. J. Gaston, F. He, *Proc. R. Soc. B* **269**, 1079–1086 (2002).
26. G. Chowell, L. Sattenspiel, S. Bansal, C. Viboud, *Phys. Life Rev.* **18**, 66–97 (2016).
27. C. Baraniuk, *BMJ* **369**, m1859 (2020).
28. J. L. Geoghegan *et al*., medRxiv 20168930 [preprint]. 20 August 2020.
29. J. Lu *et al*., *Cell* **181**, 997–1003.e9 (2020).
30. T. Seemann *et al*., *Nat. Commun.* **11**, 4376 (2020).
31. C. Dye, R. C. H. Cheng, J. S. Dagpunar, B. G. Williams, *R. Soc. Open Sci.* **7**, 201726 (2020).
32. H. Tian *et al*., *Science* **368**, 638–642 (2020).
33. K. Leung, J. T. Wu, D. Liu, G. M. Leung, *Lancet* **395**, 1382–1393 (2020).
34. D. C. Adam *et al*., *Nat. Med.* **26**, 1714–1719 (2020).
35. A. Endo, S. Abbott, A. J. Kucharski, S. Funk, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, *Wellcome Open Res.* **5**, 67 (2020).
36. L. Wang *et al*., *Nat. Commun.* **11**, 5006 (2020).
37. J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, W. M. Getz, *Nature* **438**, 355–359 (2005).
38. "Establishment and lineage dynamics of the SARS-Cov-2 epidemic in the UK" GitHub repository (https://github.com/COG-UK/uk-intros-analyses, DOI: 10.5281/zenodo.4311597).

ACKNOWLEDGMENTS

We are grateful to everyone worldwide involved in generating the virus genome data shared on GISAID. We thank S. Bhatt, P. Lemey, and C. Dye for insightful discussion. The contents of this publication are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission. **Funding:** COG-UK is funded by the Medical Research Council (MRC) part of UK Research & Innovation (UKRI), the National Institute of Health Research (NIHR) and Genome Research Limited, operating as the Wellcome Sanger Institute. Also supported by BBSRC grant BB/M010996/1 (V.H.); Fondation Botnar Research Award programme grant 6063 and UK Cystic Fibrosis Trust Innovation Hub Award 001 (C.R.); UKRI GCRF One Health Poultry Hub grant BB/S011269/1 (J.R.); a Branco Weiss Fellowship and EU grant 874850 MOOD (M.U.G.K.); WT fellowship 204311/Z/16/Z and MRC-FAPESP awards MR/S0195/1 and 18/14389-0 (N.R.F.); a Branco Weiss Fellowship (T.I.V.); Canadian Institutes for Health Research grant 02179-000 (I.I.B.); WT Collaborators Award 206298/Z/17/Z (J.T.M., R.M.C., N.J.L., and A.R.); ERC grant 725422 (A.R. and E.S.); NIHR Global Health Research Unit grant 16/136/111 (D.M.A.); and the Oxford Martin School (O.G.P., M.U.G.K., L.d.P., and A.E.Z.). **Author contributions:** Study design: L.d.P., J.T.M., M.U.G.K., A.R., O.G.P. Methods development/programming: L.d.P., J.T.M., M.U.G.K., A.R., O.G.P., A.E.Z., V.H., C.R., J.A., R.C., T.C., B.J., N.J.L., A.O., S.N., D.M.A., E.S. Data analysis: L.d.P., M.U.G.K., A.R., O.G.P., A.E.Z., V.H., C.R., D.M.A., J.T.M., A.W., I.I.B., K.K., B.G., K.V.P., E.S., T.I.V. Wrote paper: L.d.P., J.R., M.U.G.K., O.G.P. Edited paper/figure creation: L.d.P., V.H., A.E.Z., M.U.G.K., J.R., C.R., B.G., T.I.V., N.R.F., E.M.V. **Competing interests:** K.K. is the founder of BlueDot, a social enterprise that develops digital technologies for public health. A.W. and I.I.B. received employment or consulting income from BlueDot during this research. **Data and materials availability:** UK SARS-CoV-2 genomes and public metadata are available from www.cogconsortium.uk/data/ and deposited at gisaid.org and from the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB37886 (www.ebi.ac.uk/ena/browser/view/PRJEB37886). Non-UK genomes were obtained from gisaid.org. Raw data, code, and analysis files for this work are in supplementary materials or available from GitHub at https://github.com/COG-UK/uk-intros-analyses (DOI: 10.5281/zenodo.4311597), which also contains a list of sequence accession numbers. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/371/6530/708/suppl/DC1
Materials and Methods
Figs. S1 to S20
Tables S1 to S5
Data S1 to S5
References (*39*–*68*)

22 October 2020; accepted 18 December 2020
Published online 8 January 2021
10.1126/science.abf2946

REPORTS

GALAXIES

A massive stellar bulge in a regularly rotating galaxy 1.2 billion years after the Big Bang

Federico Lelli[1,2]*, Enrico M. Di Teodoro[3], Filippo Fraternali[4], Allison W. S. Man[5], Zhi-Yu Zhang[6], Carlos De Breuck[7], Timothy A. Davis[1], Roberto Maiolino[8,9]

Cosmological models predict that galaxies forming in the early Universe experience a chaotic phase of gas accretion and star formation, followed by gas ejection due to feedback processes. Galaxy bulges may assemble later via mergers or internal evolution. Here we present submillimeter observations (with spatial resolution of 700 parsecs) of ALESS 073.1, a starburst galaxy at redshift $z \simeq 5$ when the Universe was 1.2 billion years old. This galaxy's cold gas forms a regularly rotating disk with negligible noncircular motions. The galaxy rotation curve requires the presence of a central bulge in addition to a star-forming disk. We conclude that massive bulges and regularly rotating disks can form more rapidly in the early Universe than predicted by models of galaxy formation.

In the standard Lambda cold dark matter (ΛCDM) cosmological model, galaxies form inside dark matter (DM) halos when primordial gas cools, collapses, and begins star formation (*1*). At early times, when the Universe was only 10% of its current age (redshift $z \simeq 5$), ΛCDM models predict that galaxies undergo a turbulent phase of gas accretion from the cosmic web, leading to violent bursts of star formation and black hole growth (*2*). These are rapidly followed by massive gas outflows due to feedback from supernovae and active galactic nuclei (AGN). Strong stellar and AGN feedback are needed in ΛCDM models to reproduce the observed number of galaxies per unit volume per unit mass (the galaxy stellar mass function) and to quench star formation in the most massive DM halos, reproducing the observed population of quiescent early-type galaxies at $z = 0$ (*3*). In the hierarchical ΛCDM scenario, galaxy bulges (the spheroidal stellar components in the central parts of massive galaxies) are expected to form either after the merger of two galaxies of similar mass or via secular dynamical processes within the stellar disk (*4*). Observation of galaxies at high redshifts enables us to test these scenarios by searching for these physical processes in action.

The 158-μm emission line of singly ionized carbon, [C II], is a major coolant of the interstellar medium of galaxies and traces a combination of atomic and molecular hydrogen phases. The [C II] line has been observed up to $z \simeq 7.5$ (*5*), and several rotating disks have been identified at $z > 3$ (*6*–*8*). Most existing observations only marginally resolve the [C II] distribution and kinematics, and thus cannot distinguish between rotating disks, galaxy mergers, or gas inflows and outflows. Limited spatial resolution has prevented the measurement of well-sampled rotation curves (with more than six independent elements), which could be used to measure different mass components by fitting dynamical models.

[1]School of Physics and Astronomy, Cardiff University, Cardiff CF24 3AA, UK. [2]Arcetri Astrophysical Observatory, Istituto Nazionale di Astrofisica, Florence 50125, Italy. [3]Department of Physics & Astronomy, Johns Hopkins University, Baltimore, MD 21218, USA. [4]Kapteyn Astronomical Institute, University of Groningen, Groningen 9700 AV, Netherlands. [5]Dunlap Institute for Astronomy & Astrophysics, University of Toronto, Toronto, ON M5S 3H4, Canada. [6]School of Astronomy and Space Science, Nanjing University, Nanjing 210023, P. R. China. [7]European Southern Observatory, Garching bei München 85748, Germany. [8]Kavli Institute for Cosmology, University of Cambridge, Cambridge CB3 0HA, UK. [9]Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK.
*Corresponding author. Email: federico.lelli@inaf.it

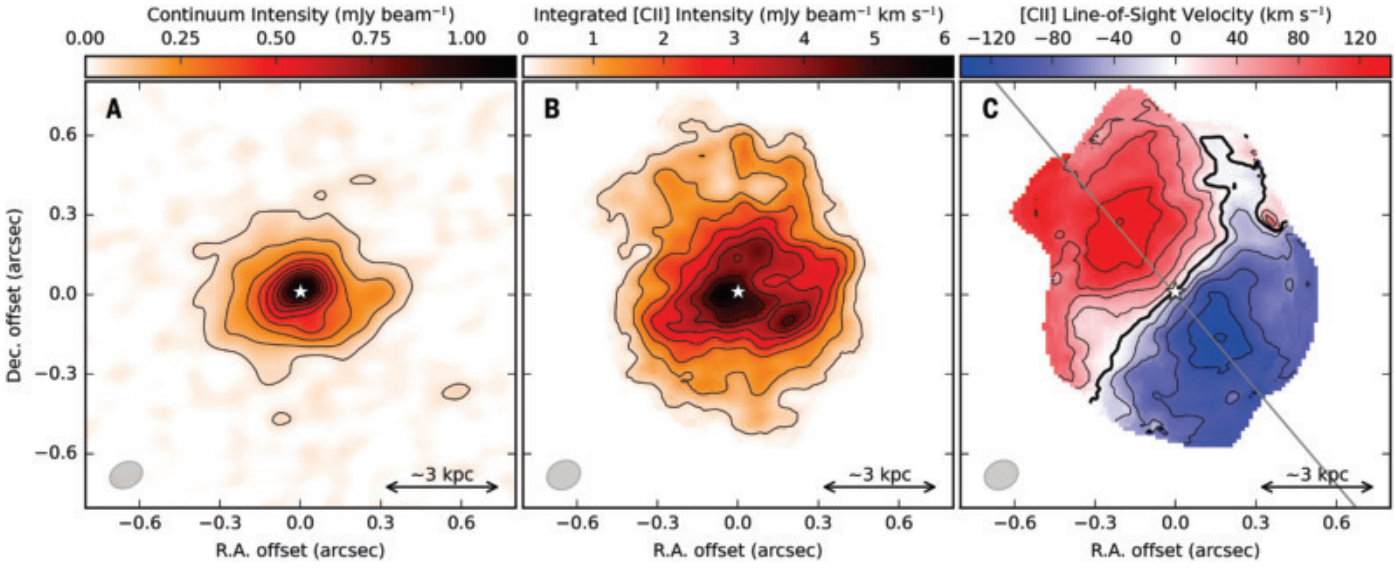


Fig. 1. ALMA observations of ALESS 073.1. (**A**) Continuum emission at 160 μm (rest-frame) tracing dust heated by young stars, (**B**) [C II] intensity map tracing cold gas, and (**C**) [C II] velocity field showing a rotating disk. North is up; east is left. The kinematic center, located at right ascension (R.A.) $03^h\ 32^m\ 29.295^s$ and declination (Dec.) −27°56′19.″60, is represented by a white star. The beam size is plotted as the gray ellipse in the bottom-left corner. The physical scale is indicated by the scale bar in the bottom-right corner. In (A), iso-emission contours range from 0.055 to 1 mJy beam^{-1} (where 1 mJy = 10^{-29} W m^{-2} Hz^{-1}) in steps of 0.11 mJy beam^{-1}. In (B), iso-emission contours range from 0.35 to 6 mJy beam^{-1} km s^{-1} in steps of 0.7 mJy beam^{-1} km s^{-1}. In (A) and (B), the lowest iso-emission contour corresponds to a signal-to-noise ratio of ~3. In (C), iso-velocity contours range from −120 to +120 km s^{-1} in steps of 30 km s^{-1}, the bold contour indicates the systemic velocity (set to zero), and the gray line shows the kinematic major axis.

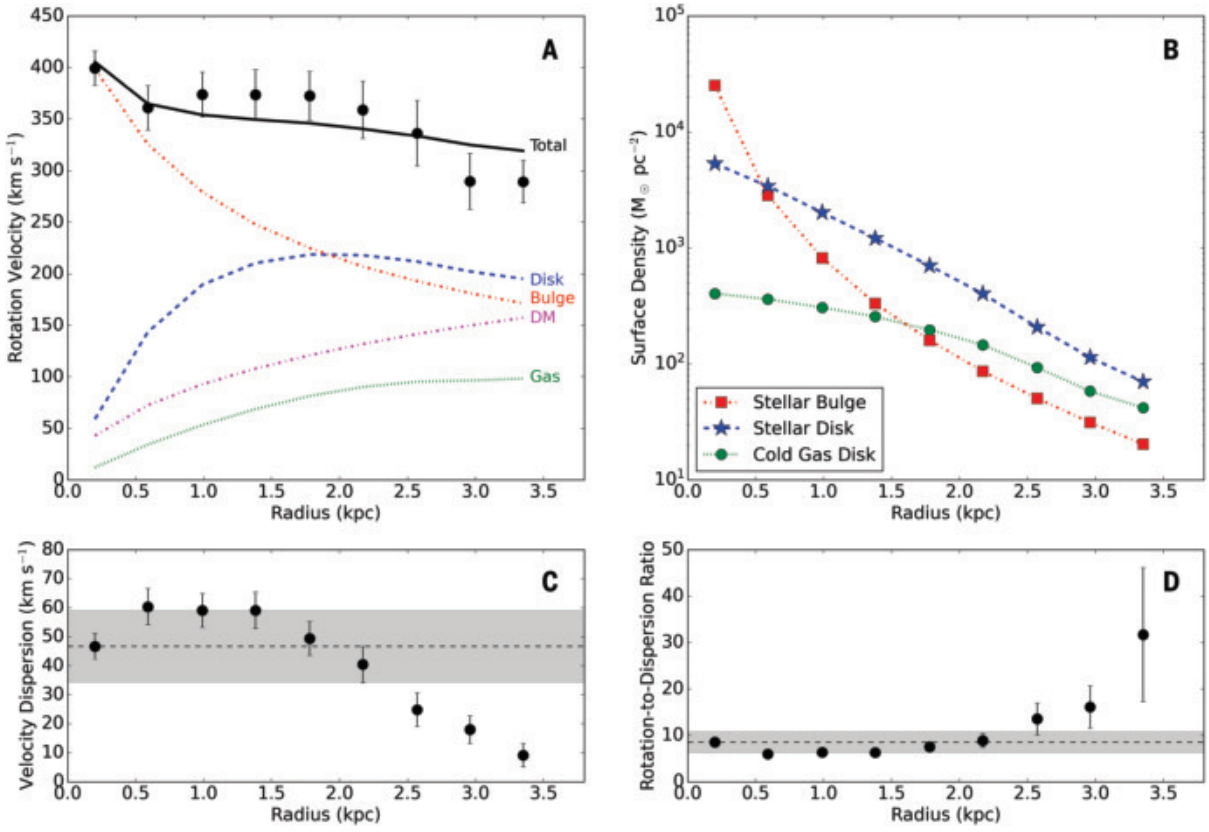


Fig. 2. Mass model of ALESS 073.1. (**A**) The observed rotation curve (black circles) adopting the best-fitting inclination angle of 22°. The model rotation curve (black solid line) is the total of contributions from the stellar bulge (red dot-dashed line), stellar disk (blue dashed line), cold gas disk (green dotted line), and DM halo (magenta dash-dotted line). (**B**) Surface density profiles of the bulge (red squares), stellar disk (blue stars), and cold gas disk (green circles) adopting the best-fitting masses (*9*). The bulge profile assumes a de Vaucouleurs' distribution (*30*). The stellar and gas disk profiles are obtained from azimuthal averages of the dust continuum and [C II] intensity maps, respectively. The uncertainties on these profiles are dominated by the systematic uncertainty on the normalizing masses (*9*). (**C**) [C II] velocity dispersion profile. (**D**) Disk rotational support as a function of radius. In (C) and (D), the dashed line and gray band show the median value and median absolute deviation, respectively. In (A), (C), and (D), the error bars show ±1σ uncertainties (*9*).

We used the Atacama Large Millimeter/submillimeter Array (ALMA) to observe a star-forming galaxy at z = 4.75 (*6*) at a spatial resolution of ~0.″11, corresponding to 0.7 kpc in a ΛCDM cosmology (*9*). The target, ALESS 073.1 (LESS J033229.3-275619), is a strong submillimeter source (*10*). Its large far-infrared luminosity suggests an ongoing starburst with an estimated star-formation rate of ~1000 solar masses ($M_\odot$) per year (*10*, *11*). ALESS 073.1 contains a dust-obscured AGN (*12*). These extreme properties indicate that the galaxy may drive a massive gas outflow (*11*). Previous [C II] observations with ALMA indicated a possible rotating disk (*6*), but the spatial resolution (0.″5, corresponding to ~3 kpc) was insufficient to determine its detailed properties or identify gas inflows and outflows.

Figure 1A shows the continuum map at rest-frame 160 μm, which traces dust heated by the star-formation activity. Figure 1B shows the integrated [C II] intensity map, which traces the cold gas distribution. The [C II] emission is about twice as extended as the dust emission and shows an asymmetry to the northeast. Similar lopsided gas disks are common in the nearby Universe, forming nearly half of the local population of atomic gas disks (*13*).

The [C II] velocity field (moment one map; Fig. 1C) shows a regularly rotating disk. Rotating disks and galaxy mergers can be difficult to distinguish when observed at low spatial resolutions (*14*). This is not the case for ALESS 073.1 because the area with detectable [C II] emission is covered by ~45 resolution elements. The kinematic major axis is perpendicular to the kinematic minor axis, implying that non-circular motions are negligible in the inner parts of the gas disk (*15*). The gas kinematics are less regular in the northeastern extension: This may be due to recently accreted gas, a warped outer disk, or streaming motions along a spiral arm (*9*). The limited signal-to-noise ratio in the outer regions does not allow us to distinguish between these possibilities. Overall, the kinematic regularity of the [C II] disk is unexpected for a starburst AGN-host galaxy at $z \simeq 0$.

We model the [C II] kinematics using the software ^{3D}BAROLO (*16*), which fits the observational data with a rotating disk model (*9*). This determines the [C II] surface brightness profile, the rotation curve, and the intrinsic velocity dispersion profile. The best-fitting rotation curve (Fig. 2A) rises steeply in the central parts and declines across the disk. Similar rotation curves are observed in bulge-dominated disk galaxies in the nearby Universe (*17*–*19*), so there might be a bulge component in ALESS 073.1. The normalization of the rotation curve depends on the disk inclination, which we treat as a free parameter below.

We use the derived rotation curve to constrain the mass distribution within the galaxy. Our fiducial mass model has four components: cold gas disk, stellar disk, stellar bulge, and DM halo. We also investigate mass models with three components, which we find either cannot fit the observations or are less plausible in the ΛCDM paradigm (*9*). We assume that the [C II] density profile traces the distribution of the cold gas disk, whereas the dust density profile traces the distribution of the stellar disk (Fig. 2B), because dust absorbs ultraviolet radiation from young stars and reemits it at

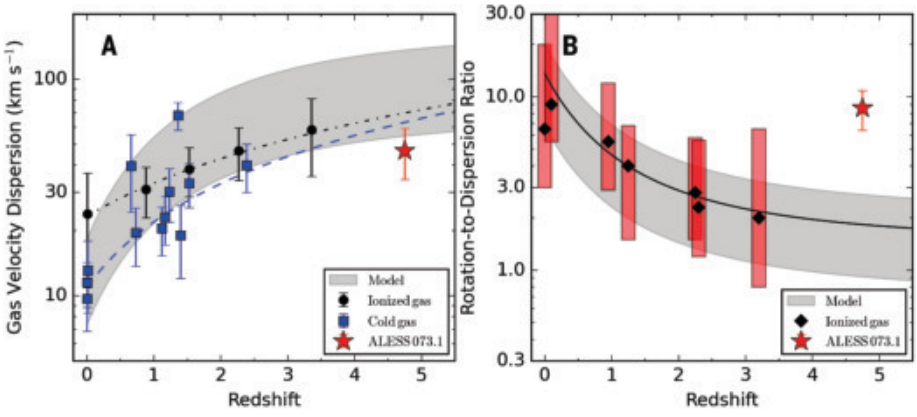


Fig. 3. Turbulence and rotational support of galaxies as a function of cosmic time. (**A**) Gas velocity dispersion and (**B**) V_{rot}/σ_V versus redshift. Gray bands show a semi-empirical model based on Toomre's disk instability parameter (*21*). In (A), the black dot-dashed line and the blue dashed line show the velocity dispersion evolution inferred from warm ionized gas data (black circles; averages from galaxy samples) and cold neutral gas data (blue squares; averages for $z = 0$ and individual galaxies for $z > 0.5$), respectively (*24*). In (B), black diamonds and red bars show, respectively, the median and 90% distribution of ionized gas surveys (*21*). The red star represents ALESS 073.1. Error bars correspond to ±1σ uncertainties.

far-infrared wavelengths. The two additional components (bulge and DM halo) are modeled with analytic functions (*9*). The bulge component implicitly includes the contribution of the central supermassive black hole, which contributes <10% of the central mass in high-z galaxies (*8*).

For the best-fitting inclination angle of 22°, the [C II] disk rotates at velocities (V_{rot}) of 300 to 400 km s^{-1} (Fig. 2A). Similar rotational speeds are observed in the most massive early-type galaxies at $z = 0$ that host molecular gas disks (*20*), so we conclude that they are the likely descendants of galaxies similar to ALESS 073.1. The intrinsic velocity dispersion (σ_V) reaches 50 to 60 km s^{-1} at small radii and decreases to ~10 km s^{-1} in the outer parts (Fig. 2C). The ratio of rotation velocity to velocity dispersion (Fig. 2D) is ~10 within 2.5 kpc, similar to that of cold gas disks at $z = 0$ (*21*). The [C II] disk of ALESS 073.1 is supported by rotation, not turbulence.

The bulge component of the model is necessary to reproduce the observed high rotation speeds at small radii. We obtain a bulge–to–total mass ratio $M_{bul}/M_{baryon} = 0.44$, where M_{baryon} is given by the sum of all baryonic (non-DM) components within 3.5 kpc. This ratio is nearly independent of the disk inclination because it is largely driven by the shapes of the observed rotation curve and of the baryonic gravitational contributions, not by their absolute normalizations. Baryons dominate the gravitational potential within 3.5 kpc, similar to massive galaxies at $z = 0$ (*19*) and $z = 1$ to 3 (*22*, *23*). In ΛCDM cosmology, DM halos are expected to have low concentrations at high z, whereas baryons can efficiently cool and collapse to the bottom of the potential well, reaching high concentrations. Thus, DM halos become gravitationally dominant at large radii.

Gas disks in the early Universe are expected to be more turbulent than their local analogs: the gas velocity dispersion is thought to increase with redshift, whereas the degree of rotational support (V_{rot}/σ_V) decreases (*21*, *24*). The enhanced turbulence may be driven by gravitational instabilities from cold gas flows, and/or by stellar and AGN feedback (*25*). At $z \geq 1$, however, the limited spatial resolution can bias the measurements of V_{rot} and σ_V (beam-smearing effects), leading to systematic overestimates of σ_V and underestimates of V_{rot}/σ_V (*26*). The gas velocity dispersion of ALESS 073.1 is consistent with the extrapolations from data at $z < 3.5$ (Fig. 3A) but lies at the low σ_V boundary of the prediction of a semi-empirical model based on Toomre's disk instability parameter (*21*), despite the galaxy hosting a starburst and a dust-obscured AGN. The rotational support of the gas disk of ALESS 073.1 is much higher than the value predicted by the same model (Fig. 3B), which has been used to describe kinematic measurements at $z < 3.5$ (*21*). The V_{rot}/σ_V ratio is similar to that of gas disks in the local Universe (Fig. 3B). This suggests that both starburst and AGN feedback have only a gentle effect on the cold interstellar medium of ALESS 073.1 because the [C II] disk is regular and unperturbed, contrary to galaxy formation models with violent feedback.

It remains unknown whether bulges exist in the early Universe ($z > 4$), given the lack of high-resolution infrared observations probing the stellar mass distribution. In ALESS 073.1, we use the gas kinematics to infer the presence of a central massive component that is not traced by the dust. This is likely a stellar bulge hosting a supermassive black hole. Bulges form either from the secular evolution of stellar disks or after major galaxy mergers (*4*). Both mechanisms could be at play in ALESS 073.1, but they must act on short time scales because the Universe was only 1.2 billion years old at $z \simeq 4.75$. Any major merger must have happened at even earlier cosmic epochs to provide time for the gas kinematics to relax to regular rotation. The orbital time at the last measured point of the rotation curve is about 5×10^7 years, and about five revolutions may be required to relax the disk, so any major merger must have happened at $z > 5.5$. ALMA observations of a quasar-host galaxy at $z \simeq 6.6$ with similar spatial resolution show that the [C II] kinematics can be heavily disturbed at this younger epoch (~0.8 billion years after the Big Bang), possibly owing to mergers or interactions (*27*). Bulges may also develop from star formation inside AGN-driven gas outflows (*28*, *29*), which could occur on short time scales. Although we observe only a single object, we conclude that the Universe produced regularly rotating galaxy disks with prominent bulges at <10% of its current age. This implies that the formation of massive galaxies and their central bulges must be a fast and efficient process.

REFERENCES AND NOTES

1. S. D. M. White, M. J. Rees, *Mon. Not. R. Astron. Soc.* **183**, 341–358 (1978).
2. A. Dekel, Y. Birnboim, *Mon. Not. R. Astron. Soc.* **368**, 2–20 (2006).
3. M. Vogelsberger *et al.*, *Nature* **509**, 177–182 (2014).
4. J. Kormendy, R. C. Kennicutt Jr., *Annu. Rev. Astron. Astrophys.* **42**, 603–683 (2004).
5. E. Bañados *et al.*, *Astrophys. J.* **881**, L23 (2019).
6. C. De Breuck *et al.*, *Astron. Astrophys.* **565**, A59 (2014).
7. G. C. Jones *et al.*, *Astrophys. J.* **850**, 180 (2017).
8. A. Pensabene *et al.*, *Astron. Astrophys.* **637**, A84 (2020).
9. Materials and methods are available as supplementary materials.
10. K. E. K. Coppin *et al.*, *Mon. Not. R. Astron. Soc.* **395**, 1905–1914 (2009).
11. R. Gilli *et al.*, *Astron. Astrophys.* **562**, A67 (2014).
12. R. Gilli *et al.*, *Astrophys. J.* **730**, L28 (2011).
13. R. Sancisi, F. Fraternali, T. Oosterloo, T. van der Hulst, *Astron. Astrophys. Rev.* **15**, 189–223 (2008).
14. S. M. Sweet *et al.*, *Mon. Not. R. Astron. Soc.* **485**, 5700–5714 (2019).
15. F. Fraternali, T. Oosterloo, R. Sancisi, G. van Moorsel, *Astrophys. J.* **562**, L47–L50 (2001).
16. E. M. Di Teodoro, F. Fraternali, *Mon. Not. R. Astron. Soc.* **451**, 3021–3033 (2015).
17. S. Casertano, J. H. van Gorkom, *Astron. J.* **101**, 1231 (1991).
18. E. Noordermeer, J. M. van der Hulst, R. Sancisi, R. S. Swaters, T. S. van Albada, *Mon. Not. R. Astron. Soc.* **376**, 1513–1546 (2007).
19. F. Lelli, S. S. McGaugh, J. M. Schombert, *Astron. J.* **152**, 157 (2016).
20. T. A. Davis *et al.*, *Mon. Not. R. Astron. Soc.* **455**, 214–226 (2016).
21. E. Wisnioski *et al.*, *Astrophys. J.* **799**, 209 (2015).
22. F. Lelli *et al.*, *Mon. Not. R. Astron. Soc.* **479**, 5440–5447 (2018).
23. R. Genzel *et al.*, *Nature* **543**, 397–401 (2017).
24. H. Übler *et al.*, *Astrophys. J.* **880**, 48 (2019).
25. M. D. Lehnert *et al.*, *Astrophys. J.* **699**, 1660–1678 (2009).
26. E. M. Di Teodoro, F. Fraternali, S. H. Miller, *Astron. Astrophys.* **594**, A77 (2016).
27. B. P. Venemans *et al.*, *Astrophys. J.* **874**, L30 (2019).

28. W. Ishibashi, A. C. Fabian, *Mon. Not. R. Astron. Soc.* **441**, 1474–1478 (2014).
29. R. Maiolino *et al*., *Nature* **544**, 202–206 (2017).
30. G. de Vaucouleurs, *Ann. d'Astrophys.* **11**, 247 (1948).

ACKNOWLEDGMENTS

We thank L. Legrand and H. Übler for providing tabular data from their publications. F.L. thanks P. Li for technical support with the Markov Chain Monte Carlo fitting code. ALMA is a partnership of ESO (representing its member states), NSF (USA), and NINS (Japan), together with NRC (Canada), MOST and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO, and NAOJ. **Funding:** F.L. was supported by an ESO fellowship during the initial stages of this project. E.M.D.T. was supported by the U.S. National Science Foundation under grant 1616177. F.F. acknowledges support from the Friedrich Wilhelm Bessel Research Award Programme of the Alexander von Humboldt Foundation. A.W.S.M. is supported by a Dunlap Fellowship at the Dunlap Institute for Astronomy & Astrophysics, funded through an endowment established by the David Dunlap family and the University of Toronto. R.M. acknowledges ERC Advanced Grant 695671 "QUENCH" and support from the Science and Technology Facilities Council (STFC). **Author contributions:** F.L.: conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, validation, visualization, writing – original draft. E.M.D.T.: conceptualization, formal analysis, methodology, software, visualization, writing – review & editing. F.F.: conceptualization, methodology, software, writing – review & editing. A.W.S.M.: formal analysis, writing – review & editing. Z.-Y.Z.: formal analysis, writing – review & editing. C.D.B.: data curation, writing – review & editing. T.A.D.: resources, writing – review & editing. R.M.: resources, writing – review & editing. **Competing interests:** We have no competing interests. **Data and materials availability:** The ALMA data are available at https://almascience.eso.org/asax/ under project codes ADS/JAO. ALMA 2017.1.01471.S, 2015.1.00456.S, and 2011.0.00124.S. The Hubble Space Telescope data are available at https://archive.stsci.edu/hst/ (proposal ID 12061, dataset ibeuglqlq). Supplementary data files contain the observed [C II] data cube (data S1), the [C II] moment maps (data S2 and S3), the model outputs from 3DBAROLO (data S4 to S6), the model outputs from GALFIT (data S7), and the best-fitting mass model (data S8).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/371/6530/713/suppl/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 and S2
References (*31*–*72*)
Data S1 to S8

10 April 2020; accepted 17 December 2020
10.1126/science.abc1893

SUPERCONDUCTIVITY

Two-dimensional superconductivity and anisotropic transport at $KTaO_3$ (111) interfaces

Changjiang Liu[1]*†, Xi Yan[1,2,3]†, Dafei Jin[4], Yang Ma[5], Haw-Wen Hsiao[6], Yulin Lin[4], Terence M. Bretz-Sullivan[1], Xianjing Zhou[4], John Pearson[1], Brandon Fisher[4], J. Samuel Jiang[1], Wei Han[5], Jian-Min Zuo[6], Jianguo Wen[4], Dillon D. Fong[1], Jirong Sun[3,7], Hua Zhou[2], Anand Bhattacharya[1]*

The distinctive electronic structure found at interfaces between materials can allow unconventional quantum states to emerge. Here we report on the discovery of superconductivity in electron gases formed at interfaces between (111)-oriented $KTaO_3$ and insulating overlayers of either EuO or $LaAlO_3$. The superconducting transition temperature, as high as 2.2 kelvin, is about one order of magnitude higher than that of the $LaAlO_3/SrTiO_3$ system. Notably, similar electron gases at $KTaO_3$ (001) interfaces remain normal down to 25 millikelvin. The critical field and current-voltage measurements indicate that the superconductivity is two-dimensional. In EuO/$KTaO_3$ (111) samples, a spontaneous in-plane transport anisotropy is observed before the onset of superconductivity, suggesting the emergence of a distinct "stripe"-like phase, which is also revealed near the critical field.

Superconductivity in two dimensions (2D) has been a central theme in condensed-matter physics and material science (*1*). In 2D, the electron-electron and electron-lattice interactions that mediate pairing can also give rise to states that compete with superconductivity, and fluctuations associated with reduced dimensionality can further suppress superconductivity. Thus, although there are many examples of two-dimensional electron gases (2DEGs) and ultrathin metallic films, only a small fraction of these are superconducting. Most of the foundational work in 2D superconductivity was carried out with amorphous thin films (*2*) and led to deep insights regarding the nature of classical and quantum phase transitions in the 2D limit. In more recent years, there have been marked developments in the realization of 2D superconductivity in crystalline materials (*1*), including layered transition-metal chalcogenides (*3*, *4*), bilayer (*5*) and trilayer (*6*) graphene, gated surfaces of single crystals (*7*, *8*), atomically thin ordered metallic layers on semiconductor surfaces (*9*), and interfaces between crystalline materials (*10*–*12*). These crystalline 2D superconductors allow us to realize and break symmetries, and tailor electronic structure in ways that are not possible in amorphous and disordered thin films. For example, in a 2D superconductor with strong spin-orbit coupling and broken inversion symmetry, a Rashba interaction may lead to a superconducting order parameter with mixed *s*-wave and *p*-wave symmetry (*13*), a candidate platform for realizing Majorana modes (*14*). It is perhaps telling that three of the most prominent examples of 2D superconductivity at crystalline interfaces—$LaAlO_3/SrTiO_3$ (LAO/STO), FeSe/$SrTiO_3$, and $La_2CuO_4/LaSrCuO_4$—involve transition-metal oxides (*10*–*12*), where strong electron-electron and electron-lattice interactions likely play a role in mediating superconducting pairing. In some of these interfacial systems, the superconductivity is believed to be "unconventional"—i.e., with a non–*s*-wave order parameter, and a pairing mechanism that is not described by the Bardeen-Cooper-Schrieffer (BCS) theory.

$KTaO_3$ (KTO) is an insulator with a cubic perovskite structure, a bandgap of 3.6 eV, and a dielectric constant that grows to >4500 upon cooling to low temperatures (*15*, *16*). It is believed that KTO is on the verge of a ferroelectric transition, thwarted by quantum fluctuations at low temperatures. It is thus called a "quantum paraelectric," much like $SrTiO_3$ (STO). STO can be easily doped into a metallic state that becomes superconducting at $T < 0.5$ K (*17*). By contrast, few studies on doped KTO exist for bulk samples (*18*, *19*), and none of them report any signatures of superconductivity. Nonetheless, it was reported that ionic liquid gating could be used to tune the KTO (001) surface into a weak superconducting state with critical temperature $T_c \sim 47$ mK (*8*). The electronic states near the Fermi level in electron-doped KTO are derived from Ta-5d states—in particular d_{xy}, d_{zx}, and d_{yz} orbitals (*20*). The large spin-orbit coupling in KTO lifts the degeneracy of these states and splits them into $J = 1/2$ (where J is the total angular momentum) and $J = 3/2$ states, with the former being ~0.4 eV higher in energy (compared to ~0.025 eV in STO). The electrons in KTO occupy the lower-energy $J = 3/2$ states, which are degenerate near the Γ-point and disperse as "light" and "heavy" electron states at finite momenta. Unlike in STO, there are no symmetry-lowering lattice distortions at low temperatures to lift this degeneracy at the Γ-point, though confinement can also lift degeneracies. 2DEGs can be

[1]Materials Science Division, Argonne National Laboratory, Lemont, IL 60439, USA. [2]Advanced Photon Source, Argonne National Laboratory, Lemont, IL 60439, USA. [3]Beijing National Laboratory for Condensed Matter and Institute of Physics, Chinese Academy of Sciences, Beijing 100190, People's Republic of China. [4]Center for Nanoscale Materials, Argonne National Laboratory, Lemont, IL 60439, USA. [5]International Center for Quantum Materials, School of Physics, Peking University, Beijing 100871, People's Republic of China. [6]Department of Materials Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [7]University of Jinan, Spintronics Institute, Jinan 250022, Shandong, People's Republic of China.
***Corresponding author. Email: anand@anl.gov (A.B.); changjiang.liu@anl.gov (C.L.)** †These authors contributed equally to this work.

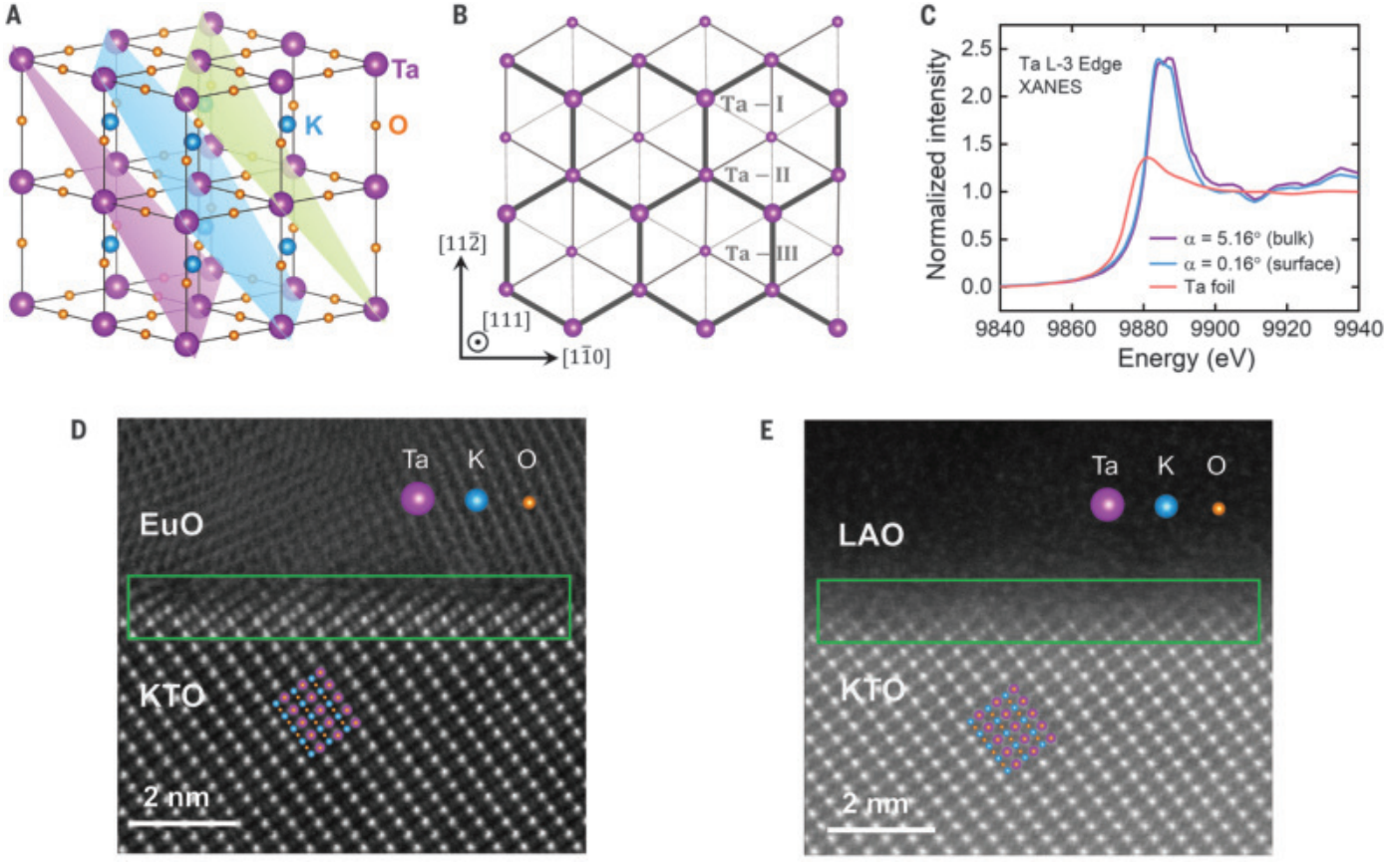


Fig. 1. Schematics of KTO (111) surface, XANES, and STEM characterizations. (**A**) $KTaO_3$ lattice structure. The relative sizes of the ions are chosen to emphasize the Ta atoms. The three adjacent (111) planes containing Ta^{5+} ions are colored in light purple, blue, and green, respectively. (**B**) Distribution of Ta^{5+} ions viewed along the [111] crystal axis. Ta^{5+} ions are shown with progressively smaller sizes in the three adjacent (111) planes, which are labeled as Ta – I, Ta – II, and Ta – III, respectively. Solid lines between Ta^{5+} ions indicate the relative distance or coupling strength—with thicker lines representing stronger couplings, giving rise to a buckled honeycomb lattice for the first bilayer comprising Ta – I and Ta – II sites. (**C**) XANES data from sample EuO/KTO(111)_4 at the Ta *L*-edge. The KTO near the surface is only slightly reduced, with Ta valence close to its bulk value (5+). XANES data for pure Ta are also shown for comparison. STEM images of the (**D**) EuO/KTO (111) and (**E**) LAO/KTO (111) interfaces, looking down the [1$\bar{1}$0] direction. Green box indicates the region near the interface.

realized at the KTO (001) and (111) interfaces through vacuum cleaving (*21*) followed by exposure to ultraviolet or synchrotron radiation. Angle-resolved photoemission spectroscopy studies on the KTO (111) surface found a distinct Fermi surface with a sixfold symmetry derived from the lattice structure (Fig. 1, A and B), spin split-off bands, and evidence for band bending and confinement near the surface (*21*, *22*). Owing to the large spin-orbit coupling (*23*, *24*), the interfacial electron gas at KTO (111) interfaces is also expected to host complex spin textures in momentum space (*22*). In the particular case of a bilayer of KTO along the [111] direction, which forms a buckled honeycomb lattice (Fig. 1B), calculations suggest that this system may support topological states (*25*, *26*).

Here, we report on the observation of 2D superconductivity in electron gases formed at oxide-insulator/KTO (111) interfaces. We have measured T_c as high as 2.2 K, and this value can be tuned by varying carrier density during the sample growth. We also observe an emergent stripe phase, which breaks the rotational symmetry in the KTO (111) surface.

The 2DEGs on KTO studied in this work were prepared by growing an overlayer of EuO with molecular beam epitaxy, or LAO using pulsed laser deposition (*27*). We characterized the chemical oxidation state of the EuO/KTO (111) interface using x-ray absorption near-edge spectroscopy (XANES) at the Ta-L_3 edge (9.88 keV). XANES measurements were carried out by adopting two different angles of incidence, one close to grazing incidence (0.16°) that probed only the first few nanometers of the KTO near the EuO/KTO interface, and another at a higher angle (5.16°) that probed the bulk of the KTO. As shown in Fig. 1C, these measurements indicate that the Ta oxidation state near the surface is very close to that in the bulk (Ta^{5+}), and a subtle difference in the absorption curves suggests that the interfacial Ta is only slightly reduced. In particular, none of our measurements show any evidence for metallic Ta, which is known to be superconducting in ultrathin films (*28*). Shown in Fig. 1, D and E, are scanning transmission electron microscopy (STEM) measurements of our samples, which indicate abrupt interfaces between an annealed KTO (111) surface and the polycrystalline EuO or amorphous LAO overlayer, as confirmed by x-ray diffraction measurements. Using aberration-corrected high-resolution TEM and scanning TEM, we find oxygen vacancies near the EuO/KTO (111) interface, and Eu substitution on the K sites within the first three unit cells (fig. S1), both of which can dope the interfacial KTO with electrons. Similar doping effects are also observed at the LAO/KTO (111), EuO/KTO (001), and LAO/KTO (001) interfaces (figs. S2 and S3).

Figure 2A shows transport measurements on 2DEGs at EuO/KTO (111) and (001) interfaces, both of which display metallic behavior from 300 to 4 K. As temperature is lowered further, the EuO/KTO (111) interface shows superconductivity, which is presented in Fig.

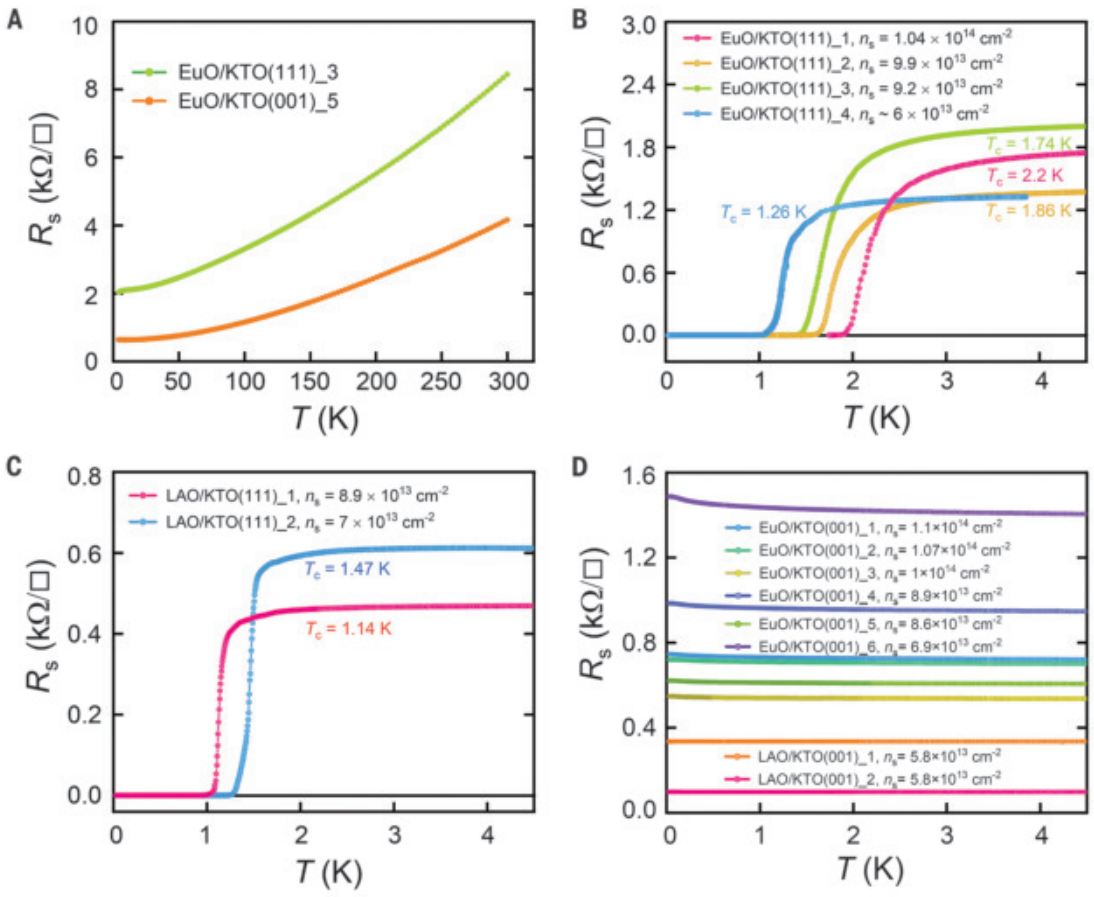


Fig. 2. Transport measurements on 2DEGs formed at different KTO interfaces. (**A**) Metallic temperature dependence of the sheet resistance of EuO/KTO (111) and (001) samples measured from 300 to 4 K. (**B**) Measurement at lower temperatures shows superconducting transitions in EuO/KTO (111) samples (current along [11$\bar{2}$]) with varying carrier densities, which are determined from Hall measurement at T = 10 K for samples EuO/KTO(111)_1, 2, and 3. The carrier density in EuO/KTO(111)_4 is estimated from growth conditions. (**C**) Similar measurements on LAO/KTO (111) samples also show superconductivity. (**D**) No superconductivity is observed in samples with (001)-oriented KTO interfaces with overlayers of either EuO or LAO down to 25 mK. The range of the carrier density is similar to those of the (111)-oriented samples shown in (B) and (C).

2B with four samples. These samples were grown at different temperatures and oxygen pressures, resulting in different carrier densities and mobilities (*27*). The T_c, determined at half of the normal state resistance R_N, reaches as high as 2.2 K in sample EuO/KTO (111)_1. This transition temperature is about one order of magnitude higher than those (T_c ~ 200 mK) found at the LAO/STO interfaces. To investigate whether the EuO overlayer is necessary for superconductivity, we prepared 2DEGs using LAO overlayers. As shown in Fig. 1C, superconductivity with T_c > 1.1 K is also observed in LAO/KTO (111) samples, demonstrating that the emergence of superconductivity does not depend on a specific overlayer of the 2DEG.

We also performed low-temperature measurements on 2DEGs formed at KTO (001) interfaces. Figure 2D shows the measurement results for a series of (001) samples including both EuO and LAO overlayers. The sheet carrier density n_s of these samples varies from 5.8 × 10^{13} to 1.1 × 10^{14} cm^{-2}, covering the same range of densities as the (111) samples shown in Fig. 2, B and C, but slightly lower than the n_s values in (*8*). Unlike the (111) samples, no superconductivity is observed in these (001) samples down to 25 mK. This crystallographic orientation–dependent interfacial superconductivity is in sharp contrast with that of 2DEGs at STO interfaces, where superconductivity, with about the same T_c, occurs for all orientations (*10*, *29*, *30*). Our STEM characterizations of the KTO (001) interfaces show an interfacial chemical composition similar to that of the (111) samples. This leads to the possibility that the distinctive electronic structure in the KTO (111) surface may play a role in inducing strong Cooper pairing among the interfacial charge carriers, which needs further investigation.

The superconductivity at the KTO (111) interface is two-dimensional. This is borne out by upper critical fields $B_{c\perp}$ and $B_{c||}$ measured perpendicular and parallel to the KTO (111) plane, respectively, and defined at the half of R_N. Figure 3, A and B, show the magnetic field dependence of R_S measured on sample EuO/KTO(111)_3. The temperature dependence of $B_{c\perp}$ is well described by the Ginzburg-Landau (G-L) theory, which yields a linearized equation: $B_{c\perp}(T) = \Phi_0(1 - T/T_c)/[2\pi(\xi_{GL})^2]$, where Φ_0 is the magnetic flux quantum and ξ_{GL} is the G-L coherence length at T = 0 K. The linear fit to $B_{c\perp}(T)$ in Fig. 3C gives ξ_{GL} ≈ 13 nm. The extrapolated value of $B_{c\perp}(0)$ ≈ 1.8 T is > 10^3 times higher than the critical field observed in ionic-liquid gated KTO (001) devices (*8*). Further, we find that the critical fields are highly anisotropic. For fields applied in the sample plane, $B_{c||}$ is much higher than $B_{c\perp}$, with the ratio $B_{c||}/B_{c\perp}$ > 10 at T = 0.6 K (Fig. 3C). The temperature dependence of $B_{c||}$ shows a characteristic square-root dependence, as expected from Tinkham's model (*31*): $B_{c||}(T) = \Phi_0[12(1 - T/T_c)]^{1/2}/(2\pi d\xi_{GL})$, where d is the effective thickness of the superconducting state. From the fit to $B_{c||}(T)$, we obtain d ≈ 5.1 nm < ξ_{GL}, which is consistent with 2D superconductivity.

We note that $B_{c||}(0)$ ≈ 17.0 T extrapolated from the fit is substantially larger than the paramagnetic pair-breaking field $B_P = \Delta_0/(2^{1/2}\mu_B)$ ≈ 3.2 T based on BCS theory in the weak-coupling limit (*32*, *33*), where Δ_0 ≈ 1.76k_BT_c is the zero-temperature superconducting gap, k_B is the Boltzmann constant, and μ_B is the Bohr magneton. High values of $B_{c||}$ in excess of B_P can be realized in the presence of spin-orbit coupling owing to elastic scattering (*33*, *34*) in the limit where the mean free path l_{mfp} << ξ_{GL}, though other spin-relaxation mechanisms may apply even in the clean limit in the presence of broken inversion symmetry and high spin-orbit coupling (*35*). Using a single-band model, where $l_{mfp} = \hbar/(e^2k_FR_S)$, $\hbar$ is

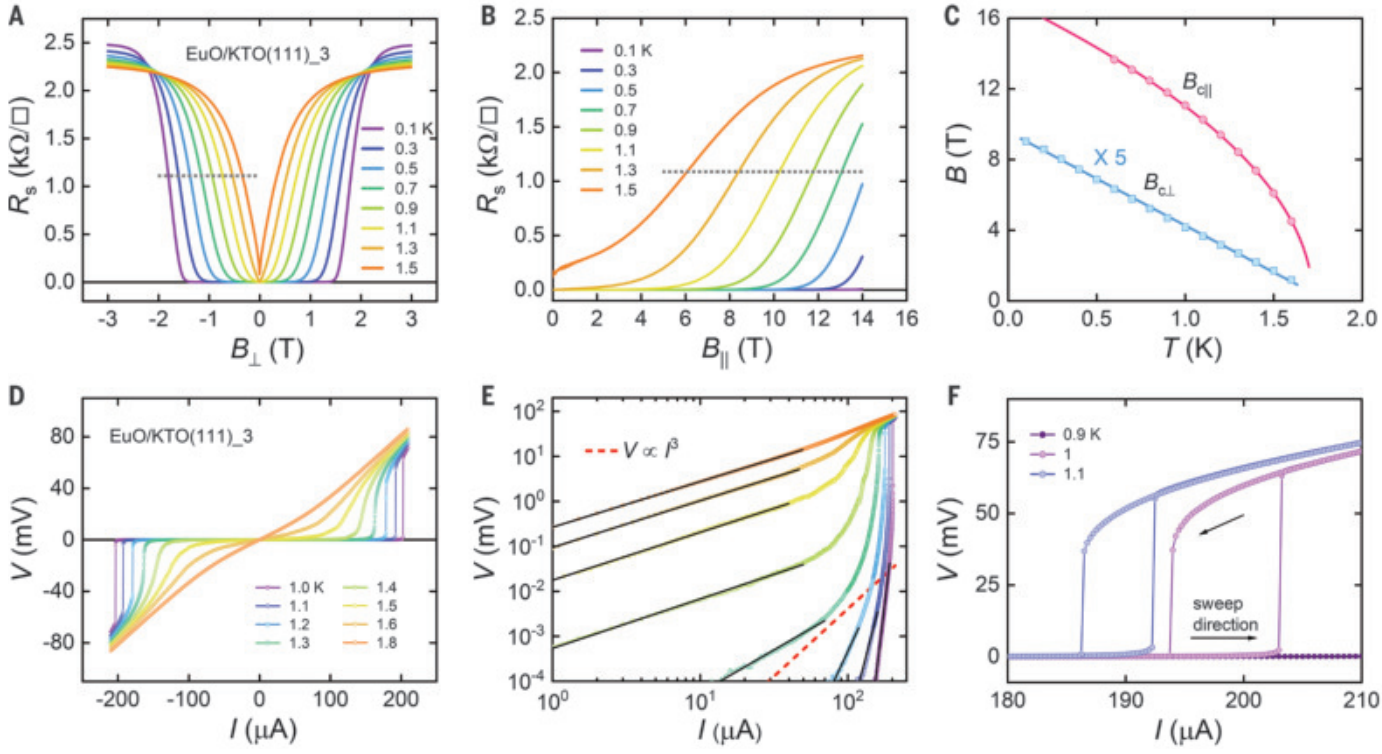


Fig. 3. Critical field and current-voltage measurements on sample EuO/KTO(111)_3. (**A** and **B**) Sheet resistance measured at different temperatures as a function of the out-of-plane and in-plane magnetic fields, respectively. (**C**) Temperature dependence of critical fields, which are determined at half of R_N [dotted horizontal line in (A) and (B)]. (**D**) *I-V* curves measured at different temperatures. (**E**) *I-V* curves plotted on a logarithmic scale using the same color codes as in (D). Black solid lines are linear fits to the data. Red dashed line is $V \propto I^3$, which is used to infer the BKT transition temperature. (**F**) Hysteresis of *I-V* curves near the critical current measured at temperatures below T_{c0}. In all the measurements (A) to (F), the current is applied along the [11$\bar{2}$] direction.

the Planck constant, e is the electron charge, and $k_F = (2\pi n_s)^{1/2}$ is the Fermi wave vector, our estimates for different samples suggest l_{mfp} ~ 0.4 to 0.7 times ξ_{GL}, which is intermediate between the clean and dirty limits. Enhanced $B_{c\parallel}$ is also expected in superconductors that are believed to have a spin-triplet p-wave order parameter, though these were thought to be unstable in the presence of disorder. However, recent theoretical developments indicate that in the presence of strong spin-orbit coupling, criteria that might apply to KTO (111) interfaces such as broken inversion symmetry (*36*), spin-momentum locking (*37*), or multi-orbital superconducting states (*38*) can stabilize unconventional superconductivity against disorder scattering. Thus, at this time, the mechanism for realizing $B_{c\parallel}$ values in excess of B_P at KTO (111) interfaces remains an open question. In Fig. 3A, we also observe an approximate crossing point near $B_\perp$ = 2.1 T, which implies the existence of a nearly temperature-independent separatrix in R_s versus T traces that show either a positive or negative temperature coefficient of resistance upon cooling, depending on magnetic field (fig. S4). Such data may be interpreted in terms of a field-induced quantum phase transition between a 2D superconductor and a correlated metal (*39*).

The superconducting state in the EuO/KTO sample shows a robust critical-current behavior. The measured *I-V* characteristics at different temperatures are shown in Fig. 3D, where the critical current I_c approaches 200 μA at T = 1 K. The I_c exceeds 200 μA at lower temperatures (Fig. 3F). A lower bound of the critical current density, K_c(1 K) = 500 μA cm^{-1}, is obtained by assuming the current spreads over the entire width of the sample (4 mm). A more precise determination of K_c requires measurements using patterned Hall bar devices. As the temperature is raised to near T_c, we observe a gradual onset of a resistive state at low currents. In Fig. 3E, we plot the *I-V* on a log-log scale and observe that the slope of the *I-V* characteristics evolves smoothly from the normal ohmic state, $V \propto I$, toward a steeper power law, $V \propto I^\alpha$, as superconductivity sets in at lower temperatures. This evolution may be interpreted in terms of a Berezinskii–Kosterlitz–Thouless (BKT) transition in a 2D superconductor (*40*), where the onset of a nonlinear *I-V* characteristic in the superconducting state is caused by the current-driven unbinding of vortex-antivortex pairs that are created by thermal fluctuations at finite temperatures. This causes the exponent α to smoothly evolve from a value of 1 above and near the onset of superconductivity to higher values as the temperature is lowered (*41*, *42*), with the α = 3 at T_{BKT}. The red dashed line in Fig. 3E corresponds to an I^3 dependence, which indicates that the T_{BKT} of this sample EuO/KTO(111)_3 is between 1.2 and 1.3 K, which is slightly below the zero-resistance T_{c0} of about 1.34 K. At temperatures below T_{c0}, we observe hysteresis in the *I-V* curves near I_c (Fig. 3F), which suggests that the 2D superconductivity is inhomogeneous, with superconducting regions joined by weak links (*43*).

In low–carrier density EuO/KTO (111) samples, we observe the emergence of a distinct phase near the superconducting state, as revealed by a spontaneous in-plane anisotropy in resistance. Shown in Fig. 4A is the measurement in sample EuO/KTO(111)_5, with n_s = 3.5 × 10^{13} cm^{-2} and a relatively high mobility of about 279 cm^2 V^{-1} s^{-1} at T = 10 K. The measurements are carried out in a van der Pauw geometry (Fig. 4B). As temperature is lowered below about 2.2 K (Fig. 4A), the resistance increases by almost 50% for current along the [1$\bar{1}$0] crystal axis, whereas it decreases by over 50% for current flowing in the [11$\bar{2}$] direction. We note that although the van der Pauw method can amplify the transport anisotropy, the qualitative feature

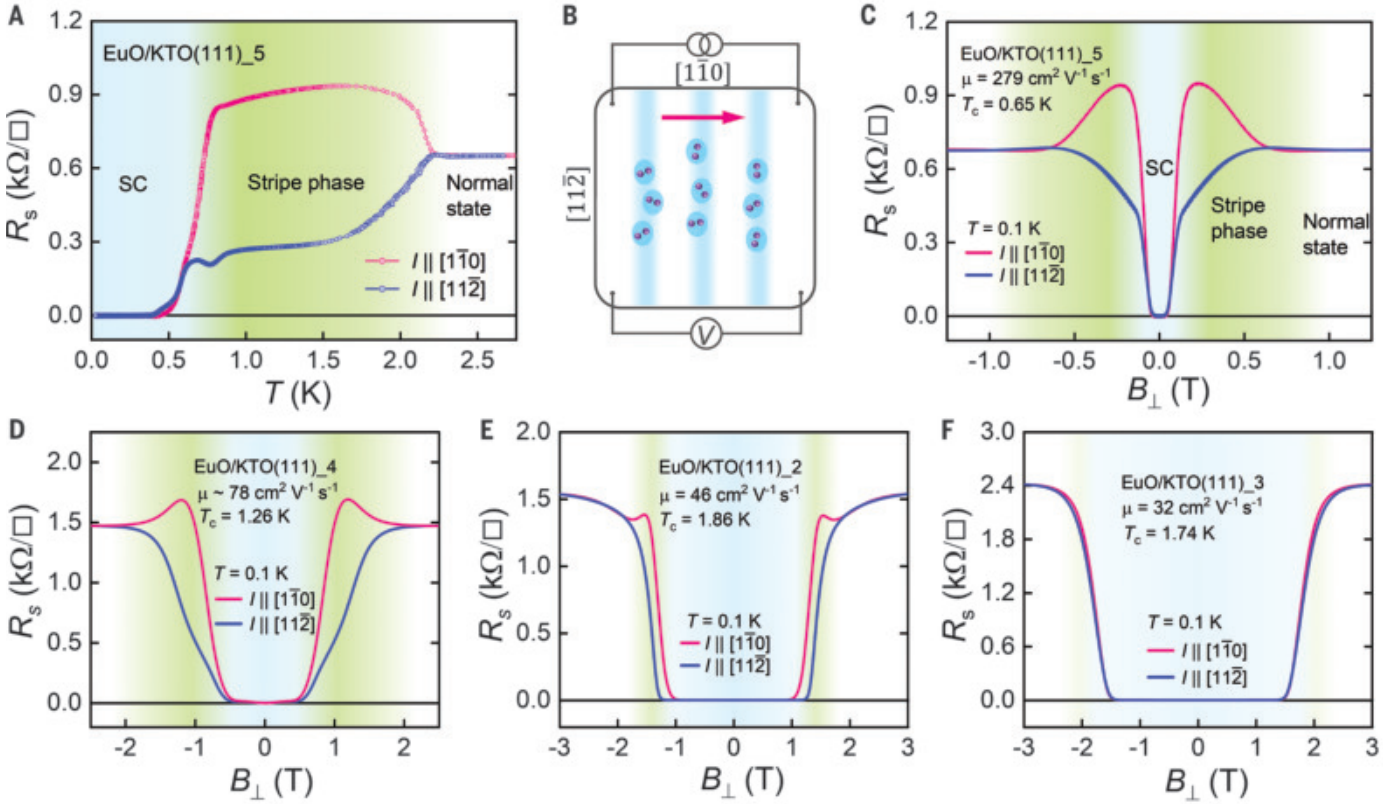


Fig. 4. Stripe phase measured in different EuO/KTO(111) samples. (**A**) Sheet resistance of sample EuO/KTO(111)_5 measured with electric current along [1$\bar{1}$0] (red) and [11$\bar{2}$] (blue) crystal axes under zero field. The light blue and green regions indicate superconducting (SC) and "stripe" state, respectively. (**B**) Illustration of the measurement geometry for the case of current (red arrow) along the [1$\bar{1}$0] direction perpendicular to the stripes. These stripes may be composed of Cooper pairs, which are shown in light blue bubbles. (**C** to **F**) Magnetic field dependence of the sheet resistance measured along both current directions at T = 0.1 K in samples with decreasing mobilities. The stripe phase is revealed in all samples around the critical field (green region). Note that EuO/KTO(111)_2 has a higher T_c than EuO/KTO(111)_3 but also shows a more prominent transport anisotropy.

reported here should be unaffected (*44*). A similar approach has been used in the past to characterize resistance anisotropy in high-mobility 2DEGs in the quantum Hall regime (*45*). The formation of this anisotropy in our samples suggests the emergence of a distinct phase (*46*) that breaks rotational symmetry over macroscopic length scales (fig. S5). The observed anisotropy persists over a broad temperature range from 2.2 K down to about 0.7 K, which is shown with a green background in Fig. 4A. At still lower temperatures, the resistance in both the [1$\bar{1}$0] and [11$\bar{2}$] current directions drops rapidly to zero, and the superconducting state is obtained (light blue region).

The increase in resistance along the [1$\bar{1}$0] direction (Fig. 4A) upon lowering temperature from the normal state is similar to observations in 2D granular superconductors (*47*), where superconducting puddles develop first. This occurs owing to the gapping out of quasiparticles by a superconducting gap (*48*), which inhibits transport between weakly coupled superconducting regions. Global superconductivity is restored at still lower temperatures presumably via Josephson coupling between these regions. The decrease in resistance along [11$\bar{2}$] at the same temperature suggests the possibility of stronger coherent coupling between the superconducting regions in that direction. This may occur if the underlying superconductivity is anisotropic, making superconducting regions organize themselves into "stripes," as illustrated in Fig. 4B. Our measurements also imply that the orientation of the stripes breaks the threefold rotational symmetry of the KTO (111) surface. A possible explanation is that these stripes are pinned by the sample edges and remain coherently aligned over macroscopic length scales. We note that preformed superconducting puddles have also been suggested in the LAO/STO system (*49*).

The magnetic field dependence of sheet resistance provides further evidence for an anisotropic stripe like phase. Figure 4C shows R_s measured in the same EuO/KTO(111)_5 sample in an out-of-plane magnetic field at T = 0.1 K, which is well below T_{c0}. As the magnetic field is increased to near the critical field, the resistance along both directions of current rises sharply. The initial high- and low-resistance values along [1$\bar{1}$0] and [11$\bar{2}$], respectively, in the resistive state are approximately equal to those in the stripe phase shown in Fig. 4A, obtained by just raising the temperature to above T_{c0}. This suggests that the stripe phase is obtained by a suppression of global superconductivity upon the application of a magnetic field. By increasing the magnetic field further, the resistance decreases along [1$\bar{1}$0] and increases along [11$\bar{2}$], as the stripe phase is driven into the normal state at $B_\perp \geq 0.8$ T, where the resistance anisotropy also disappears, presumably owing to a more complete suppression of the superconducting gap within the stripes.

Figure 4, D to F, shows magnetotransport measurements in other EuO/KTO (111) samples, with higher carrier densities and lower mobilities than that of EuO/KTO(001)_5. The resistance anisotropy, showing the same dependence on the crystal axes, is observed in all of these samples to varying degrees around the critical field. The magnitude of the transport anisotropy, normalized by the

normal-state value, becomes progressively smaller as the mobility of the sample decreases. In sample EuO/KTO(111)_3, which has the lowest mobility among these samples, the transport is nearly isotropic (Fig. 4F). This suggests that the stripe phase depends on the disorder and/or carrier density within the sample. We point out that in general, nematic or stripe-like states are known to be very sensitive to disorder in 2D (*50*)—for example, in the layered material $Sr_3Ru_2O_7$ (*51*), the stripe-like phase is only observed in very clean samples. Furthermore, the stripe phase in our samples seems to not compete with the superconductivity. For example, sample EuO/KTO(111)_2 clearly shows a stripe-like anisotropy (Fig. 4E), but it also has a higher T_c than that of EuO/KTO(111)_3 (see Fig. 2B), which shows very weak anisotropy (Fig. 4F). We note that our limited transport measurements in LAO/KTO (111) samples, where the overlayer LAO is nonmagnetic, show that the transport is mostly isotropic when an out-of-plane field is applied (fig. S6). However, a large anisotropy is observed upon the application of an in-plane magnetic field approaching $B_{c||}$. Our findings suggest that the stripe phase may be controlled by the interaction between electron spins and an effective magnetic field—either through the exchange coupling with the magnetic EuO overlayer in EuO/KTO (111) samples, or the external in-plane magnetic field, as for the case of LAO/KTO (111).

The 2D superconductivity observed in the KTO (111) interfaces shows substantially higher T_c than that in the widely studied LAO/STO system, and the T_c increases monotonically with n_s in the measured range. These higher T_c values make fundamental studies and device applications of superconducting 2DEGs using crystalline interfaces more readily accessible. The origin of the superconductivity is still an open question. Possible links between superconductivity and electronic structure including the large spin-orbit interaction in KTO (111) 2DEGs, as has been suggested in LAO/STO (111) (*52*, *53*), could be explored in future work.

When global superconductivity is suppressed by either temperature or a magnetic field, transport measurements reveal a stripe phase that produces large anisotropic transport in the (111) KTO plane. Notably, the directions for the high- and low-resistance states in the vicinity of superconductivity are oriented along the same crystal axes in the KTO and STO (111) interfaces (*54*, *55*). We observe no appreciable dependence on thermal cycling across the onset temperature of the stripe phase (fig. S7), as has been observed in LAO/STO (*56*), where these effects were caused by the formation of structural domain walls. Although our x-ray and STEM measurements carried out at 300 K do not indicate any structural distortions that would break the threefold symmetry of the KTO (111) surface, we cannot rule out weak distortions below our measurement resolution, or the emergence of such distortions at lower temperatures. In addition, the stripe phase shows strong dependence on magnetic field, mobility, and/or carrier densities, which suggests a magnetic and/or electronic origin. A possible explanation for the observed anisotropy may be a spatially modulated or inhomogeneous superconductivity. This can happen, for example, in a Fulde-Ferrel-Larkin-Ovchinnikov state, where a spin-split Fermi surface formed owing to a magnetic proximity effect or an external magnetic field resulting in Cooper pairs with nonzero total crystal momentum (*36*, *46*), whose density is periodically modulated in real space. It is also possible that the superconductivity at KTO (111) interfaces is inherently composed of a spatially modulated density of Cooper pairs or a pair-density wave (*57*). Further experiments, including probes of the spatial structure of superconductivity, will be helpful for elucidating the nature of superconductivity and transport anisotropy that we observe.

REFERENCES AND NOTES

1. Y. Saito, T. Nojima, Y. Iwasa, *Nat. Rev. Mater.* **2**, 16094 (2016).
2. A. M. Goldman, *Int. J. Mod. Phys. B* **24**, 4081–4101 (2010).
3. M. M. Ugeda *et al.*, *Nat. Phys.* **12**, 92–97 (2015).
4. X. X. Xi *et al.*, *Nat. Phys.* **12**, 139–143 (2016).
5. Y. Cao *et al.*, *Nature* **556**, 43–50 (2018).
6. G. Chen *et al.*, *Nature* **572**, 215–219 (2019).
7. J. T. Ye *et al.*, *Nat. Mater.* **9**, 125–128 (2010).
8. K. Ueno *et al.*, *Nat. Nanotechnol.* **6**, 408–412 (2011).
9. T. Zhang *et al.*, *Nat. Phys.* **6**, 104–108 (2010).
10. N. Reyren *et al.*, *Science* **317**, 1196–1199 (2007).
11. A. Gozar *et al.*, *Nature* **455**, 782–785 (2008).
12. J.-F. Ge *et al.*, *Nat. Mater.* **14**, 285–289 (2015).
13. L. P. Gor'kov, E. I. Rashba, *Phys. Rev. Lett.* **87**, 037004 (2001).
14. A. C. Potter, P. A. Lee, *Phys. Rev. B* **83**, 184520 (2011).
15. Y. Fujii, T. Sakudo, *J. Phys. Soc. Jpn.* **41**, 888–893 (1976).
16. C. Ang, A. S. Bhalla, L. E. Cross, *Phys. Rev. B* **64**, 184104 (2001).
17. J. F. Schooley *et al.*, *Phys. Rev. Lett.* **14**, 305–307 (1965).
18. S. H. Wemple, *Phys. Rev.* **137**, A1575–A1582 (1965).
19. A. Sakai, T. Kanno, S. Yotsuhashi, H. Adachi, Y. Tokura, *Jpn. J. Appl. Phys.* **48**, 097002 (2009).
20. G. E. Jellison, I. Paulauskas, L. A. Boatner, D. J. Singh, *Phys. Rev. B* **74**, 155130 (2006).
21. C. Bareille *et al.*, *Sci. Rep.* **4**, 3586 (2014).
22. F. Y. Bruno *et al.*, *Adv. Electron. Mater.* **5**, 1800860 (2019).
23. H. Zhang *et al.*, *Nano Lett.* **19**, 1605–1612 (2019).
24. H. Zhang *et al.*, *ACS Nano* **13**, 609–615 (2019).
25. M. S. Scheurer, J. Schmalian, *Nat. Commun.* **6**, 6005 (2015).
26. M. S. Scheurer, D. F. Agterberg, J. Schmalian, *npj Quantum Mater.* **2**, 9 (2017).
27. See supplementary materials.
28. Y. Qin, C. L. Vicente, J. Yoon, *Phys. Rev. B* **73**, 100505 (2006).
29. Y. L. Han *et al.*, *Appl. Phys. Lett.* **105**, 192603 (2014).
30. A. M. R. V. L. Monteiro *et al.*, *Phys. Rev. B* **96**, 020504 (2017).
31. M. Tinkham, *Phys. Rev.* **129**, 2413–2422 (1963).
32. A. M. Clogston, *Phys. Rev. Lett.* **9**, 266–267 (1962).
33. N. R. Werthamer, E. Helfand, P. C. Hohenberg, *Phys. Rev.* **147**, 295–302 (1966).
34. M. Kim, Y. Kozuka, C. Bell, Y. Hikita, H. Y. Hwang, *Phys. Rev. B* **86**, 085121 (2012).
35. D. Pesin, A. H. MacDonald, *Nat. Mater.* **11**, 409–416 (2012).
36. K. Michaeli, A. C. Potter, P. A. Lee, *Phys. Rev. Lett.* **108**, 117003 (2012).
37. K. Michaeli, L. Fu, *Phys. Rev. Lett.* **109**, 187003 (2012).
38. M. S. Scheurer, M. Hoyer, J. Schmalian, *Phys. Rev. B* **92**, 014518 (2015).
39. A. Yazdani, A. Kapitulnik, *Phys. Rev. Lett.* **74**, 3037–3040 (1995).
40. J. M. Kosterlitz, D. J. Thouless, *J. Phys. Chem.* **5**, L124–L126 (1972).
41. M. R. Beasley, J. E. Mooij, T. P. Orlando, *Phys. Rev. Lett.* **42**, 1165–1168 (1979).
42. K. Epstein, A. M. Goldman, A. M. Kadin, *Phys. Rev. Lett.* **47**, 534–537 (1981).
43. W. J. Skocpol, M. R. Beasley, M. Tinkham, *J. Appl. Phys.* **45**, 4054–4066 (1974).
44. I. Miccoli, F. Edler, H. Pfnür, C. Tegenkamp, *J. Phys. Condens. Matter* **27**, 223201 (2015).
45. M. P. Lilly, K. B. Cooper, J. P. Eisenstein, L. N. Pfeiffer, K. W. West, *Phys. Rev. Lett.* **82**, 394–397 (1999).
46. V. Barzykin, L. P. Gor'kov, *Phys. Rev. Lett.* **89**, 227002 (2002).
47. H. M. Jaeger, D. B. Haviland, B. G. Orr, A. M. Goldman, *Phys. Rev. B* **40**, 182–196 (1989).
48. V. F. Gantmakher, V. T. Dolgopolov, *Phys. Uspekhi* **53**, 1–49 (2010).
49. Z. Chen *et al.*, *Nat. Commun.* **9**, 4008 (2018).
50. E. Fradkin, S. A. Kivelson, M. J. Lawler, J. P. Eisenstein, A. P. Mackenzie, *Annu. Rev. Condens. Matter Phys.* **1**, 153–178 (2010).
51. S. A. Grigera *et al.*, *Science* **306**, 1154–1157 (2004).
52. P. K. Rout, E. Maniv, Y. Dagan, *Phys. Rev. Lett.* **119**, 237002 (2017).
53. J. Hutchinson, J. E. Hirsch, F. Marsiglio, *Phys. Rev. B* **97**, 184513 (2018).
54. S. Davis *et al.*, *Phys. Rev. B* **96**, 134502 (2017).
55. S. Davis *et al.*, *Phys. Rev. B* **98**, 024504 (2018).
56. Y. Frenkel *et al.*, *ACS Appl. Mater. Interfaces* **8**, 12514–12519 (2016).
57. D. F. Agterberg *et al.*, *Annu. Rev. Condens. Matter Phys.* **11**, 231–270 (2020).
58. C. Liu *et al.*,Data for "Two-dimensional superconductivity and anisotropic transport at $KTaO_3$ (111) interfaces," Harvard Dataverse (2020); https://doi.org/10.7910/DVN/ATP5G5.

ACKNOWLEDGMENTS

We thank I. Martin, M. Norman, and P. Littlewood for discussions. We thank C. Sun for the reference XANES spectral of Ta metal foil and M. Eblen-Zayas for discussions about EuO growth. **Funding:** All work at Argonne was supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division. The use of facilities at the Center for Nanoscale Materials and the Advanced Photon Source, both Office of Science user facilities, was supported by the U.S. Department of Energy, Basic Energy Sciences, under Contract no. DE-AC02-06CH11357. The support provided by China Scholarship Council (CSC) during a visit of X.Y. to Argonne National Laboratory is acknowledged. J.S. acknowledges the support of the National Natural Science Foundation of China (no. 11934016). J.-M.Z. is supported by the Energy & Biosciences Institute through the EBI-Shell program. The work at Peking University was supported by National Natural Science Foundation of China (grant no. 11974025). **Author contributions:** Superconductivity was discovered in EuO/KTO (111) and LAO/KTO (111) samples by C.L. with assistance from X.Y., T.M.B.-S., X.Z., D.J., and A.B. EuO/KTO samples were grown at Peking University by Y.M. and W.H. (see fig. S8 and table S1) and at Argonne by C.L. with assistance from J.P. and A.B.; LAO/KTO samples were made at Argonne by X.Y. with assistance from D.D.F. All transport measurements were planned by C.L. and A.B. and carried out by C.L. with assistance from T.M.B.-S., J.S.J., X.Z., B.F., and D.J. Samples from Peking University were measured in collaboration with X.Y. and J.S. X-ray diffraction and XANES measurement were carried out by H.Z. and X.Y. at the Advanced Photon Source at Argonne, with assistance from D.D.F. TEM sample preparation and HRTEM measurements were carried out by Y.L. and J.W. at CNM, Argonne. STEM and STEM-EELS measurements were carried out by H.-W.H. and J.-M.Z. at UIUC. The paper was written by C.L. and A.B. with inputs from all coauthors. A.B. supervised the project. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data for all graphs presented in this paper are available from the Harvard Dataverse (*58*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/371/6530/716/suppl/DC1
Materials and Methods
Figs. S1 to S8
Table S1

13 December 2019; resubmitted 14 April 2020
Accepted 8 January 2021
Published online 21 January 2021
10.1126/science.aba5511

THERMOELECTRICS

Enhanced atomic ordering leads to high thermoelectric performance in $AgSbTe_2$

Subhajit Roychowdhury[1]*, Tanmoy Ghosh[1]*, Raagya Arora[2], Manisha Samanta[1], Lin Xie[3], Niraj Kumar Singh[4], Ajay Soni[4], Jiaqing He[3], Umesh V. Waghmare[2,5,6], Kanishka Biswas[1,5,6]†

High thermoelectric performance is generally achieved through either electronic structure modulations or phonon scattering enhancements, which often counteract each other. A leap in performance requires innovative strategies that simultaneously optimize electronic and phonon transports. We demonstrate high thermoelectric performance with a near room-temperature figure of merit, *ZT* ~ 1.5, and a maximum *ZT* ~ 2.6 at 573 kelvin, by optimizing atomic disorder in cadmium-doped polycrystalline silver antimony telluride ($AgSbTe_2$). Cadmium doping in $AgSbTe_2$ enhances cationic ordering, which simultaneously improves electronic properties by tuning disorder-induced localization of electronic states and reduces lattice thermal conductivity through spontaneous formation of nanoscale (~2 to 4 nanometers) superstructures and coupling of soft vibrations localized within ~1 nanometer around cadmium sites with local strain modulation. The strategy is applicable to most other thermoelectric materials that exhibit inherent atomic disorder.

Thermoelectric materials, for their ability to convert waste heat into electricity without much maintenance, offer an efficient alternative route of electrical and thermal energy utilization and management (*1*–*6*). In the mid-temperature range, high-performance thermoelectric materials can be used in conversion of untapped heat to electricity in, for example, automobile exhausts, various industrial sectors, and power plants. The best performing thermoelectrics for mid- to high temperatures are highly doped conventional semiconductors, such as PbTe (*3*), GeTe (*7*), and SnSe (*4*, *8*, *9*); composed of one or more heavy elements; or have complex crystal structures, such as Zn_4Sb_3 (*10*), $CoSb_3$ (*11*), and Mg_3Bi_2 (*12*). The waste heat–to–electrical energy conversion efficiency of these materials primarily depends on the dimensionless thermoelectric figure of merit, $ZT = \frac{S^2\sigma}{\kappa_e+\kappa_L}T$, where S is the Seebeck coefficient, σ is the electrical conductivity, T is the absolute temperature, and κ_e and κ_L are the electronic and lattice thermal conductivities, respectively (*1*–*4*). Achieving high thermoelectric performance in a material usually involves increasing the power factor (σS^2) or decreasing the lattice thermal conductivity (κ_L).

The established electronic structure modulation strategies, such as valence band convergence (*13*) and resonant doping (*14*), increase σS^2. Similarly, various strategies are also available for phonon scattering optimizations to reduce κ_L, such as the introduction of nano-, meso-, or all-scale hierarchical architectures (*3*); anharmonic vibrations by using various intrinsic effects such as layered structure (*9*); lone pairs–induced rattling (*15*); ferroelectric instability (*16*); and liquid-like cationic motion (*17*). However, most of these electronic structure and phonon transport optimization strategies almost independently improve σS^2 and reduce κ_L, respectively, or adversely affect one another. Therefore, the challenge of improving the thermoelectric performance of a material is to devise a single strategy to optimize σS^2 and κ_L simultaneously.

We developed a strategy that is based on engineering of atomic disorder on a lattice to achieve ultrahigh thermoelectric performance. The atomic disorder, which is ubiquitous in most of the high-performance thermoelectric materials, strongly modulates the electronic and thermal transports. The disorder and impurities present in the lattice scatter charge carriers, limiting their mean free path (l) and σ to a finite value as $T \to 0$. In a noninteracting electronic system, σ reaches to its minimum value (σ_{min}) when l becomes equal to the interatomic spacing (a), as given by the Mott-Ioffe-Regel limit $\sigma_{min} \approx \frac{0.03e^2}{\hbar a}$, where $\hbar$ is Planck's constant h divided by 2π (*18*, *19*). However, σ vanishes in three-dimensional systems beyond a critical level of disorder owing to Anderson localization of electronic states (*20*–*22*). The Anderson metal-insulator transition by disorder-induced localization of electronic states can be understood through the concept of "mobility edge" (*18*). At a given disorder strength, a critical energy (E_c), called mobility edge, separates the localized electronic states from the extended states (Fig. 1A) (*18*, *23*). With increasing disorder, E_c shifts (Fig. 1A, horizontal red arrow) deeper into valence band (VB) [or conduction band (CB)] and closer to the Fermi energy, E_F, and σ decreases as a consequence. A material becomes an insulator when E_F falls within E_c of the band edge (within the localized electronic states). Therefore, by tuning the atomic disorder or the degree of atomic ordering, we can engineer the electronic properties. Furthermore, electronic density of states (DOS) deviates from the normal parabolic energy dependence, leading to a stronger one as E_c moves closer to E_F (*24*). Such distorted electronic DOS, and consequently enhanced effective mass (m^*) and S, are well known to have beneficial effects on the thermoelectric properties (*14*, *25*).

However, disorder-induced electronic localization has conflicting effects on σ and S. As the localization increases with increasing disorder, σ decreases, whereas S increases (Fig. 1A). Therefore, achieving optimally high σS^2 requires tuning the disorder. Several studies have been carried out that use Anderson localization to achieve high thermoelectric performance (*24*, *26*–*28*). However, σS^2 would be suboptimal deep inside the localization regime because σ continues to decrease with increasing disorder and vanishes at E_c, whereas S increases and assumes a constant value, determined by the conductivity exponent, at E_c (*29*). Therefore, we investigated thermoelectric properties in the asymptotic regime of Anderson localization, called weak localization (WL) (*30*, *31*). In the limit of weak disorder, coherent backscattering of diffusive electrons causes partial localization of electronic states because of quantum interference effects. This WL is often regarded as a precursor of Anderson localization (*22*). In the WL regime, although electrical conductivity decreases because of quantum interference effect (*30*, *31*), the electrical conductivity is still expected to be much higher in a three-dimensional solid as compared with that in the insulating state with strong disorder.

We demonstrated the optimization of atomic ordering while achieving high thermoelectric performance in the p-type silver antimony telluride ($AgSbTe_2$). $AgSbTe_2$ is well known in thermoelectrics for its good thermoelectric properties (*32*–*34*) and for its applications when alloyed with GeTe and PbTe, in materials such as TAGS and LAST (*35*, *36*). Despite being such an important thermoelectric material, many of its physical properties are yet to be resolved unambiguously. Although $AgSbTe_2$ assumes a rocksalt-type cubic crystal structure at room temperature, the exact nature of Ag or Sb occupancy in the cationic sites—whether they are randomly distributed or have a particular ordered arrangement—is still ambiguous (*37*–*40*). $AgSbTe_2$ has a high Seebeck coefficient

[1]New Chemistry Unit, Jawaharlal Nehru Centre for Advanced Scientific Research (JNCASR), Jakkur P.O., Bangalore 560064, India. [2]Theoretical Sciences Unit, JNCASR, Jakkur P.O., Bangalore 560064, India. [3]Department of Physics, Southern University of Science and Technology, Shenzhen 518055, China. [4]School of Basic Science, Indian Institute of Technology Mandi, Mandi, Himachal Pradesh 175075, India. [5]School of Advanced Materials, JNCASR, Jakkur P.O., Bangalore 560064, India. [6]International Centre for Materials Science, JNCASR, Jakkur P.O., Bangalore 560064, India.
*These authors contributed equally to this work.
†Corresponding author. Email: kanishka@jncasr.ac.in

(S ~ 260 μV K^{-1} at 300 K) because of highly degenerate and flat valence bands (*40*–*42*). Another intriguing property of $AgSbTe_2$ is its low κ_L (κ_L ~ 0.5 W m^{-1} K^{-1} at 300 K), which has been attributed to spontaneously formed nanoscale superstructures (*38*, *43*), low frequency optical phonons (*44*), and anharmonicity (*45*). Nanoscale superstructures in $AgSbTe_2$ are highly effective in lowering its κ_L because the phonon mean free path in $AgSbTe_2$ is ~3 to 6 nm (*38*). However, a few reports did not observe such superstructures in $AgSbTe_2$, particularly in polycrystalline samples (*46*–*48*).

We present evidence for increased cationic ordering and the consequent formation of nanoscale superstructures upon Cd doping in polycrystalline $AgSbTe_2$ by use of temperature-dependent electronic transport, transmission electron microscopy (TEM), and first-principles density functional theory (DFT) calculations. By optimizing such atomic ordering, we achieved ultrahigh thermoelectric performance (Fig. 1, B and C) in 6 mol % Cd–doped polycrystalline $AgSbTe_2$. This value is one of the highest among state-of-the-art bulk thermoelectric materials from room temperature to 573 K. We report a near–room temperature ZT of ~1.5 and the maximum ZT of ~2.6 at 573 K. The high ZT values are reproducible (within the error bar) with respect to different batches of samples and several heating-cooling cycles (figs. S1 and S2) (*49*). With this high ZT over a wide temperature range, we calculated device thermoelectric figure of merit ZT_{dev} ~ 1.9 (*49*) for a 6 mol % Cd–doped polycrystalline $AgSbTe_2$ sample in the temperature range from room temperature to 573 K (Fig. 1C).

Disorder-induced WL has discernible effects in the temperature variation of electrical resistivity (ρ) at lower temperatures with a resistivity minimum at a temperature T_m, below which ρ increases with decreasing temperature. T_m increases as disorder increases in the system (*30*, *50*, *51*). The resistivity minimum is a result of competitive effects of WL and disorder-enhanced electron-electron interaction (EEI) from the diffusive electronic transport with inelastic scattering processes such as electron-phonon scattering (*30*, *50*, *51*). Disorder-induced localized electronic states and enhanced EEI cause increase in ρ with lowering of temperature, whereas the inelastic scattering process has an opposite effect and tends to diminish electronic localization by destroying phase coherence. At very low temperatures, disorder-enhanced EEI dominates the electronic transport with a $\sqrt{T}$-dependent contribution to the resistivity (*30*, *50*, *51*). Temperature variation of ρ below room temperature for both the pristine $AgSbTe_2$ and 6 mol % Cd–doped $AgSbTe_2$ exhibits such resistivity minima and $\sqrt{T}$ dependence at low temperatures, indicating the presence of disorder-induced localized electronic states (Fig. 2, A to C, and fig. S3A). A strong suppression of T_m from ~61 K in pristine $AgSbTe_2$ to ~10 K in 6 mol % Cd–doped $AgSbTe_2$ indicates drastic reduction in disorder in $AgSbTe_2$ through Cd substitution. The temperature-dependent Seebeck coefficient of pristine $AgSbTe_2$ (fig. S3B) exhibits a broad hump at low temperatures, which indicates enhanced Seebeck coefficient because of the presence of localized electronic states. The broad hump in temperature-dependent Seebeck coefficient is, however, strongly suppressed to lower temperatures in 6 mol % Cd–doped $AgSbTe_2$ (fig. S3B), which further corroborates to the reduction of disorder in $AgSbTe_2$ upon Cd substitution.

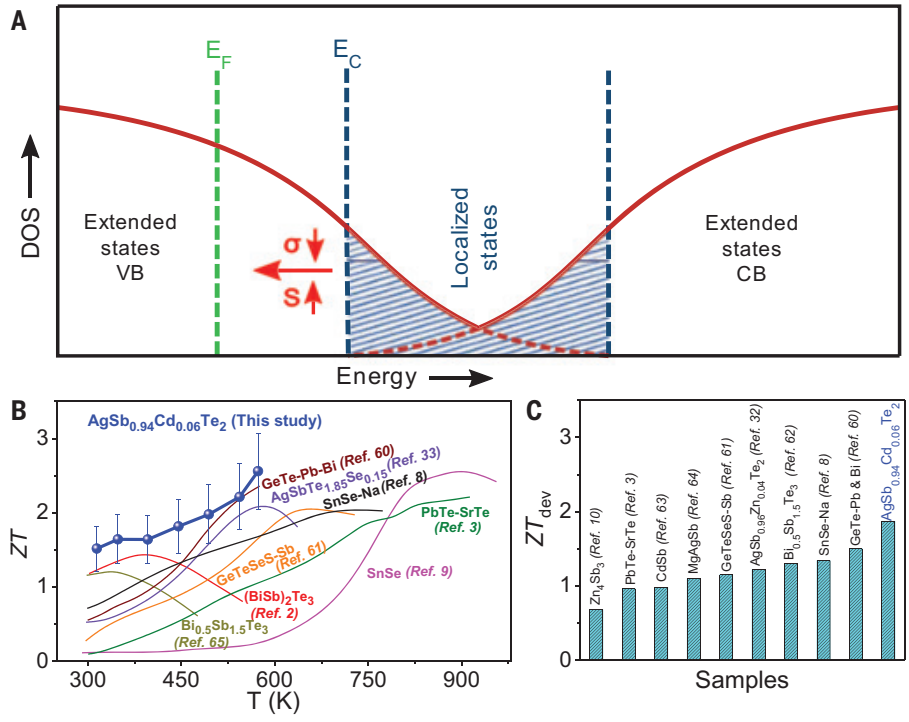


Fig. 1. Schematic of disorder-tuning of electronic structure and thermoelectric performance of Cd-doped polycrystalline $AgSbTe_2$. (**A**) A schematic of disorder-tuning of mobility edge (E_c) separating the localized electronic states from the extended ones and how it affects the electrical conductivity (σ) and Seebeck coefficient (S). E_F denotes the Fermi energy. The horizontal red arrow denotes the shift in E_c toward the Fermi energy E_F with increasing disorder. (**B** and **C**) Comparison of (B) thermoelectric figure of merit (ZT) and (C) device figure of merit (ZT_{dev}), respectively, of Cd-doped polycrystalline $AgSbTe_2$ with other state-of-the-art thermoelectric materials (*2*, *3*, *8*–*10*, *32*, *33*, *60*–*65*). ZT_{dev} is calculated for the entire measured temperature range, as given in the respective references.

Pristine $AgSbTe_2$ has two distinct types of disorder: (i) the precipitation of impurity phases, such as Ag_2Te and Sb_2Te_3, and (ii) Ag or Sb cationic disorder (*34*, *52*, *53*). The impurity phases appear because the cubic structure of $AgSbTe_2$ forms in a slightly off-stoichiometric part of the Ag_2Te-Sb_2Te_3 pseudobinary phase diagram (*46*, *54*, *55*). Cadmium doping suppresses both types of disorder. Similar to the earlier reports (*32*, *34*, *46*), we observed the presence of Ag_2Te impurity phase in our pristine polycrystalline $AgSbTe_2$ specimen, as evident from the room-temperature powder x-ray diffraction (PXRD) patterns (fig. S4B) and the endothermic peak corresponding to the α-Ag_2Te → β-Ag_2Te transition ~425 K in differential scanning calorimetry (DSC) (fig. S5). With increasing Cd concentration, the gradual suppression of Ag_2Te impurity phase is evident (figs. S4B and S5) and nearly disappears in 6 mol % Cd–doped $AgSbTe_2$.

We found evidence for the enhancement in cationic ordering upon Cd doping by using high-resolution scanning transmission electron microscopy (HR-STEM) high-angle annular dark field (HAADF) images (Fig. 2D and fig. S7C) and selected area electron diffraction (SAED) patterns (Fig. 2E and figs. S7 to S9). Whereas a few earlier experimental reports (*37*, *38*) revealed evidence for nanoscale and even partial long-range ordering, other experimental reports (*46*–*48*) did not find evidence for such nanoscale superstructures in pristine polycrystalline $AgSbTe_2$. We have not observed the ordered nanoscale domains in

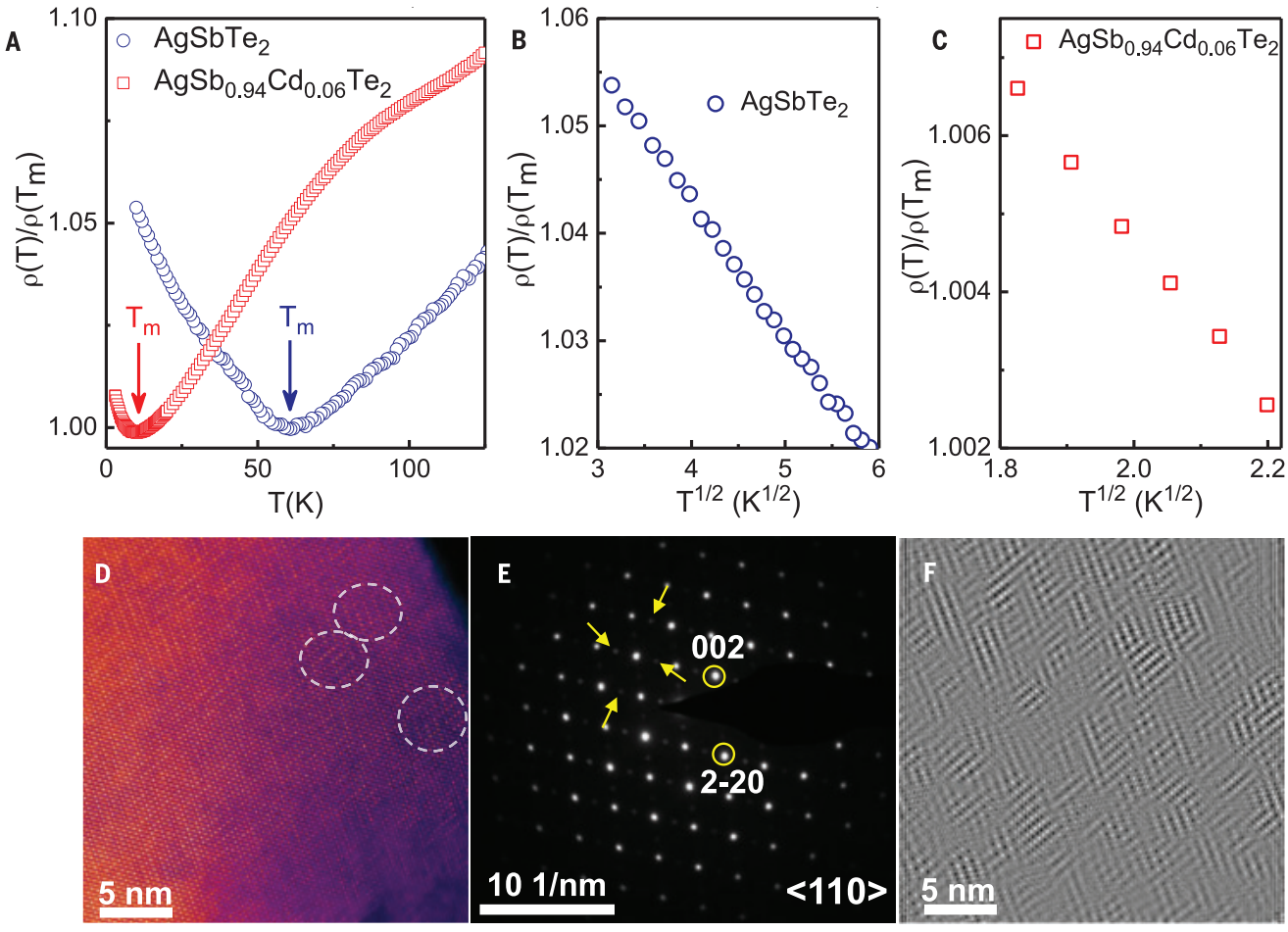


Fig. 2. Enhanced cationic ordering in Cd-doped polycrystalline $AgSbTe_2$. (**A**) Temperature dependence of electrical resistivity (ρ) of pristine $AgSbTe_2$ and 6 mol % Cd–doped $AgSbTe_2$ at low temperature, where T_m represents the temperature at which resistivity minimum occurs. (**B** and **C**) Temperature variation in ρ of (B) pristine $AgSbTe_2$ and (C) 6 mol % Cd–doped $AgSbTe_2$ exhibiting linearity with $T^{½}$ as an evidence of dominance of disorder-enhanced EEI. (**D**) HRSTEM-HAADF image and (**E**) SAED pattern exhibit the formation of cation ordering in 6 mol % Cd–doped $AgSbTe_2$. The dashed white circles in (D) indicate typical cation ordered regions. Yellow arrows in (E) indicate signature of cation ordering. (**F**) Inverse FFT image of the HRSTEM-HAADF image shown in (D), after filtering the superlattice reflections from FFT.

STEM-HAADF images or superlattice reflections in SAED pattern in pristine polycrystalline $AgSbTe_2$ (fig. S6), although our pristine sample may have some local cationic ordering, as previously found (*37*–*40*). However, STEM-HAADF images (Fig. 2D and fig. S7C) and HRTEM images (figs. S7A; S8, A and D; and S9A) of 6 mol % Cd–doped $AgSbTe_2$ show cationic ordering and nanoscale domains, respectively. Superlattice reflections in the SAED pattern for the cationic ordering (Fig. 2E and figs. S7B, S8B, and S9B) and nanoscale superstructures (figs. S8B and S9B) are clearly evident for 6 mol % Cd–doped $AgSbTe_2$. We confirmed the presence of superlattice reflections for the nanoscale superstructures with fast Fourier transform (FFT) images of the HRTEM images (fig. S8, C and E). The corresponding inverse FFT (IFFT) images also reveal the presence of nanoscale superstructure domains (fig. S8F). We estimated the size of nanoscale ordered superstructure domains to be ~2 to 4 nm. Moreover, in the SAED patterns (Fig. 2E and figs. S7B, S8B, and S9B) and FFT of the HR-STEM and HRTEM images (figs. S7D and S8C), we observed the presence of low-intensity spots halfway between the fundamental spots, which indicates doubling of the unit cell because of the cation ordering in Cd-doped $AgSbTe_2$. In $AgSbTe_2$, multiple ordered structures are nearly degenerate in energy (*39*), and all of these ordered structures have the possibility of formation instead of the lowest-energy ordered structure because of local fluctuations in lattice potential (*37*–*40*). The slight differences between the lattices of various ordered structures cause strain in the lattice, and we observed such strain ripples in IFFT of an HRSTEM-HAADF image (Fig. 2F) and IFFT of an HRTEM image (fig. S10) of 6 mol % Cd–doped $AgSbTe_2$. These strain ripples are also visible in a (220) dark-field TEM image of 6 mol % Cd–doped $AgSbTe_2$ (fig. S9C) and its corresponding IFFT image (fig. S9D).

We determined formation energies of ordered and various disordered pristine and Cd-doped $AgSbTe_2$ configurations using DFT to further elucidate the influence of Cd doping on increased cationic ordering (Fig. 3A). We started with the cubic AF-IIb structure with the *Fd-3m* space group, which is the lowest-energy ordered $AgSbTe_2$ structure (*39*, *40*), and created cationic antisite disorder by exchanging nearest-neighbor (1N) and next-nearest-neighbor (2N) Ag or Sb cations. In pristine $AgSbTe_2$, we found that the ordered structure is 0.46 and 0.34 eV lower in energy per formula unit than the disordered 1N and 2N configurations, respectively. This is consistent with earlier DFT results (*39*, *40*, *42*), indicating a greater stability of ordered $AgSbTe_2$ phase than that of the disordered phase. The formation

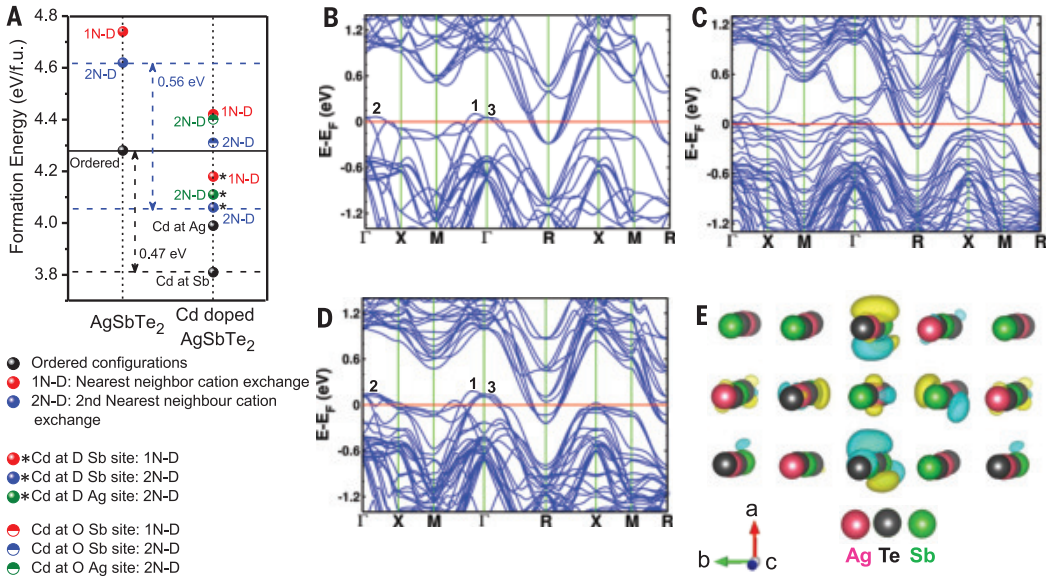


Fig. 3. DFT-calculated formation energy, electronic structure, and charge distribution. (**A**) Formation energies of ordered (*O*) and disordered (*D*) configurations of pristine and Cd-doped $AgSbTe_2$. (**B** to **D**) Electronic structures of (B) ordered $AgSbTe_2$, (C) disordered $AgSbTe_2$, and (D) Cd-doped ordered $AgSbTe_2$. In (B) and (D), first, second, and third valence band maxima are indicated with 1, 2, and 3. (**E**) Isosurface of wave function of the defect band (first band above the Fermi energy) at Γ point of disordered $AgSbTe_2$ exhibiting localization of electronic states at Te atoms.

energy of ordered $AgSbTe_2$ further reduces by 0.29 and 0.47 eV per formula unit when Cd is substituted at Ag and Sb sites, respectively, establishing that Cd atoms preferentially occupy Sb sites. Similarly, formation energy of disordered $AgSbTe_2$ is lower when Cd atoms are substituted at Sb sites than at Ag sites. The 2N configuration of disordered $AgSbTe_2$ remains more favorable than the 1N configuration after Cd substitution. In disordered $AgSbTe_2$, Cd substitution at the disordered Sb site of 2N configuration has the lowest formation energy as well. Thus, Cd atoms preferentially occupy the disordered Sb sites rather than the ordered Sb sites within the disordered $AgSbTe_2$ structure, eliminating Sb atoms from the disordered sites and promoting cationic order in $AgSbTe_2$. Also, Cd doping at the Sb site of ordered $AgSbTe_2$ has the lowest formation energy, and the reduction in formation energy from the pristine to Cd-doped phases in the presence of disorder (0.56 eV) is greater than that in the ordered configuration (0.47 eV). Fundamental origin of Cd doping–induced enhanced cationic ordering involves the cationic radii of Ag^+, Sb^{3+}, and Cd^{2+}, which are 115, 76, and 95 pm, respectively. Large cationic charge and size differences between Ag^+ and Sb^{3+} in their octahedral environment are not so favorable for an ordered structure. Cd^{2+}, with its intermediate cationic radii and charge state, promotes cationic ordering and alleviate large potential fluctuations of the lattice-disorder suppressing localization.

The Cd doping–induced atomic ordering has profound effects on the thermoelectric properties of $AgSb_{1-x}Cd_xTe_2$ (x = 0 to 0.08) (Fig. 4 and fig. S11). Cd doping up to 6 mol % increases σ throughout the temperature range from near room temperature to 573 K (Fig. 4A), with near–room temperature σ increasing from ~121 S cm^{-1} for pristine $AgSbTe_2$ to ~222 S cm^{-1} in 6 mol % Cd–doped $AgSbTe_2$. Temperature variation of S of $AgSb_{1-x}Cd_xTe_2$ (x = 0 to 0.06) (Fig. 4B) exhibits overall reduction in S with increasing Cd concentration. The near–room temperature S decreases from ~279 μV K^{-1} for pristine $AgSbTe_2$ to ~248 μV K^{-1} in 6 mol % Cd–doped $AgSbTe_2$. As a result of this large enhancement in σ while maintaining a high S, σS^2 of Cd-doped $AgSbTe_2$ remains higher throughout the temperature range from near room temperature to 573 K than that of the pristine $AgSbTe_2$ (Fig. 4C). The near–room temperature value of σS^2 of 6 mol % Cd–doped $AgSbTe_2$ composition (~13.6 μW cm^{-1} K^{-2}) is substantially higher than that of the pristine $AgSbTe_2$ (~9.5 μW cm^{-1} K^{-2}). The maximum of σS^2 of 6 mol % Cd–doped $AgSbTe_2$ is ~18.3 μW cm^{-1} K^{-2} at 543 K.

Electrical transports of $AgSbTe_2$ have the contribution from both low mobility heavy holes (higher concentration) and highly mobile light electrons (lower concentration) as it has semimetallic electronic structure with flat valence band and narrow conduction band (*41*). We estimated the concentrations of electrons (n) and holes (p) and mobilities of electron (μ_n) and hole (μ_p) using the two-carrier model (fig. S12) (*41*, *49*). At room temperature, the pristine $AgSbTe_2$ has a hole concentration of ~ 4.3 × 10^{19} cm^{-3} and hole mobility of ~15.1 cm^2 V^{-1} s^{-1}, which is similar to previous results (*41*). In $AgSbTe_2$, 6 mol % Cd doping increases the room-temperature hole concentration to ~8.3 × 10^{19} cm^{-3}. However, the room-temperature hole mobility in 6 mol % Cd–doped $AgSbTe_2$ is ~ 12.5 cm^2 V^{-1} s^{-1}, which is similar to the pristine $AgSbTe_2$ value. Two important aspects are clear from the temperature-dependent Hall mobility of holes (fig. S12), which are the majority carrier in $AgSbTe_2$ system. First, hole mobility (μ_p) monotonously increases with decreasing temperature in pristine $AgSbTe_2$ until the temperature at which we observed the resistivity minimum (Fig. 2A) then subsequently exhibits a broad hump. However, the hole mobility (μ_p) in 6 mol % Cd–doped $AgSbTe_2$ increases monotonously with decreasing temperature in the measured temperature range. Second, although 6 mol % Cd–doping increases the hole concentration to ~ 2.4 × 10^{19} cm^{-3} from the value of 1.3 × 10^{19} cm^{-3} in pristine $AgSbTe_2$ at 10 K, the degradation in hole mobility is not as large. The values of hole mobility at 10 K in pristine $AgSbTe_2$ and 6 mol % Cd–doped $AgSbTe_2$ are ~74.4 and 61.2 cm^2 V^{-1} s^{-1}, respectively. Both of these aspects demonstrate the substantial role of Cd doping–induced enhanced cationic ordering in electrical transport of $AgSbTe_2$.

From the electronic structure, we uncovered mechanisms and major roles of Cd doping and cationic ordering on electronic transport. Pristine $AgSbTe_2$ in the ordered phase is a semimetal with a pseudogap of –0.29 eV (Fig. 3B). Its valence band near E_F is predominantly constituted of Te 5*p* orbitals, whereas the conduction band has dominant contributions from Sb 5*p* and Ag 5*s* orbitals (fig. S13B). As we introduced cationic disorder in $AgSbTe_2$, very flat (localized) mid-gap electronic states emerged (Fig. 3C). Visualization of the isosurfaces of wave functions at Γ and M points (Fig. 3E and figs. S14 to 20) of disordered structures clearly shows the localized nature of these mid-gap states (near E_F) centered on Te and disordered sites of Sb atoms in the disordered structure. Similarly, electronic DOS of disordered $AgSbTe_2$ (fig. S13D) shows many states at the band edges

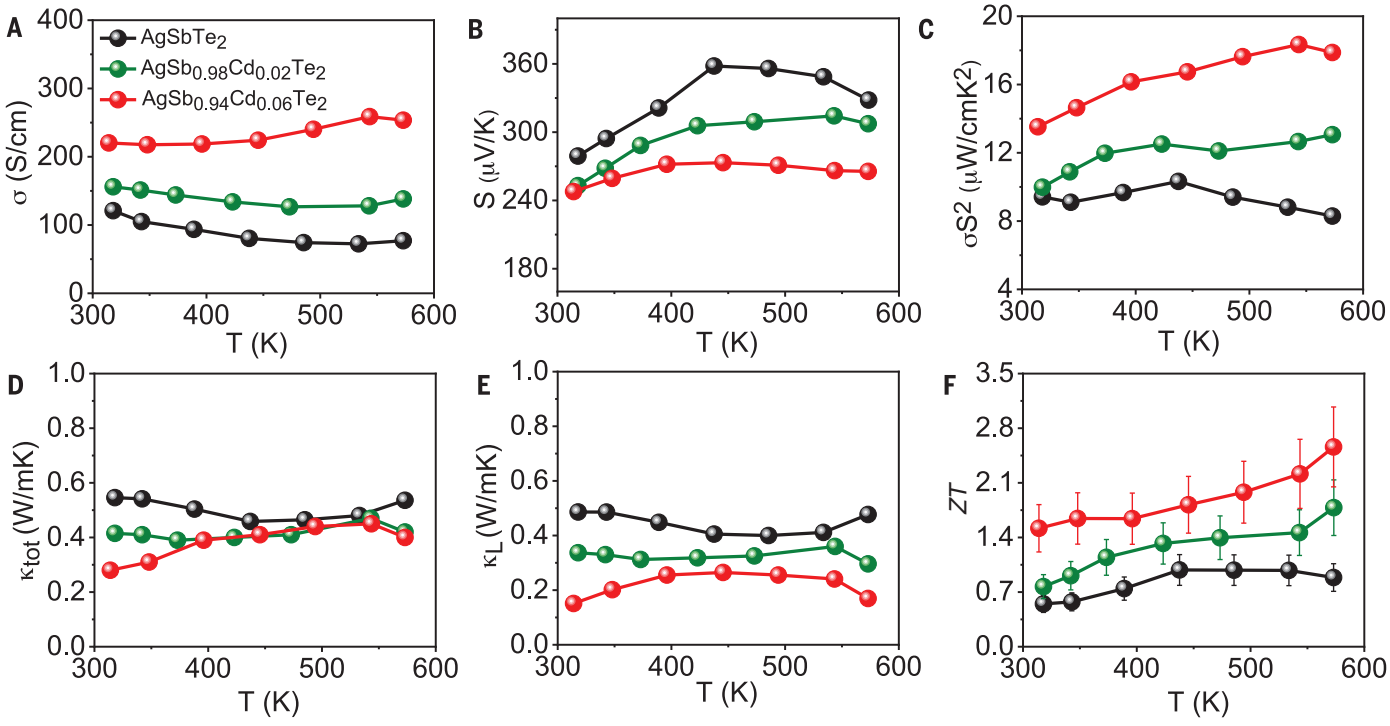


Fig. 4. Thermoelectric properties of $AgSb_{1-x}Cd_xTe_2$ (x = 0 to 0.06). (**A** to **F**) Temperature variation of (A) electrical conductivity (σ), (B) Seebeck coefficient (S), (C) power factor (σS^2), (D) total thermal conductivity (κ_{tot}), (E) lattice thermal conductivity (κ_L), and (F) thermoelectric figure of merit (ZT) of $AgSb_{1-x}Cd_xTe_2$. The uncertainty of ZT measurement is ~20%.

that are confirmed to be localized. Thus, Ag or Sb cationic disorder produces localized electronic states at the valence and conduction band edges lining the gap, verifying the schematic electronic structure shown in Fig. 1A. Cd doping, which favors cationic ordering and shifts E_F away from the mobility edge E_c, thus improves electrical conductivity by delocalizing these electronic states.

The high S values of pristine $AgSbTe_2$ have been attributed to its highly degenerate band structure and the presence of closely spaced multiple valence bands (*41*, *42*). However, earlier calculations of electronic structure estimate lower m^* (~0.9 m_0) (*33*) than what (~1.7 m_0 to 2 m_0) is required to account for the observed S of pristine $AgSbTe_2$ (*33*, *41*). The high S values of pristine $AgSbTe_2$ originate from the presence of the mobility edge in the vicinity of E_F, which distorts the electronic structure, enhancing m^*. Although S decreases with increasing Cd concentration up to 6 mol % because of the increased p-type carrier density, the decrement is negligible. High S of Cd-doped $AgSbTe_2$ also arises from electronic structure modulations and convergence of multiple valence bands caused by Cd doping (Fig. 3D). The energy differences between VB1 and VB2 reduce by 7 meV, and between VB1 and VB3, that reduce by 9 meV. Further, new heavy hole bands arise in the valence band manifold, pushing E_F deeper into the valence band, as we found particularly obvious at the X and M points (Fig. 3D).

Cd doping up to 6 mol % in polycrystalline $AgSbTe_2$ improves its thermoelectric performance further by reducing κ_{tot} and κ_L (Fig. 4, D and E). The near–room temperature κ_{tot} ~ 0.55 W m^{-1} K^{-1} of pristine $AgSbTe_2$ decreases with increasing Cd doping and reaches a low value of ~0.28 W m^{-1} K^{-1} in 6 mol % Cd–doped $AgSbTe_2$. Lattice thermal conductivity reduces upon Cd doping up to 6 mol % in $AgSbTe_2$ throughout our measured temperature range. The κ_L of ~0.16 W m^{-1} K^{-1} in 6 mol % Cd–doped $AgSbTe_2$ nearly equals the recently developed diffuson model of minimum κ_L (~0.20 W m^{-1} K^{-1}) (*56*), which was estimated by using the measured sound velocity (table S1). The ultralow κ_L we observed in Cd-doped polycrystalline $AgSbTe_2$ is comparable with the κ_L of single crystalline SnSe [~0.23 W m^{-1} K^{-1} at ~923 K (*9*)] and other low–thermal conductive materials such as Tl_3VSe_4 [~0.3 W m^{-1} K^{-1} at 300 K (*57*)] and Ag_9GaSe_6 [0.15 W m^{-1} K^{-1} at 300 K (*58*)]. Because the mean free path of phonons (~3 to 6 nm) in pristine $AgSbTe_2$ (*38*) is on the order of the nanoscale superstructure size (~2 to 4 nm), the dominant role of the nanoscale superstructures in phonon scattering is more plausible. Our observations of Cd doping–induced formation of the nanoscale superstructures and enhanced lattice strain (Fig. 2F and figs. S9 and S10) concurrent with strong suppression of κ_L in 6 mol % Cd–doped $AgSbTe_2$ validate such a scenario.

Additionally, we found that the variation in heat capacity (C_p) of 6 mol % Cd–doped $AgSbTe_2$ with temperature (2 to 35 K) could only be accounted properly by including several Einstein modes along with the Debye modes (fig. S22 and table S2). The frequency of the lowest Einstein mode is ~19.3 cm^{-1}. The presence of a broad Boson-like peak in the C_p/T^3 versus T plot at low temperature (fig. S22B) further signifies the presence of an excess phonon density of states and contribution to C_p from low-energy optical phonon modes (*15*). Our first-principles calculations of phonons confirm the presence of many low-frequency Γ-point optical modes in the frequency range from 30 to 80 cm^{-1} (table S3). We analyzed effects of Cd doping on thermal conductivity by estimating the interactions of Cd substituents in $AgSb_{0.94}Cd_{0.06}Te_2$ with phonons at the lowest two orders: (i) large forces on atoms in the neighborhood of Cd (fig. S23A), revealing which phonons are excited by Cd substitution, and (ii) mixing or coupling between the phonons (fig. S23B) of $AgSbTe_2$ that arise from Cd substitution. Projection of the former on normal modes shows that many of the low-frequency phonons are excited, causing local strains to nucleate at the Cd site, and the latter shows that these soft modes mix strongly because of Cd-induced coupling. Thus, Sb-Ag

ordering and associated superstructures are favored by Cd substitution, and coupled soft vibrations localized within 1 nm of Cd sites act as heat sinks in the ordered regions. With large Gruneisen parameters of these soft modes (*45*) and consequent strain modulation, acoustic phonons are thus scattered effectively, resulting in ultralow κ_L. Although Cd doping strongly affects κ_L of $AgSbTe_2$, we found that the effect of disorder is much smaller, particularly on the low-frequency phonons (fig. S24).

As a result of this reduced κ_L and enhanced σS^2, *ZT* of Cd-doped (2 to 6 mol %) polycrystalline $AgSbTe_2$ improves tremendously (Fig. 4F and fig. S11). The near–room temperature *ZT* ~0.6 of $AgSbTe_2$ improves to a value of ~1.5 in 6 mol % Cd–doped $AgSbTe_2$. Cadmium doping improves *ZT* throughout our measured temperature range, and we achieved a maximum *ZT* ~ 2.6 at 573 K in 6 mol % Cd–doped $AgSbTe_2$. The measured *ZT* value in 6 mol % Cd–doped $AgSbTe_2$ is one of the highest in the mid-temperature region (from room temperature to 573 K) (Fig. 1B). As a result of this ultrahigh thermoelectric performance for a wide temperature range, 6 mol % Cd–doped $AgSbTe_2$ results in device figure of merit ZT_{dev} of ~1.9 for the measured temperature range from near room temperature to 573 K (Fig. 1C). We have experimentally measured a high output power of ~97 mW for a temperature difference of 300 K, with the cold end being at room temperature from a single-leg thermoelement made by using 6 mol % Cd–doped polycrystalline $AgSbTe_2$ (fig. S25A). The experimentally estimated conversion efficiency of the fabricated single-leg thermoelectric module is ~9.8% for a temperature difference of 300 K, with the cold end being at room temperature (fig. S25B). The experimentally estimated conversion efficiency is lower than the theoretically calculated value using ZT_{dev} (fig. S25B), which can be further improved by optimizing the contact materials and interfacial resistance (*59*). The high thermoelectric performance in the mid-temperature region opens the opportunities to convert heat to electrical energy efficiently from various waste heat sources, including power plants and industrial processes.

We found evidence for the positive effects of atomic disorder tuning on electronic and phonon transports, which are simultaneously optimized toward high thermoelectric performance. Optimal Cd doping in $AgSbTe_2$ enhances the cation ordering, which delocalizes the electronic states and thereby moves E_F away from the mobility edge and enhances the power factor. Further, nanoscale ordered superstructures and enhanced anharmonic scattering of phonons induced by Cd doping in $AgSbTe_2$ lower its κ_L. The demonstrated strategy opens new avenues of achieving high thermoelectric performance in bulk materials in particular, because most of the high-performance thermoelectric materials have inherent atomic disorder.

REFERENCES AND NOTES

1. L.-D. Zhao, V. P. Dravid, M. G. Kanatzidis, *Energy Environ. Sci.* **7**, 251–268 (2014).
2. B. Poudel *et al*., *Science* **320**, 634–638 (2008).
3. K. Biswas *et al*., *Nature* **489**, 414–418 (2012).
4. C. Chang *et al*., *Science* **360**, 778–783 (2018).
5. G. Tan, L.-D. Zhao, M. G. Kanatzidis, *Chem. Rev.* **116**, 12123–12149 (2016).
6. J. He, T. M. Tritt, *Science* **357**, eaak9997 (2017).
7. S. Roychowdhury, M. Samanta, S. Perumal, K. Biswas, *Chem. Mater.* **30**, 5799–5813 (2018).
8. L.-D. Zhao *et al*., *Science* **351**, 141–144 (2016).
9. L.-D. Zhao *et al*., *Nature* **508**, 373–377 (2014).
10. G. J. Snyder, M. Christensen, E. Nishibori, T. Caillat, B. B. Iversen, *Nat. Mater.* **3**, 458–463 (2004).
11. X. Shi *et al*., *J. Am. Chem. Soc.* **133**, 7837–7846 (2011).
12. J. Mao *et al*., *Science* **365**, 495–498 (2019).
13. Y. Pei *et al*., *Nature* **473**, 66–69 (2011).
14. J. P. Heremans *et al*., *Science* **321**, 554–557 (2008).
15. M. K. Jana, K. Biswas, *ACS Energy Lett.* **3**, 1315–1324 (2018).
16. A. Banik *et al*., *Energy Environ. Sci.* **12**, 589–595 (2019).
17. H. Liu *et al*., *Nat. Mater.* **11**, 422–425 (2012).
18. N. F. Mott, *Phys. Today* **31**, 42–47 (1978).
19. O. Gunnarsson, M. Calandra, J. E. Han, *Rev. Mod. Phys.* **75**, 1085–1099 (2003).
20. P. W. Anderson, *Phys. Rev.* **109**, 1492–1505 (1958).
21. E. Abrahams, P. W. Anderson, D. C. Licciardello, T. V. Ramakrishnan, *Phys. Rev. Lett.* **42**, 673–676 (1979).
22. A. Lagendijk, B. van Tiggelen, D. S. Wiersma, *Phys. Today* **62**, 24–29 (2009).
23. G. Semeghini *et al*., *Nat. Phys.* **11**, 554–559 (2015).
24. K. Yamamoto, A. Aharony, O. Entin-Wohlman, N. Hatano, *Phys. Rev. B* **96**, 155201 (2017).
25. G. D. Mahan, J. O. Sofo, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 7436–7439 (1996).
26. C. Villagonzalo, R. A. Römer, M. Schreiber, *Eur. Phys. J. B* **12**, 179–189 (1999).
27. M. H. Lee *et al*., *ACS Nano* **13**, 3806–3815 (2019).
28. Z. Tian, *ACS Nano* **13**, 3750–3753 (2019).
29. J. E. Enderby, A. C. Barnes, *Phys. Rev. B Condens. Matter* **49**, 5062–5064 (1994).
30. P. A. Lee, T. V. Ramakrishnan, *Rev. Mod. Phys.* **57**, 287–337 (1985).
31. G. Bergmann, *Phys. Rep.* **107**, 1–58 (1984).
32. S. Roychowdhury, R. Panigrahi, S. Perumal, K. Biswas, *ACS Energy Lett.* **2**, 349–356 (2017).
33. M. Hong *et al*., *Adv. Energy Mater.* **8**, 1702333 (2018).
34. B. Du, H. Li, J. Xu, X. Tang, C. Uher, *Chem. Mater.* **22**, 5521–5527 (2010).
35. K. F. Hsu *et al*., *Science* **303**, 818–821 (2004).
36. F. D. Rosi, J. P. Dismukes, E. F. Hockings, *Electr. Eng.* **79**, 450–459 (1960).
37. E. Quarez *et al*., *J. Am. Chem. Soc.* **127**, 9177–9190 (2005).
38. J. Ma *et al*., *Nat. Nanotechnol.* **8**, 445–451 (2013).
39. S. V. Barabash, V. Ozolins, C. Wolverton, *Phys. Rev. Lett.* **101**, 155704 (2008).
40. K. Hoang, S. D. Mahanti, J. R. Salvador, M. G. Kanatzidis, *Phys. Rev. Lett.* **99**, 156403 (2007).
41. V. Jovovic, J. P. Heremans, *Phys. Rev. B Condens. Matter Mater. Phys.* **77**, 245204 (2008).
42. S. V. Barabash, V. Ozolins, *Phys. Rev. B Condens. Matter Mater. Phys.* **81**, 075212 (2010).
43. J. Ma *et al*., *Phys. Rev. B Condens. Matter Mater. Phys.* **90**, 134303 (2014).
44. L.-H. Ye *et al*., *Phys. Rev. B Condens. Matter Mater. Phys.* **77**, 245203 (2008).
45. D. T. Morelli, V. Jovovic, J. P. Heremans, *Phys. Rev. Lett.* **101**, 035901 (2008).
46. J. D. Sugar, D. L. Medlin, *J. Alloys Compd.* **478**, 75–82 (2009).
47. J. D. Sugar, D. L. Medlin, *J. Mater. Sci.* **46**, 1668–1679 (2011).
48. P. A. Sharma, J. D. Sugar, D. L. Medlin, *J. Appl. Phys.* **107**, 113716 (2010).
49. Materials and methods are available as supplementary materials.
50. B. L. Altshuler, A. G. Aronov, Electron-electron interaction in disordered conductors, in *Electron-Electron Interactions in Disordered Systems*, A. L. Efros, M. Pollak, Eds. (North-Holland, 1985), pp. 1–153.
51. H. Fukuyama, Interaction effects in the weakly localized regime of two- and three-dimensional disordered systems, in *Electron-Electron Interactions in Disordered Systems*, A. L. Efros, M. Pollak, Eds. (North-Holland, 1985), pp. 155–230.
52. H. Wang, J.-F. Li, M. Zou, T. Sui, *Appl. Phys. Lett.* **93**, 202106 (2008).
53. K. T. Wojciechowski, M. Schmidt, *Phys. Rev. B Condens. Matter Mater. Phys.* **79**, 184202 (2009).
54. R. W. Armstrong, J. W. Faust Jr., W. A. Tiller, *J. Appl. Phys.* **31**, 1954–1959 (1960).
55. J. P. McHugh, W. A. Tiller, S. E. Haszko, J. H. Wernick, *J. Appl. Phys.* **32**, 1785–1785 (1961).
56. M. T. Agne, R. Hanus, G. J. Snyder, *Energy Environ. Sci.* **11**, 609–616 (2018).
57. S. Mukhopadhyay *et al*., *Science* **360**, 1455–1458 (2018).
58. S. Lin *et al*., *Joule* **1**, 816–830 (2017).
59. P. Jood, M. Ohta, A. Yamamoto, M. G. Kanatzidis, *Joule* **2**, 1339–1355 (2018).
60. J. Li *et al*., *Joule* **2**, 976–987 (2018).
61. M. Samanta, K. Biswas, *J. Am. Chem. Soc.* **139**, 9382–9391 (2017).
62. S. I. Kim *et al*., *Science* **348**, 109–114 (2015).
63. S. Wang *et al*., *Chem. Mater.* **27**, 1071–1081 (2015).
64. H. Zhao *et al*., *Nano Energy* **7**, 97–103 (2014).
65. R. Deng *et al*., *Sci. Adv.* **4**, eaar5606 (2018).

ACKNOWLEDGMENTS

The authors acknowledge J. Ghatak (JNCASR) for help during TEM measurement. The authors also acknowledge P. Yanda (JNCASR) for help during preliminary low-temperature resistivity measurements. **Funding:** K.B. acknowledges generous funding support from Swarnajayanti Fellowship, Science and Engineering Research Board (SERB) (SB/SJF/2019-20/06), Department of Science and Technology (DST) (DST/SJF/CSA-02/2018-19), India; Sheik Saqr Fellowship, SSL, ICMS, and JNCASR, India. S.R. and M.S. acknowledge CSIR and UGC, India, for research fellowship, respectively. J.H. acknowledges support from the National Natural Science Foundation of China (11934007) and the Science and Technology Innovation Committee Foundation of Shenzhen (KQTD2016022619565991). U.V.W. acknowledges support from IKST, Sheikh Saqr Fellowship, and J. C. Bose National Fellowship of the DST, India. **Author contributions:** K.B. conceived the idea and designed the study. S.R., T.G., M.S., and K.B. carried out the synthesis, structural, and other characterizations, thermoelectric experiments, and analysis of data. L.X. and J.H. carried out the HR-STEM-HAADF imaging and analysis. N.K.S. and A.S. carried out low-temperature transport experiments. R.A. and U.V.W. carried out the theoretical calculations. All authors contributed to writing the manuscript. S.R. and T.G. contributed equally to this work. **Competing interests:** K.B., S.R., and T.G. have filed an Indian Patent Application (pending) (201941028467) related to the synthesis and experimental high-temperature thermoelectric properties. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/371/6530/722/suppl/DC1
Materials and Methods
Figs. S1 to S25
Tables S1 to S3
References (*66*–*73*)

18 February 2020; resubmitted 5 November 2020
Accepted 15 December 2020
10.1126/science.abb3517

GEOLOGY

Orogenic quiescence in Earth's middle age

Ming Tang[1]*, Xu Chu[2], Jihua Hao[3,4], Bing Shen[1]

Mountain belts modulate denudation flux and hydrologic processes and are thus fundamental to nutrient cycling on Earth's surface. We used europium anomalies in detrital zircons to track mountain-building processes over Earth's history. We show that the average thickness of active continental crust varied on billion-year time scales, with the thickest crust formed in the Archean and Phanerozoic. By contrast, the Proterozoic witnessed continuously decreasing crustal thickness, leaving the continents devoid of high mountains until the end of the eon. We link this gradually diminished orogenesis to the long-lived Nuna-Rodinia supercontinent, which altered the mantle thermal structure and weakened the continental lithosphere. This prolonged orogenic quiescence may have resulted in a persistent famine in the oceans and stalled life's evolution in Earth's middle age.

Earth's continents have a highly skewed elevation distribution. Most of the continental areas lie close to sea level because of the balance between erosion and deposition (*1*). However, at convergent plate boundaries, an active mountain-building process known as orogenesis generates substantial uplift. These mountainous terrains, though minor by area, profoundly influence global denudation and hydrologic processes on land (*2*–*5*).

Mountain belts owe their high elevations to crustal thickening, which in turn is driven by tectonic compression and magmatic inflation (*6*). Meanwhile, uplift increases erosion efficiency and induces gravitational spreading (*7*), which counters crustal thickening. After magmatism and tectonic compression terminate, erosion and collapse of the orogen prevail, and the crust quickly loses its thickness and elevation as it ages. Thereby, mountains are ephemeral, and the history of mountain building has been constantly erased and overprinted. The thickness of the remnant is not representative of the crust when it formed. This preservation issue poses great challenges to tracking orogenesis in deep time.

Here, we reconstructed mountain-building history using a recently calibrated zircon-based crustal thickness proxy (*8*). Detrital zircons are derived from a variety of crustal rocks and naturally sample large tracts of the continental crust exposed to erosion. Because of their refractory nature, zircons survive most erosion and weathering processes, remain chemically intact, and thus provide a continuous record of magmatic history where fragmented rock records fail (*9*). The zircon-based crustal thickness proxy uses the pressure-sensitive Eu systematics during magmatic differentiation, which is recorded as Eu anomalies (Eu/Eu*, chondrite normalized Eu/$\sqrt{\text{Sm} \times \text{Gd}}$) in crystallizing zircons. We filtered out metamorphic zircons and zircons derived from S-type granites based on zircon Th/U ratios and P contents, respectively (*10*).

Our approach calculates the thickness of all magmatically active crust, which we call the active continental crust. Most of this crust forms at convergent plate margins because of oceanic plate subduction and continent-continent collision. The overriding continental plate in which orogenic magmatisms develop could be either reworked preexisting crust or juvenile crust. In both cases, the resulting mountain belts interact with the surface environment in essentially the same way. Thus, we did not filter our compilation with zircon Hf or O isotope data. The scope of this study and our approach thus differ from a previous study by Dhuime *et al.* (*11*), who used Sr isotopes and Nd model ages to specifically constrain the thickness of juvenile crust.

The reconstructed thickness of active continental crust averaged 50 to 60 km in the Phanerozoic and showed a thickening trend toward the present (Fig. 1A). The thick active continental crust in the Phanerozoic is consistent with the observation that most of the

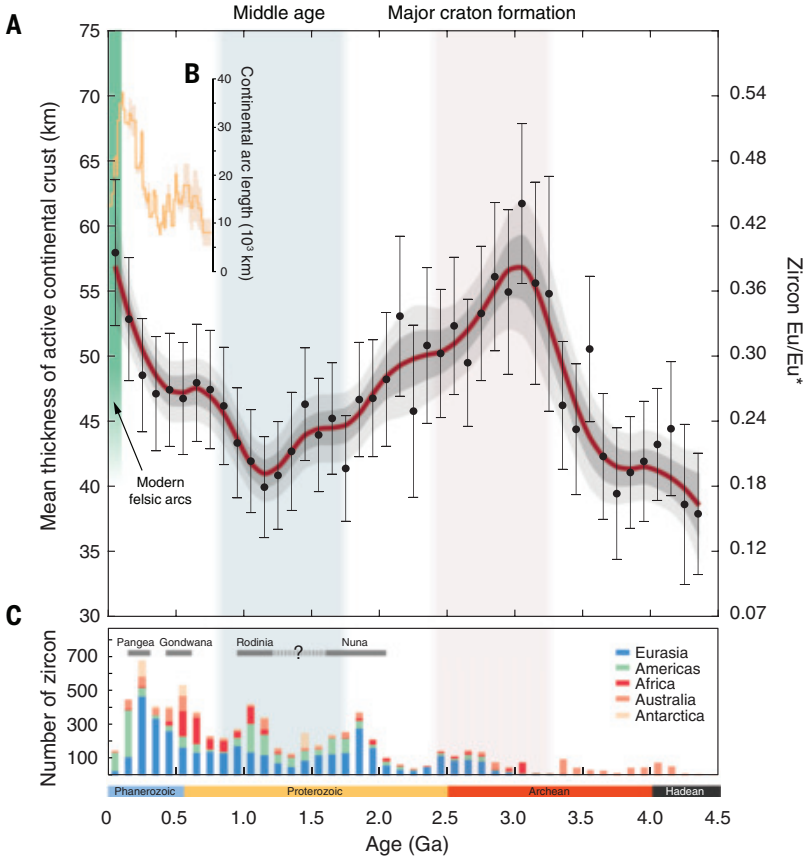


Fig. 1. Reconstructed thickness of active continental crust over Earth's history. (**A**) Reconstruction based on >14,000 analyses of detrital zircons from around the globe (*10*). Data are plotted as binned averages (bin size, 100 million years), with error bars indicating ± 2 SEM. A smoothed trend bracketed by 68 and 95% confidence intervals is shown by the red curve with shaded envelopes. We take zircon crystallization ages as the ages of synmagmatic orogenesis. The time window of major craton formation (pink band) is from (*13*). (**B**) Length of continental arcs in the last 750 million years (means enveloped by uncertainty intervals) is from (*49*). (**C**) Number of detrital zircons from each continent within each 100-million-year bin.

[1]Key Laboratory of Orogenic Belt and Crustal Evolution, MOE; School of Earth and Space Science, Peking University, Beijing 100871, China. [2]Department of Earth Sciences, University of Toronto, Toronto, Ontario M5S 3B1, Canada. [3]Department of Marine and Coastal Sciences, Rutgers University, New Brunswick, NJ 08901, USA. [4]CAS Key Laboratory of Crust-Mantle Materials and Environments, School of Earth and Space Sciences, University of Science and Technology of China, Hefei 230026, China.
*Corresponding author. Email: mingtang@pku.edu.cn

modern felsic arcs are built on crust >40 km (*12*). We also found that the pattern of reconstructed crustal thickness mimics the trend of continental arc length in the last 750 million years (Fig. 1B). Because continental arcs represent the thick end member of active crust, this pattern similarity affirms that the crustal thickness reconstructed from detrital zircons faithfully tracks the overall mountain-building activity in the past.

The Precambrian active continental crust varied substantially in thickness over time (Fig. 1). The active crust became progressively thicker from the Hadean to the Archean and reached a maximum average thickness of 55 to 65 km in the Mesoarchean to Neoarchean [3.2 to 2.5 billion years ago (Ga)]. The emergence of thick felsic crust in the Mesoarchean to Neoarchean coincided with the peak of craton formation (*13*), which may also result from compressive tectonics (*14*, *15*). The lack of thick active crust and thus orogenesis in the Paleoarchean and Hadean (Fig. 1) suggests that lateral plate convergence may have been weak in the first billion years of Earth's history. We note that this speculation is based on a limited number of detrital zircons from this time period and may be tested when additional detrital zircon data become available.

In the Proterozoic, the thickness of active continental crust exhibited a "V"-shaped temporal pattern. Crustal thickness declined continuously from the Paleoproterozoic through the end of the Mesoproterozoic. At 1.3 to 1.0 Ga, the average thickness of active crust may have been as low as 40 km, close to the crustal thickness of the heavily eroded continental interior. This would imply that, on 100-million-year time scales, the continents at that time were far less mountainous than today.

This long-term quiescence in mountain building coincides with a substantial reduction of subduction flux in the Proterozoic (*16*) inferred from Nb/Th of the depleted mantle (*17*) and $^{4}He/^{3}He$ of ocean island basalts (*18*). Reduced orogenesis and subduction flux may be linked to the unusual supercontinent cycles in the Proterozoic. Supercontinents insulate the underlying mantle, and this "blanketing" effect can profoundly alter mantle thermal structure (*19*–*21*). As a consequence, the mantle beneath a supercontinent becomes hotter, whereas the mantle beneath an oceanic domain cools down for a thermal balance (*21*). Cooling increases suboceanic mantle viscosity and thus decreases oceanic plate velocity.

The Proterozoic witnessed two supercontinents, Nuna (Columbia) and Rodinia. Nuna was assembled between 2.1 and 1.8 Ga (*22*) and then broke up between 1.6 and 1.2 Ga, followed by the amalgamation of Rodinia at 1.2 to 0.9 Ga (*22*). However, a growing body of evidence suggests that the breakup of Nuna was limited, and it transitioned to Rodinia with only minor reconfiguration (*23*, *24*). This unconventional transition is also supported by the paucity of passive margins in the Mesoproterozoic (*25*). Therefore, Nuna and Rodinia may be viewed as one largely coherent supercontinent cycle [also known as Nudinia (*26*)] spanning from the late Paleoproterozoic to the early Neoproterozoic (*27*). The mantle thermal structure, altered by the long-lived supercontinent lid, may have led to substantial cooling of the suboceanic mantle to the point that plate tectonics operated intermittently until the breakup of Rodinia.

Prolonged heating of the continental lithosphere may prompt widespread low-pressure–ultra-high-temperature metamorphism and intraplate anhydrous magmatism within the continents (*19*), which are distinctive in the mid-Proterozoic metamorphic (*28*) and igneous (*29*) records. The Grenville orogen, formed between 1250 and 980 million years ago (Ma), has long been regarded as a prototype of the Himalaya-Tibet orogen (*30*). However, the massive anorthosite massifs, dike swarms, peralkaline shoshonite, and ubiquitous A-type granite suites that characterize the Grenville orogen (and the Sveconorwegian orogen) have been rare in Phanerozoic orogens (*27*, *31*). Intense heating could thermally weaken the lithosphere, causing thickened crust to relax rapidly. Detrital zircons of broadly defined Grenvillian ages (1250 to 980 Ma) from North America further show that the Grenville orogen differs from its Phanerozoic counterparts. In southern Tibet and the North American Cordillera, most detrital zircons record 50- to 70-km crustal thicknesses, whereas in the Grenville, crustal thickness peaked at 40 km (Fig. 2).

Long-term quiescence in mountain building may have a profound influence on the hypsography of Earth's surface. The elevation of continent freeboard is determined by a number of factors, including the thickness of the continental crust and oceanic crust, densities of the crust and mantle, and seafloor depth (*32*, *33*). For simplicity, we extrapolate the conditions of isostasy on modern Earth to the past. We found that if the sea level was not substantially different, most of the active continental crust in the mid-Proterozoic would have elevations of no more than 1 to 2 km, in contrast to 3 to 5 km in the succeeding Phanerozoic era (Fig. 3A).

The loss of elevation contrasts between the active continental crust and oceans would substantially reduce erosion rate (*2*) and the intensity of the hydrologic cycle (*34*) on the continents, leading to a subdued weathering flux in this time period. This effect is shown by the similar active continent elevation and seawater Sr isotope curves in the Proterozoic (Fig. 3, A and B).

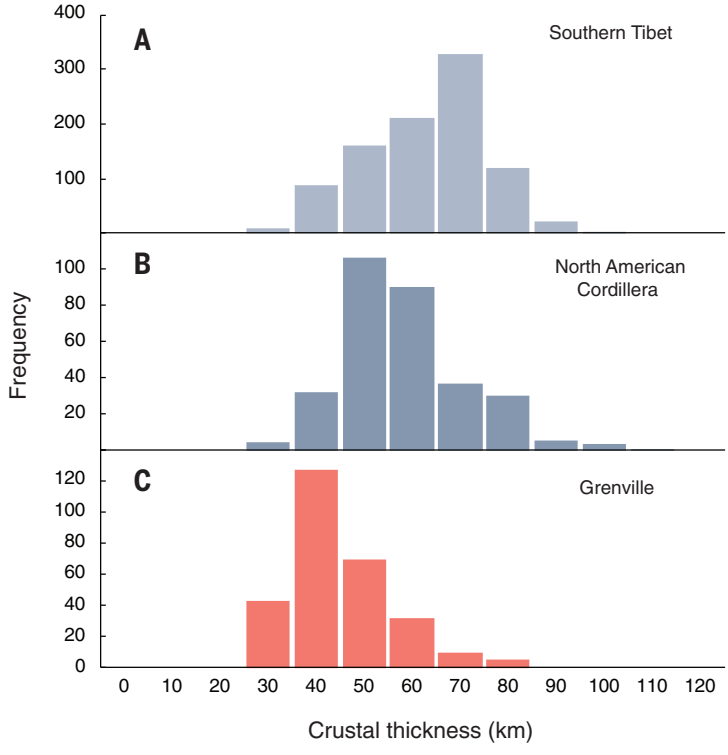


Fig. 2. Reconstructed crustal thicknesses for southern Tibet, North American Cordillera, and North American Grenville. (**A**) Southern Tibet detrital zircon data (100 Ma to present) are from (*8*). (**B** and **C**) We assumed that North American detrital zircon populations at 230 to 50 Ma and 1250 to 980 Ma were primarily sourced from the North American Cordillera (B) and Grenville (C) orogens, respectively.

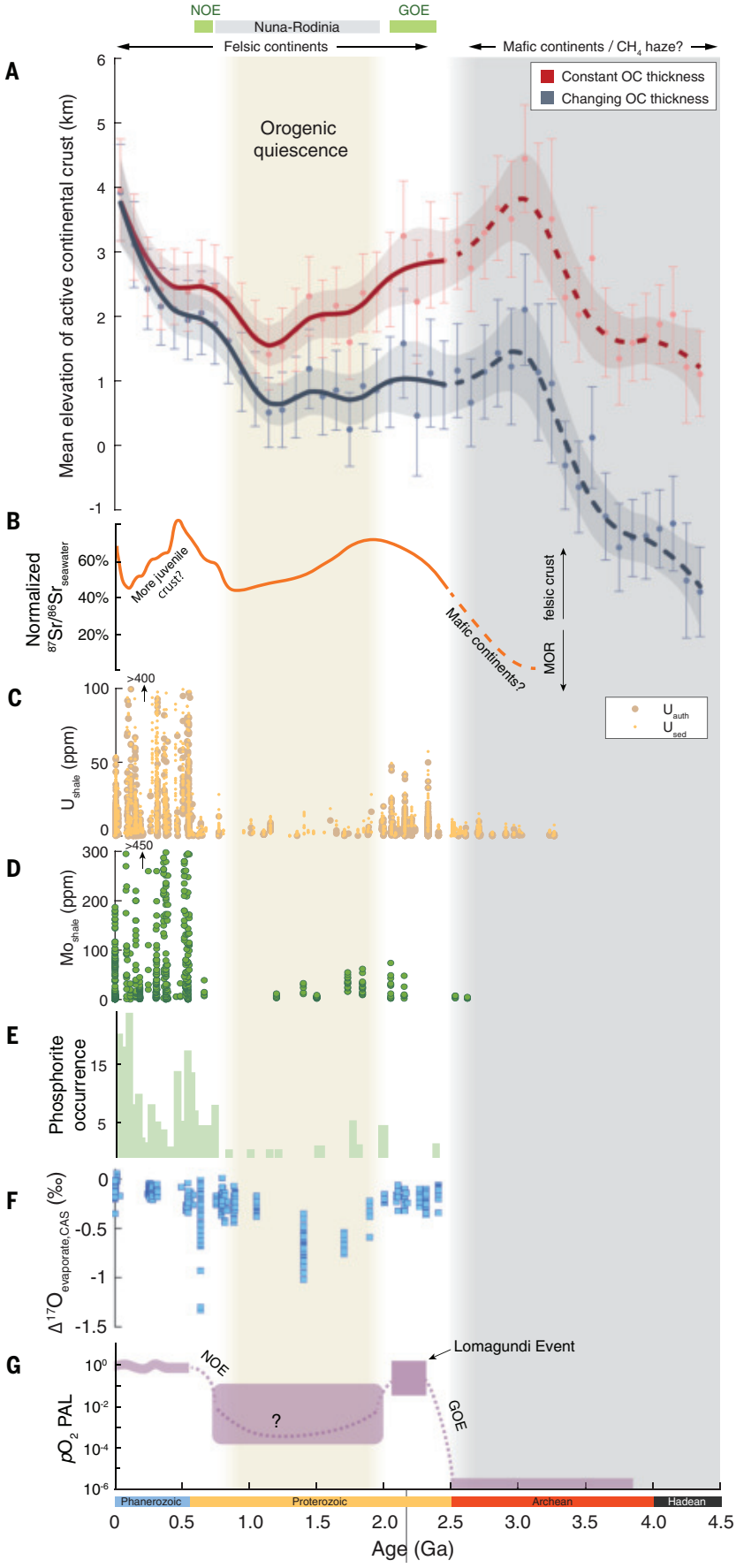


Fig. 3. Elevation of the active continental crust over Earth's history and evolution of Earth's surface environment. (**A**) Elevation plotted as binned averages (bin size, 100 million years), with error bars indicating ± 2 SEM. Also shown are the smoothed trends bracketed by 95% confidence intervals. The elevation of active continental crust can be calculated from our reconstructed crustal thickness using an isostasy model (*10*). We consider two end-member scenarios of oceanic crust thickness, one with constant thickness over time and the other decreasing thickness from the Hadean to the present day (*10*). (**B**) The normalized seawater Sr isotope curve is from (*50*), which removes the radiogenic decay effect and thus reflects the contributions from the midocean ridge (MOR) and the felsic continental crust. The Sr isotope curve has decoupled from that of the elevation since ~450 Ma, which was probably caused by an increasing contribution from juvenile crust with MOR-like Sr isotopes (*51*). (**C**) Sediment U contents and authigenic U concentrations in black shales (*42*). (**D**) Mo concentrations in black shales (*36*). (**E**) Sedimentary phosphorite occurrence (*39*). (**F**) $\Delta^{17}O$ in evaporate and carbonate-associated sulfate (*41*). (**G**) Atmospheric O_2 partial pressure relative to present atmospheric level (PAL). The purple fields are from (*52*) and (*53*). The dashed curve in the Proterozoic field is our proposed path (schematic). NOE, Neoproterozoic Oxidation Event; GOE, Great Oxidation Event.

With muted continental weathering, nutrient supply to the oceans declines. The scarcity of P, Mo, and other trace metals would dampen primary productivity and reduce O_2 production (*35*), which would in turn further decrease the Mo flux from the continents (*36*) and enhance Mo and P removal from the Proterozoic oceans (*36*–*38*).

Eventually, a widespread famine followed by a collapse of primary productivity may have occurred in the Proterozoic oceans. This is reflected by the extremely low Mo in black shales, disappearance of sedimentary phosphorite, and strongly negative $\Delta^{17}O$ evaporates and carbonate-associated sulfate between 1.8 and 0.8 Ga (*36*, *39*–*41*) (Fig. 3, D to F). The systematic decline in atmospheric O_2 into the mid-Proterozoic is corroborated by the decreasing U concentrations in black shales (*42*) (Fig. 3C) and the fall of seawater sulfate level (*43*, *44*) after the initial rise of atmospheric O_2 between 2.5 and 2.0 Ga, the Great Oxidation Event. Biological evolution may have been largely stalled during this 1-billion-year orogenic quiescence, a time period often referred to as the "boring billion" (*45*, *46*).

The middle-age orogenic quiescence came to an end in the Neoproterozoic (Fig. 1A), a time corresponding to the termination of the long-lived Nuna-Rodinia supercontinent. The

breakup of Nuna-Rodinia may have relaxed the thermal contrast between the suboceanic and subcontinental mantle and established modern-style plate tectonics. As mountains reappeared on the continents, nutrient supply to the oceans was enhanced, which catalyzed surges in biological productivity and resumed surface oxidation. Efficient orogenesis appears to have been maintained ever since (Fig. 1). The sustained high erosion and weathering rates promoted organic C burial, as evidenced by a systematic ^{13}C enrichment in Phanerozoic carbonates (*47*). With the emergence of a fully oxidized atmosphere-ocean system, the planet was eventually primed for the arrival of metazoans in the Cambrian (*48*).

REFERENCES AND NOTES

1. D. B. Rowley, *J. Geol.* **121**, 445–454 (2013).
2. I. J. Larsen, D. R. Montgomery, H. M. Greenberg, *Geology* **42**, 527–530 (2014).
3. J. D. Milliman, K. L. Farnsworth, *River discharge to the coastal ocean: A global synthesis*. (Cambridge Univ. Press, 2013).
4. L. R. Kump, S. L. Brantley, M. A. Arthur, *Annu. Rev. Earth Planet. Sci.* **28**, 611–667 (2000).
5. A. J. West, A. Galy, M. Bickle, *Earth Planet. Sci. Lett.* **235**, 211–228 (2005).
6. C.-T. A. Lee, S. Thurner, S. Paterson, W. Cao, *Earth Planet. Sci. Lett.* **425**, 105–119 (2015).
7. C.-T. A. Lee, in *Treatise on Geochemistry*, H. D. Holland, K. K. Turekian, Eds. (Elsevier, ed. 2, 2014), vol. 4, pp. 423–456.
8. M. Tang, W.-Q. Ji, X. Chu, A. Wu, C. Chen, *Geology* **49**, 76–80 (2021).
9. C. J. Hawkesworth *et al*., *J. Geol. Soc. London* **167**, 229–248 (2010).
10. Materials and methods are available as supplementary materials.
11. B. Dhuime, A. Wuestefeld, C. J. Hawkesworth, *Nat. Geosci.* **8**, 552–555 (2015).
12. M. J. Farner, C.-T. A. Lee, *Earth Planet. Sci. Lett.* **470**, 96–107 (2017).
13. D. G. Pearson, N. Wittig, in *Treatise on Geochemistry*, H. D. Holland, K. K. Turekian, Eds. (Elsevier, ed. 2, 2014), vol. 3, pp. 255–292.
14. H. Wang, J. van Hunen, D. G. Pearson, *Tectonophysics* **746**, 562–571 (2018).
15. M. Tang, C.-T. A. Lee, R. L. Rudnick, K. C. Condie, *Geochim. Cosmochim. Acta* **278**, 6–15 (2020).
16. P. G. Silver, M. D. Behn, *Science* **319**, 85–88 (2008).
17. K. D. Collerson, B. S. Kamber, *Science* **283**, 1519–1522 (1999).
18. S. W. Parman, *Nature* **446**, 900–903 (2007).
19. N. Coltice, B. R. Phillips, H. Bertrand, Y. Ricard, P. Rey, *Geology* **35**, 391–394 (2007).
20. A. Lenardic, *Nat. Geosci.* **10**, 4–5 (2017).
21. A. Lenardic *et al*., *Geochem. Geophys. Geosyst.* **12**, Q10016 (2011).
22. G. Zhao, P. A. Cawood, S. A. Wilde, M. Sun, *Earth Sci. Rev.* **59**, 125–162 (2002).
23. N. M. W. Roberts, *Geosci. Front.* **4**, 681–691 (2013).
24. P. A. Cawood, R. A. Strachan, S. A. Pisarevsky, D. P. Gladkochub, J. B. Murphy, *Earth Planet. Sci. Lett.* **449**, 118–126 (2016).
25. D. C. Bradley, *Earth Sci. Rev.* **91**, 1–26 (2008).
26. P. A. Cawood, *Geology* **48**, 525–526 (2020).
27. P. A. Cawood, C. J. Hawkesworth, *Geology* **42**, 503–506 (2014).
28. M. Brown, C. L. Kirkland, T. E. Johnson, *Geology* **48**, 488–492 (2020).
29. C. Liu, A. H. Knoll, R. M. Hazen, *Nat. Commun.* **8**, 1950 (2017).
30. T. Rivers, *Precambrian Res.* **167**, 237–259 (2008).
31. T. Rivers, N. Culshaw, A. Hynes, J. Martignole, in *Tectonic Styles in Canada: The LITHOPROBE Perspective*, J. A. Percival, F. A. Cook, R. M. Clowes, Eds. (Geological Association of Canada, Special Paper 49, 2012), pp. 97–236.
32. C.-T. A. Lee *et al*., *Int. Geol. Rev.* **60**, 431–448 (2018).
33. J. Korenaga, N. J. Planavsky, D. A. D. Evans, *Philos. Trans. A Math Phys. Eng. Sci.* **375**, 20150393 (2017).
34. D. Viviroli, R. Weingartner, B. Messerli, *Mt. Res. Dev.* **23**, 32–40 (2003).
35. A. L. Zerkle, C. H. House, R. Cox, D. E. Canfield, *Geobiology* **4**, 285–297 (2006).
36. C. Scott *et al*., *Nature* **452**, 456–459 (2008).
37. M. A. Kipp, E. E. Stüeken, *Sci. Adv.* **3**, eaao4795 (2017).
38. T. A. Laakso, E. A. Sperling, D. T. Johnston, A. H. Knoll, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 11961–11967 (2020).
39. N. J. Planavsky, *Nat. Geosci.* **7**, 855–856 (2014).
40. P. W. Crockford *et al*., *Nature* **559**, 613–616 (2018).
41. P. W. Crockford *et al*., *Chem. Geol.* **513**, 200–225 (2019).
42. C. Partin *et al*., *Earth Planet. Sci. Lett.* **369–370**, 284–293 (2013).
43. S. Schröder, A. Bekker, N. Beukes, H. Strauss, H. Van Niekerk, *Terra Nova* **20**, 108–117 (2008).
44. M. Fakhraee, O. Hancisse, D. E. Canfield, S. A. Crowe, S. Katsev, *Nat. Geosci.* **12**, 375–380 (2019).
45. M. D. Brasier, J. F. Lindsay, *Geology* **26**, 555–558 (1998).
46. H. D. Holland, *Philos. Trans. R. Soc. London B Biol. Sci.* **361**, 903–915 (2006).
47. J. Krissansen-Totton, R. Buick, D. C. Catling, *Am. J. Sci.* **315**, 275–316 (2015).
48. C. R. Marshall, *Annu. Rev. Earth Planet. Sci.* **34**, 355–384 (2006).
49. W. Cao, C.-T. A. Lee, J. S. Lackey, *Earth Planet. Sci. Lett.* **461**, 85–95 (2017).
50. G. A. Shields, *eEarth* **2**, 35–42 (2007).
51. C. P. Bataille, A. Willis, X. Yang, X.-M. Liu, *Sci. Adv.* **3**, e1602183 (2017).
52. T. W. Lyons, C. T. Reinhard, N. J. Planavsky, *Nature* **506**, 307–315 (2014).
53. A. P. Martin, D. J. Condon, A. R. Prave, A. Lepland, *Earth Sci. Rev.* **127**, 242–261 (2013).

ACKNOWLEDGMENTS

We thank R. L. Rudnick and C.-T. A. Lee, whose comments and suggestions greatly improved the early versions of the manuscript; R. McKenzie, Y. Shen, M. Li, and P. Liu for insightful discussions; and E. J. Chin, P. A. Cawood, and an anonymous reviewer for constructive comments. **Funding:** This work was supported by the National Natural Science Foundation of China (grants 41888101, 91755000, and 42073026). M.T. was supported by a Beijing "Double-First Class" initiative grant (7101302526) and a starting fund from Peking University. J.H. gratefully acknowledges a NASA Astrobiology Institute grant (80NSSC18M0093). **Author contributions:** M.T. initiated the idea and wrote the first version of the manuscript. All authors contributed to interpreting data and revising the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/371/6530/728/suppl/DC1
Materials and Methods
Figs. S1 and S2
References (*54*–*60*)
Database 1
MDAR Reproducibility Checklist

9 October 2020; accepted 5 January 2021
10.1126/science.abf1876

OCEANOGRAPHY

Seismic crustal imaging using fin whale songs

Václav M. Kuna[1,2]* and John L. Nábělek[1]

Fin whale calls are among the strongest animal vocalizations that are detectable over great distances in the oceans. We analyze fin whale songs recorded at ocean-bottom seismometers in the northeast Pacific Ocean and show that in addition to the waterborne signal, the song recordings also contain signals reflected and refracted from crustal interfaces beneath the stations. With these data, we constrain the thickness and seismic velocity of the oceanic sediment and basaltic basement and the *P*-wave velocity of the gabbroic lower crust beneath and around the ocean bottom seismic stations. The abundant and globally available fin whale calls may be used to complement seismic studies in situations where conventional air-gun surveys are not available.

The fin whale (*Balaenoptera physalus*) is an endangered (*1*) cetacean that populates global oceans from polar regions to equatorial waters (*2*), with an estimated population of 100,000 (*3*). Fin whale vocalizations are among the strongest animal calls in the ocean (*4*–*6*). Reaching up to 189 dB (1 μPa at 1 m) (*5*, *6*), the source levels are comparable to those of noise produced by large ships (*4*). The calls can be monitored hundreds of miles away from the source (*4*), enabling studies of whale behavior, abundance, distribution, and migration patterns (*7*–*13*).

Fin whale vocalizations include short, 1-s calls with a dominant frequency of around 20 Hz (*14*). Calls are characterized by a sinusoidal, downward frequency-sweeping signal, with about a 5-Hz drop in frequency over the duration of the call (*7*). These calls repeat every 7 to 40 s (*7*, *15*, *16*), forming songs that last up to tens of hours, with short interruptions about every 15 min when the whale surfaces (*7*) (fig. S1). The songs may be composed of a single repeating call but often consist of two or three alternating calls with different downward-sweeping character (*16*) (Fig. 1, A and B, and figs. S2 and S3). Ocean-bottom seismometer (OBS) stations deployed for earthquake monitoring are commonly configured to record vibrations of 50 Hz and lower and often capture vocalizations of whales with low-frequency calls, including those of fin whales (*8*, *11*–*13*). We show that the fin whale calls recorded at OBS stations contain, in addition to waterborne energy, signals reflected and refracted from crustal interfaces beneath the station

[1]College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR 97331, USA. [2]Institute of Geophysics of the CAS, 141 31 Prague, Czech Republic.
*Corresponding author. Email: kuna@ig.cas.cz

that can be used for seismic imaging of the oceanic crust.

From 2012 to 2013, a network of 54 OBS stations was deployed in the northeast Pacific Ocean to monitor the seismicity of the Blanco transform fault (*17*, *18*). Besides fault seismicity, OBS stations recorded numerous fin whale songs (fig. S4), which often lasted up to 10 hours. We analyzed fin whale songs recorded at the BS080, BS100, and BS110 stations of the OBS network (Fig. 2). The stations were deployed 20 to 40 km north of the Blanco fault at a depth of about 3000 m, in a flat region without a substantial variation in bathymetry. The sites were 270 to 330 km from the Juan de Fuca spreading ridge, which corresponds to a crustal age of about 5 million to 7 million years (*19*). The three OBS stations were equipped with short-period, three-component seismic sensors and differential pressure gauges, recording data streams at the rate of 100 samples per second (*20*).

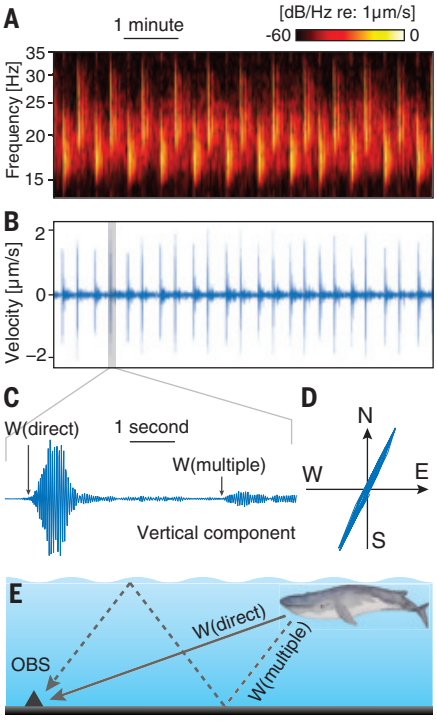


Fig. 1. Characteristics of fin whale vocalizations. (**A**) Spectrogram of a 7-min section of an unfiltered fin whale song, recorded by the vertical seismometer component at the station BS080 (November 2012 whale song; see Fig. 2 for location). re, relative to. (**B**) Song section in the time domain. (**C**) Magnification of a single fin whale call. Waterborne wave arrivals are denoted with arrows. (**D**) An example of a direct-wave arrival recorded at the station's horizonal components, showing polarization along the azimuth to the whale. (**E**) A sketch illustrating travel paths of waterborne phases used for estimating the distance to the whale.

We analyzed a total of six songs, each formed by vocalizations of a single whale. Two were recorded in October and November 2012 at BS080, two in September 2012 and September 2013 at BS100, and two in September 2012 and May 2013 at BS110 (Fig. 2, fig. S5, and table S1). The song durations range from 2.5 to 4.9 hours, during which time the whales covered distances between 16 and 38 km with mean cruising speeds between 2.2 and 5.6 knots (4.1 to 10.3 km hour^{-1}). The whales traveled on oblique paths around the stations, swimming from as far as 19 km from the stations to as close as 700 m. The songs primarily consisted of calls of two distinct types: a downward-sweeping call with dominant frequencies between 25 and 20 Hz and a lower-pitch call that dropped in frequency from 20 to 15 Hz (Fig. 1A and figs. S2 and S3). Owing to lower background noise levels in the higher-frequency range, we used only the higher-pitch calls. The songs contained 212 to 593 calls, with intercall times between 30 and 40 s (fig. S5B). To enhance the useful signal, we filtered the data using a tapered Butterworth bandpass filter with a passband from 21 to 25 Hz.

We determined the distance between the whale (i.e., the source) and the station from the difference in arrival times of the direct water wave and the waterborne multiple, which is the reflected water wave that bounces between the seafloor and the sea surface before reaching the station [Fig. 1, C and E; (*20*)]. We assumed that the whales swam at a depth of 10 m, given that fin whales typically call while swimming at depths between 0 and 20 m (*21*). With a depth range of 10 ± 10 m, the estimated uncertainty of the source-station distance at 20 km is 30 m. To find the actual whale swim path, we calculated the whale location by combining the distance and the azimuth between the whale and the station. We determined the station-source azimuth from the directivity of the direct water wave that aims toward the source [Fig. 1D; (*20*)]. The final travel path was smoothed with a moving average filter based on reasonable assumptions that the whale trajectory is expected to be relatively smooth and the cruising velocity steady. Changing seafloor bathymetry may influence the source-station distance determination, and we correct for this effect by using an iterative location method (*20*). However,

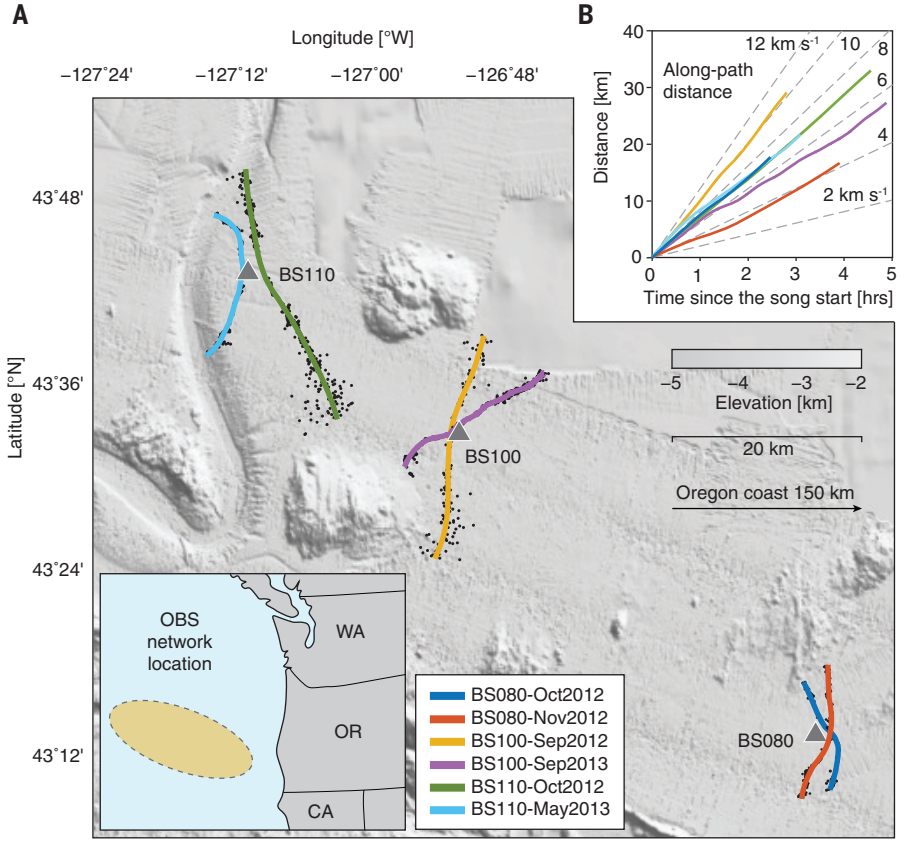


Fig. 2. Analyzed whale songs and travel paths. (**A**) Seafloor bathymetry, station location, and the whale travel path during the analyzed songs. Black dots denote location of individual calls, colored lines are smoothed whale travel paths, and gray triangles represent stations. The orientation map on the bottom left shows the OBS network location in the northeast Pacific. (**B**) Source-station offset plotted as a function of time elapsed since the beginning of the song.

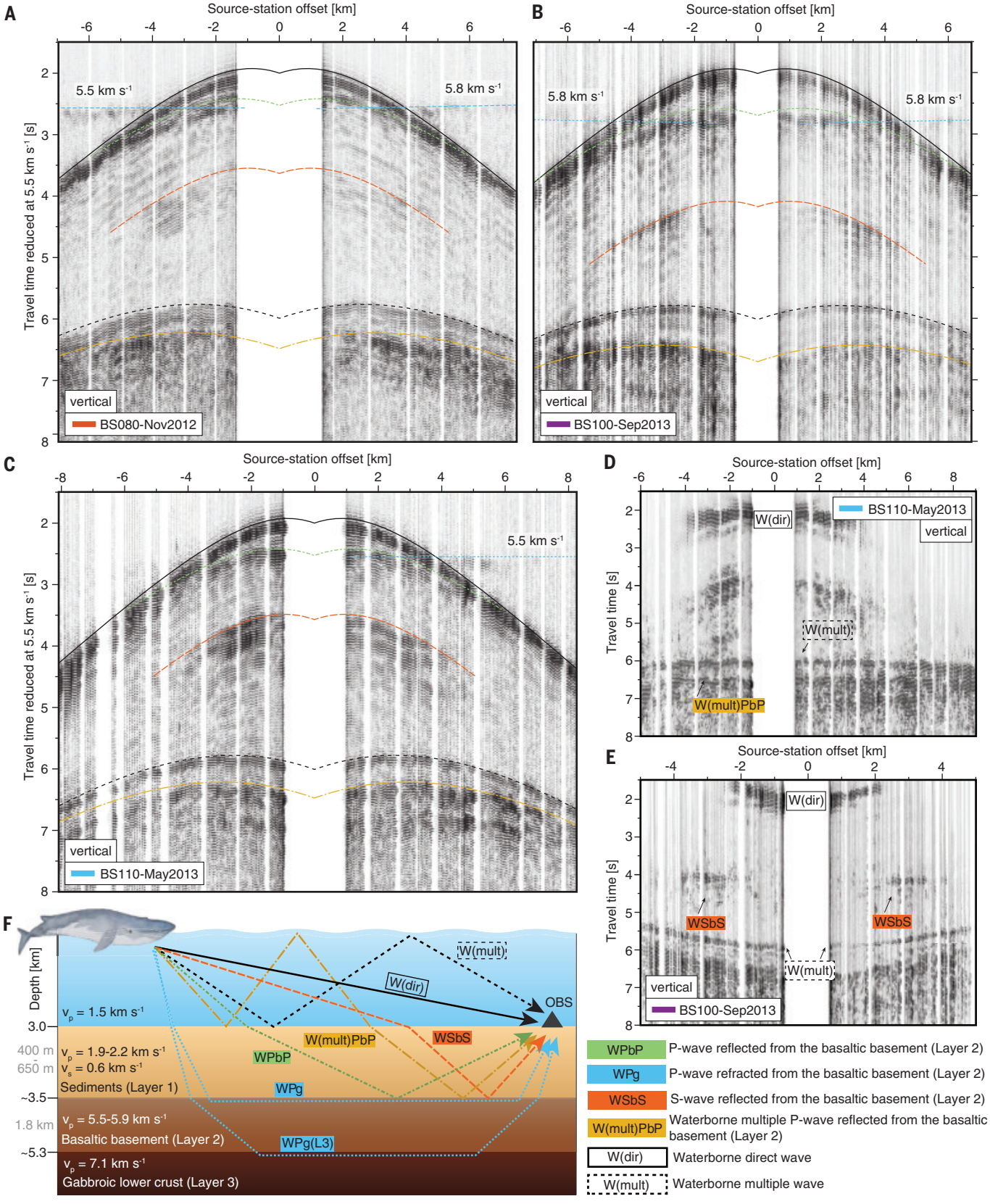


Fig. 3. Interpreted song gathers from stations BS080, BS100, and BS110. (**A** to **C**) Song gathers with identified phase travel times (A) BS080-Nov2012, (B) BS100-Sep2013, and (C) BS110-Oct2012. The theoretical phase travel times (colored lines) are based on the seismic velocities and layer thicknesses given in the main text and Table 1. (**D** and **E**) Optimal NMO correction that flattens the (D) W(mult)PbP phase arrival at BS110-May2013 and (E) WSbS phase arrival at BS100-Sep2013. (**F**) Model of the upper oceanic crust, with seismic velocities and ray paths. The model depth is not to scale.

the bathymetry around the studied stations is generally very flat.

When a fin whale call impinges on the ocean bottom, part of the call energy is transmitted in the ground as a seismic wave. The seismic wave travels through the oceanic crust, where it is reflected and refracted by layers within the crust. When recorded at an OBS station, it enables us to estimate the crustal structure. To show arrivals of individual seismic phases, we organized the whale calls into the receiver gathers. A receiver gather represents a collection of individual whale calls (i.e., seismic traces) displayed according to the distance between the whale and the station at the time of the call (i.e., source-station offset). Each of the receiver gathers comprises whale calls recorded during one song, which we refer to as a song gather. Individual phase arrivals appear in song gathers as intervals of a relatively high-amplitude signal. With increasing distance, the phases arrive progressively later, a process termed normal moveout (NMO). The accurate medium velocity corrects for the NMO by making phase arrival times equivalent at all distances, which can be used to estimate medium velocity. The seismic phase arrivals are best displayed in song gathers from the vertical seismometer component (Fig. 3). The most prominent arrivals in song gathers are the waterborne waves—the direct wave W(dir) and the reflected (multiple) wave W(mult). In addition to the waterborne waves, we identified four dominant seismic phases belonging to wave reflections and refractions in the basaltic basement (i.e., layer 2) and refractions in the gabbroic lower crust (i.e., layer 3). The best-fitting crustal model for each station (Table 1) was found by regression of arrival times of the individual phases and modeled using a ray-modeling software (*20*, *22*). Despite near-horizontal bathymetry, we observe signal fluctuations and differences in the appearance of waterborne and seismic phases in the individual song gathers at horizontal scales of 200 to 500 m. These differences could be due to scattering from ripples at the sediment surface caused by ocean-bottom currents, heterogeneity in the sediment layering, and rough topography on layer 2.

P-wave reflections from the top of the basement (layer 2) can be identified after the arrival of the direct waterborne wave W(dir) or the waterborne multiple W(mult). The WPbP corresponds to the direct wave reflection and the W(mult)PbP to the waterborne multiple reflection. The basement reflections enable us to calculate the thickness and the *P*-wave velocity of the oceanic sediment, layer 1. When the direct water wave impinges obliquely on the interface between the water and the sediment layer, part of the wave energy is converted to an *S* wave. *S*-wave reflections from layer 2 (WSbS) arrive in the song gather between the waterborne wave arrivals (Fig. 3 and figs. S6 to S8). The WSbS provides an independent estimate of the sediment thickness and constrains the *S*-wave velocity and the Poisson ratio of layer 1. At offsets greater than about 4 km, the *P* wave refracted along the basaltic basement (referred to as WPg) becomes the first arriving phase and constrains the *P*-wave velocity of layer 2 as well as the layer 1 thickness. At offsets greater than about 12 km, the velocity of the WPg phase increases, which indicates the arrival of a *P* wave refracted along the gabbroic lower crust (layer 3). The WPg(L3) enables us to calculate the layer 2 thickness and the *P*-wave velocity of layer 3.

The basaltic basement reflections WPbP and W(mult)PbP are observed at all stations, arriving about 0.5 to 0.7 s after the respective waterborne waves at 2-km source-station offsets (figs. S6 to S8). For longer offsets, the time difference between the reflected phases and the waterborne wave arrivals decreases, indicating faster NMO velocity of the reflected phase arrival. The optimal NMO correction of the W(mult)PbP phase uses a velocity between 1.53 and 1.56 km s^{-1}, which indicates that the *P*-wave velocity in the sediment layer is between 1.9 and 2.2 km s^{-1} (Fig. 3D and fig. S9). The difference between the phase travel time of W(mult)PbP and W(mult) suggests a total sediment thickness of about 500, 650, and 400 m at BS080, BS100, and BS110, respectively, with an estimated uncertainty of about 30 to 40 m (Table 1 and figs. S10 to S12).

Appearing at short offsets, about 1.5 to 2 s after the W(dir) at 2 km, the *S*-wave basement reflection WSbS is a dominant phase on all song gathers. The NMO of the reflected WSbS phase (Fig. 3E) suggests that the *S*-wave velocity of the oceanic sediment layer is roughly 0.6 km s^{-1}. The sediment Poisson ratio is in the range of 0.44 to 0.46, and the ratio of *P*- to *S*-wave velocity (v_p/v_s) ratio is between 3.2 and 3.7. The WSbS phase arrivals are consistent with the sediment thickness determined from the W(mult)PbP phase.

The WPg phase, representing the *P* wave refracted along the basaltic basement (layer 2), indicates a basaltic basement velocity between 5.5 and 5.9 km s^{-1} at all stations (figs. S10 to S12), with an uncertainty of about 0.2 km s^{-1}. Differing slopes of WPg recorded at song gathers at opposite sides of a station indicate that the top of layer 2 may be dipping. The effect of the offset uncertainty on the model that results from the uncertainty in the whale dive depth is negligible (~0.2%). The BS110-Oct2012 song gather displays WPg arrival up to 19-km offset south of the station (figs. S8 and S12). At offsets larger than 12 km, the phase arrivals suggest that velocity increases to roughly 7.1 ± 0.1 km s^{-1}, which is in agreement with the oceanic layer 3. With this result, the estimated thickness of layer 2 is 1.8 ±

Table 1. Crustal seismic velocities and layer thicknesses with 95% uncertainty intervals estimated from individual song gathers. A "–" indicates parameters that could not be estimated from the song gather.

		Layer 1			Layer 2			Layer 3
		Seismic velocity (km s^{-1})			*P*-wave velocity (km s^{-1})			
OBS station	Song ID	*P*-wave velocity based on W(mult)PbP NMO correction	*S*-wave velocity based on WSbS NMO correction	Layer thickness (m)	Based on WPg - north arrival time	Based on WPg - south arrival time	Layer thickness (km) based on WPg(L3) intercept time	*P*-wave velocity (km s^{-1}) based on WPg(L3) arrival time
BS080	BS080-Oct2012	2.2 ± 0.4	0.6 ± 0.1	490 ± 40	–	–	–	–
	BS080-Nov2012	2.1 ± 0.1	–	510 ± 130	5.5 ± 0.2	5.8 ± 0.2	–	–
BS100	BS100-Sep2012	2.0 ± 0.1	–	660 ± 30	–	–	–	–
	BS100-Sep2013	2.1 ± 0.1	0.6 ± 0.1	650 ± 60	–	–	–	–
BS110	BS110-Oct2012	2.0 ± 0.2	–	380 ± 60	5.9 ± 0.1	–	1.8 ± 0.3	7.1 ± 0.1
	BS110-May2013	1.9 ± 0.1	0.6 ± 0.1	400 ± 30	5.5 ± 0.3	–	–	–
Overall	–	1.9 to 2.2	0.6	380 to 650	5.5 to 5.9		1.8 ± 0.3	7.1 ± 0.1

0.3 km. The station BS110 (fig. S8) shows particularly well the direct, reflected, and converted phases that constrain the derived crustal model.

Our results with uncertainties conform to the archetypal structure of the oceanic crust, where the gabbroic lower crust (layer 3) is overlain by extrusive basalts (layer 2) and unconsolidated, hydrated sediments (layer 1) (*23*). A seismic image derived from a controlled-source transect about 200 km north of the studied site (*24*) indicates a sediment thickness of roughly 700 to 800 m. At the Blanco fault zone sites, the sediments are 100 to 300 m thinner, which may be due to the sites' greater distance from the sediment sources in the Columbia River estuary and the Juan de Fuca Strait. Our estimates of the sediment v_p/v_s ratio are lower than those suggested by the controlled-source experiment (*25*) farther north, indicating sediment over-compaction, which may have occurred because of the proximity to the Blanco fault and repeated exposure to strong earthquake shaking (*26*). The basaltic basement (layer 2) thickness and the *P*-wave velocity match values observed for this layer of the same crustal age, and the *P*-wave velocity of layer 3 falls within the globally observed range (*23*).

The three studied sites show consistent results, congruent with regional and global observations, and demonstrate that fin whale calls can be used for seismic imaging of the oceanic crust. This method provides lower resolution than air-gun surveys, owing to the narrower frequency band and lower dominant frequency of fin whale calls. Higher-pitch whale vocalizations with a broader frequency band, such as those of sperm whales, could be used for high-resolution studies of the ocean-floor sediments. The source location procedure (*20*) used here limits the application to relatively flat regions without substantial bathymetrical features. Regions of high topography and crustal heterogeneity would require a more sophisticated three-dimensional approach and source location based on multiple seismic stations. We suggest that machine-learning algorithms, such as deep neural networks, could be used for automatic whale location and song processing.

Although fin whale vocalizations do not have the potential to replace high-energy broadband air-gun signals in ocean crust studies, they may complement passive seismic experiments during deployments of ocean-bottom seismic stations. More generally, our study demonstrates that animal vocalizations are useful not only for studying the animals themselves but also for investigating the environment that they inhabit.

REFERENCES AND NOTES

1. J. V. Carretta *et al*., "U.S. Pacific Marine Mammal Stock Assessments: 2013" (NOAA Technical Memorandum NMFS, NOAA-TM-NMFS-SWFSC-532, NOAA, 2014).
2. A. Aguilar, in *Encyclopedia of Marine Mammals*, W. F. Perrin, Ed. (Elsevier, 2008), pp. 433–436.
3. J. G. Cooke, *Balaenoptera physalus*. The IUCN Red List of Threatened Species 2018: e.T2478A50349982 (2019); https://dx.doi.org/10.2305/IUCN.UK.2018-2.RLTS.T2478A50349982.en.
4. W. J. Richardson, C. R. Greene Jr., C. I. Malme, D. H. Thomson, *Marine Mammals and Noise* (Academic Press, 2013).
5. A. Širović, J. A. Hildebrand, S. M. Wiggins, *J. Acoust. Soc. Am.* **122**, 1208–1215 (2007).
6. M. J. Weirathmueller, W. S. D. Wilcock, D. C. Soule, *J. Acoust. Soc. Am.* **133**, 741–749 (2013).
7. W. A. Watkins, P. Tyack, K. E. Moore, J. E. Bird, *J. Acoust. Soc. Am.* **82**, 1901–1912 (1987).
8. M. A. McDonald, J. A. Hildebrand, S. C. Webb, *J. Acoust. Soc. Am.* **98**, 712–721 (1995).
9. S. A. Mizroch, D. W. Rice, D. Zwiefelhofer, J. Waite, W. L. Perryman, *Mammal Rev.* **39**, 193–227 (2009).
10. M. Simon, K. M. Stafford, K. Beedholm, C. M. Lee, P. T. Madsen, *J. Acoust. Soc. Am.* **128**, 3200–3210 (2010).
11. W. S. D. Wilcock, *J. Acoust. Soc. Am.* **132**, 2408–2419 (2012).
12. R. Iwase, *Jpn. J. Appl. Phys.* **54**, 07HG03 (2015).
13. O. Gaspà Rebull, J. Díaz Cusí, M. Ruiz Fernández, J. Gallart Muset, *J. Acoust. Soc. Am.* **120**, 2077–2085 (2006).
14. W. A. Watkins, *Sci. Rep. Whales Res. Inst.* **33**, 83–117 (1981).
15. D. C. Soule, W. S. Wilcock, *J. Acoust. Soc. Am.* **133**, 1751–1761 (2013).
16. M. J. Weirathmueller *et al*., *PLOS ONE* **12**, e0186127 (2017).
17. J. L. Nábělek, J. Braunmiller, Plate boundary evolution and physics at an oceanic transform fault system. International Federation of Digital Seismograph Networks (2012); https://doi.org/10.7914/SN/X9_2012.
18. V. M. Kuna, J. L. Nábělek, J. Braunmiller, *Nat. Geosci.* **12**, 138–142 (2019).
19. D. S. Wilson, *J. Geophys. Res. Solid Earth* **98**, 16053–16071 (1993).
20. Materials and methods are available as supplementary materials.
21. A. K. Stimpert *et al*., *Anim. Biotelem.* **3**, 23 (2015).
22. C. A. Zelt, R. B. Smith, *Geophys. J. Int.* **108**, 16–34 (1992).
23. G. L. Christeson, J. A. Goff, R. S. Reece, *Rev. Geophys.* **57**, 504–529 (2019).
24. G. Horning *et al*., *J. Geophys. Res. Solid Earth* **121**, 5859–5879 (2016).
25. J. Zhu *et al*., *J. Geophys. Res. Solid Earth* **125**, e2019JB019239 (2020).
26. D. E. Sawyer, J. R. DeVore, *Geophys. Res. Lett.* **42**, (2015).
27. J. L. Nábělek, R/V Melville, cruise MV1212. R2R (2012); http://dx.doi.org/10.7284/902871.

ACKNOWLEDGMENTS

We thank K. Davenport for assistance with data processing, T. Kudrnova for the whale illustration used in Figs. 1 and 3, and S. Gyles for reviewing the manuscript. Comments by A. Trehu and J. Braunmiller improved the rigor of the paper. **Funding:** The seismic stations for the project were provided by the U.S. Ocean Bottom Seismograph Instrument Pool (www.obsip.org), funded by the National Science Foundation (NSF). This research was supported by NSF grants OCE-1031858 and OCE-1131767. **Author contributions:** J.L.N. collected the dataset. V.M.K. conducted the data analysis. Both authors discussed the results and contributed to writing the manuscript. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** The OBS dataset is archived at the IRIS Data Management System (www.iris.edu) under network code X9 for the Plate Boundary Evolution and Physics at an Oceanic Transform Fault System project (*17*). Raw bathymetry data, used in Fig. 2 and for whale location, are available from the Rolling Deck to Repository (R2R), R/V Melville, cruise MV1212 (*27*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/371/6530/731/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S12
Table S1
References (*28*–*30*)

23 October 2020; accepted 5 January 2021
10.1126/science.abf3962

CORONAVIRUS

Mosaic nanoparticles elicit cross-reactive immune responses to zoonotic coronaviruses in mice

Alexander A. Cohen[1], Priyanthi N. P. Gnanapragasam[1], Yu E. Lee[1], Pauline R. Hoffman[1], Susan Ou[1], Leesa M. Kakutani[1], Jennifer R. Keeffe[1], Hung-Jen Wu[2], Mark Howarth[2], Anthony P. West[1], Christopher O. Barnes[1], Michel C. Nussenzweig[3], Pamela J. Bjorkman[1]*

Protection against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and SARS-related emergent zoonotic coronaviruses is urgently needed. We made homotypic nanoparticles displaying the receptor binding domain (RBD) of SARS-CoV-2 or co-displaying SARS-CoV-2 RBD along with RBDs from animal betacoronaviruses that represent threats to humans (mosaic nanoparticles with four to eight distinct RBDs). Mice immunized with RBD nanoparticles, but not soluble antigen, elicited cross-reactive binding and neutralization responses. Mosaic RBD nanoparticles elicited antibodies with superior cross-reactive recognition of heterologous RBDs relative to sera from immunizations with homotypic SARS-CoV-2–RBD nanoparticles or COVID-19 convalescent human plasmas. Moreover, after priming, sera from mosaic RBD–immunized mice neutralized heterologous pseudotyped coronaviruses as well as or better than sera from homotypic SARS-CoV-2–RBD nanoparticle immunizations, demonstrating no loss of immunogenicity against particular RBDs resulting from co-display. A single immunization with mosaic RBD nanoparticles provides a potential strategy to simultaneously protect against SARS-CoV-2 and emerging zoonotic coronaviruses.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a newly emergent betacoronavirus, resulted in a global pandemic in 2020, infecting millions and causing the respiratory disease COVID-19 (*1*, *2*). Two other zoonotic betacoronaviruses, SARS-CoV and Middle East respiratory syndrome coronavirus (MERS-CoV), have also resulted in outbreaks within the past 20 years (*3*). All three viruses presumably originated in bats (*4*), with SARS-CoV and MERS-CoV adapting to intermediary animal hosts before jumping

[1]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA. [2]Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK. [3]Laboratory of Molecular Immunology, The Rockefeller University, New York, NY 10065, USA.
***Corresponding author. Email: bjorkman@caltech.edu**

to humans. SARS-like viruses circulate in bats, and serological surveillance of people living near caves where bats carry diverse coronaviruses demonstrates direct transmission of SARS-like viruses with pandemic potential (*5*). This finding suggests that a pan-coronavirus vaccine is needed to protect against future outbreaks and pandemics. In particular, the bat WIV1 and SHC014 strains are thought to represent an ongoing threat to humans (*6*, *7*).

Most current SARS-CoV-2 vaccine candidates include the spike trimer (S), the viral protein that mediates target cell entry after one or more of its receptor binding domains (RBDs) adopts an "up" position to bind a host receptor (Fig. 1A). The RBDs of human coronaviruses SARS-CoV-2, SARS-CoV, and HCoV-NL63, as well as those of the related animal coronaviruses WIV1 and SHC014, use angiotensin-converting enzyme 2 (ACE2) as their host receptor (*1*, *8*, *9*), whereas other coronaviruses use receptors such as dipeptidyl peptidase 4 (*10*) or sialic acids (*11*, *12*). Consistent with its function in viral entry, S is the primary target of neutralizing antibodies (*13*–*22*), with many targeting the RBD (*14*–*18*, *21*–*26*).

Multivalent display of antigen enhances B cell responses and can provide longer-lasting immunity than monovalent antigens (*27*, *28*); thus, protein-based vaccine candidates often involve a nanoparticle that enables antigen multimerization. Many nanoparticles and coupling strategies have been explored for vaccine design (*29*), with "plug and display" strategies being especially useful (*30*, *31*). In one such approach, multiple copies of an engineered protein domain called SpyCatcher fused to subunits of a virus-like particle form spontaneous isopeptide bonds to purified antigens tagged with a 13-residue SpyTag (*29*–*32*). The SpyCatcher-SpyTag system was used to prepare multimerized SARS-CoV-2 RBD or S trimer that elicited high titers of neutralizing antibodies (*33*, *34*). Although promising for protection against SARS-CoV-2, coronavirus reservoirs in bats suggest future cross-species transmission (*6*, *7*, *35*), necessitating a vaccine that protects against emerging coronaviruses as well as SARS-CoV-2. Here, we prepared SpyCatcher003-mi3 nanoparticles (*31*, *36*) simultaneously displaying SpyTagged RBDs from human and animal coronaviruses to evaluate whether mosaic particles can elicit cross-reactive antibody responses, as previously demonstrated for influenza head domain mosaic particles (*37*). We show that mice immunized with homotypic or mosaic nanoparticles produced broad binding and neutralizing responses, in contrast to plasma antibodies elicited in humans by SARS-CoV-2 infection. Moreover, relative to homotypic SARS-CoV-2 nanoparticles, mosaic nanoparticles showed enhanced heterologous binding and neutralization properties against human and bat SARS-like betacoronaviruses (sarbecoviruses).

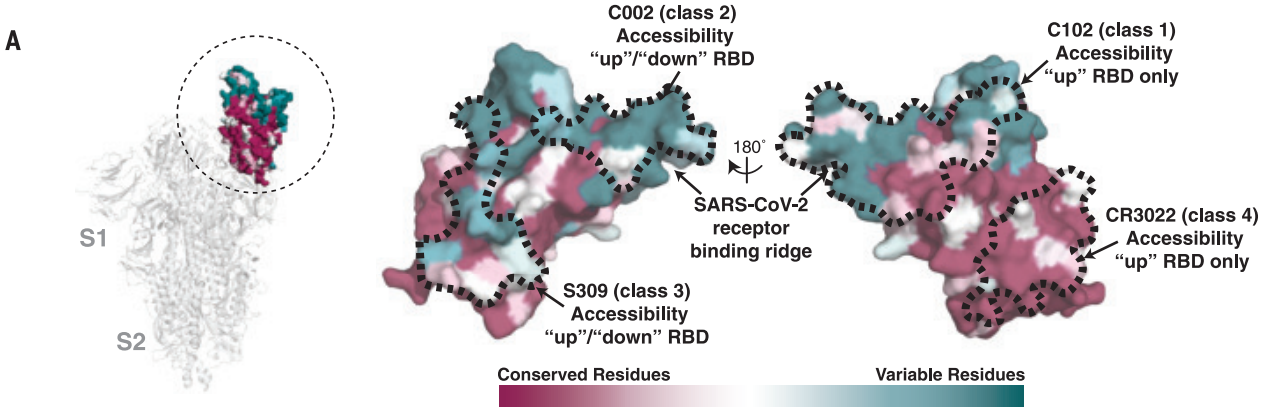
We used a study of sarbecovirus RBD receptor usage and cell tropism (*38*) to guide our choice of RBDs for co-display on mosaic particles. From 29 RBDs that were classified into distinct clades (clades 1, 2, 1/2, and 3) (*38*), we identified diverse RBDs from SARS-CoV, WIV1, and SHC014 (clade 1); SARS-CoV-2 (clade 1/2); Rs4081, Yunnan 2011 (Yun11), and Rf1 (clade 2); and BM-4831 (clade 3). Of these, SARS-CoV-2 and SARS-CoV are human coronaviruses and the rest are bat viruses originating in China or Bulgaria (BM-4831). We also included RBDs from the GX pangolin clade 1/2 coronavirus (referred to here as pang17) (*39*); RaTG13, the bat clade 1/2 virus most closely related to SARS-CoV-2 (*40*); RmYN02, a clade 2 bat virus from China (*41*); and BtKY72, a Kenyan bat clade 3 virus (*42*). Mapping of the sequence conservation across selected RBDs showed varying degrees of sequence identity (68 to 95%), with highest sequence variability in residues corresponding to the SARS-CoV-2 ACE2 receptor binding motif (Fig. 1, A to D, and fig. S1). We chose 8 of the 12 RBDs as sources for three types of mosaic nanoparticles—mosaic-4a (coupled to SARS-2, RaTG13, SHC014, and Rs4081 RBDs); mosaic-4b (coupled to pang17, RmYN02, Rf1, and WIV1 RBDs); and mosaic-8 (coupled to all eight RBDs)—and compared them with homotypic mi3 particles constructed from SARS-CoV-2 RBD alone (homotypic SARS-2). RBDs from SARS, Yun11, BM-4831, and BtKY72, which were not coupled to mosaic particles, were used to evaluate sera for cross-reactive responses.

SpyTag003-RBDs were coupled to SpyCatcher003-mi3 (60 potential conjugation sites) (*36*, *43*) to make homotypic and mosaic nanoparticles (Fig. 2A). Particles were purified by size exclusion chromatography (SEC) and analyzed by SDS–polyacrylamide gel electrophoresis (PAGE), revealing monodisperse SEC profiles and nearly 100% conjugation (Fig. 2, B and C). Representative RBDs were conjugated to SpyCatcher003-mi3 with similar or identical efficiencies (fig. S2), which suggests that mosaic particles contained approximately equimolar mixtures of different RBDs.

We immunized mice with soluble SARS-CoV-2 spike trimer (SARS-2 S), nanoparticles displaying only SARS-2 RBD (homotypic SARS-2), nanoparticles co-displaying RBDs (mosaic-4a, mosaic-4b, or mosaic-8), or unconjugated nanoparticles (mi3). Immunoglobulin G (IgG) responses were evaluated after prime or boost immunizations (Fig. 3A) by enzyme-linked immunosorbent assay (ELISA) against SARS-2 S (Fig. 3B) or a panel of RBDs (Fig. 3, C to F, and fig. S3). Sera from unconjugated nanoparticle-immunized animals (Fig. 3 and fig. S3, black) showed no responses above background. Anti–SARS-2 S trimer and anti–SARS-2 RBD serum responses were similar (Fig. 3, B and C), demonstrating that antibodies elicited against RBDs can access their epitopes on SARS-2 S trimer. We also conducted in vitro neutralization assays using a pseudotyped virus assay that quantitatively correlates with authentic virus neutralization (*44*) for strains known to infect $293T_{ACE2}$ target cells (SARS-CoV-2, SARS, WIV1, and SHC104). Neutralization and ELISA titers were significantly correlated (fig. S4), which implies that ELISAs are predictive of neutralization results when viral entry receptor usage prevents accurate pseudotyped neutralization assays.

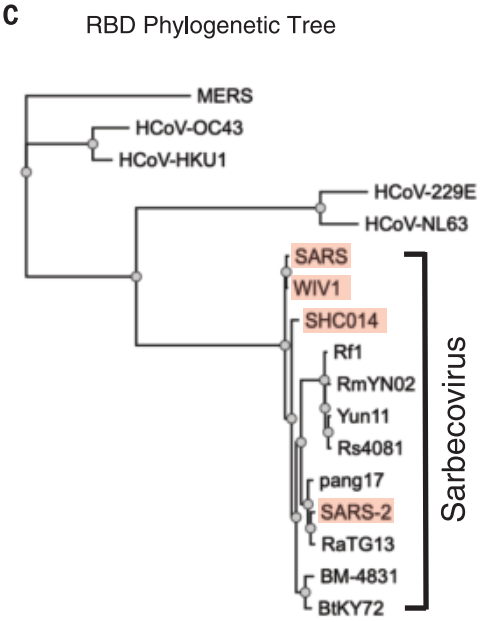
Mice immunized with soluble SARS-2 S trimer showed no binding or neutralization except for autologous responses against SARS-2 after boosting (Fig. 3, C to F, brown bars). By contrast, sera from RBD nanoparticle–immunized animals exhibited binding to all RBDs (Fig. 3, C to F, red, green, yellow, and blue bars; fig. S3A) and neutralization against all four strains after boosting (Fig. 3, C to E), consistent with increased immunogenicities of multimerized antigen on nanoparticles versus soluble antigen (*27*, *28*). Homotypic SARS-2 nanoparticles, but not soluble SARS-2 trimer, induced heterologous responses to zoonotic RBDs and neutralization of heterologous coronaviruses (Fig. 3, D to F). To address whether co-display of SARS-2 RBD along with other RBDs on mosaic-4a and mosaic-8 versus homotypic display of SARS-2 RBD (homotypic SARS-2) diminished anti–SARS-2 responses, we compared SARS-2–specific ELISA and neutralization titers for mosaic versus homotypic immunizations (Fig. 3C); there were no significant differences in IgG anti–SARS-2 titers for animals immunized with homotypic (Fig. 3C, red) versus mosaic nanoparticles (Fig. 3C, green and blue). Thus, in terms of the magnitude of immune response against SARS-2, there was no advantage of immunization with a homotypic RBD nanoparticle versus a mosaic nanoparticle that included SARS-2 RBD.

We next compared serum responses against matched RBDs (RBDs present on an injected nanoparticle) versus mismatched RBDs (RBDs not present on an injected nanoparticle) (Fig. 3 and fig. S3, gray and red horizontal shading, respectively). Although SARS-2 RBD was not presented on mosaic-4b, antibody titers elicited by mosaic-4b immunization (yellow) were not significantly different from titers elicited by matched nanoparticle immunizations [homotypic SARS-2 (red), mosaic-4a (green), and mosaic-8 (blue)], and sera from boosted mosaic-4b–immunized mice neutralized SARS-2 pseudovirus (Fig. 3C). In other matched versus mismatched comparisons, sera showed binding and neutralization of SHC014 and WIV1 regardless of whether these RBDs were included on the injected nanoparticle (Fig. 3D); this result implies sharing of common epitopes among RBDs (Fig. 1A).



B

		Virus	Accession	Residue number	Clade	Host species	Location
8	4a	SARS-CoV-2	MN985325.1	319-539	1/2	Human	Wuhan, Hubei, China
		RaTG13	QHR63300	319-541	1/2	Rhinolophus affinis (bat)	Yunnan, China
		SHC014	KC881005	307-524	1	Rhinolophus sinicus (bat)	Yunnan, China
		Rs4081	KY417143	310-515	2	Rhinolophus sinicus (bat)	Kunming, Yunnan, China
	4b	pang17	QIA48632	317-539	1/2	Manis javanica (pangolin)	Guanxi, China
		RmYN02	EPI_ISL_412977	298-503	2	Rhinolophus malayanus (bat)	Yunnan, China
		Rf1	DQ412042	310-515	2	Rhinolophus ferrumequium (bat)	Yichang, Hubei, China
		WIV1	KF367457	307-528	1	Rhinolophus sinicus (bat)	Yunnan, China
Not included		SARS-CoV	AAP13441.1	318-510	1	Human	Guangdong, China
		Yun11	JX993988	310-515	2	Chaerephon plicata (bat)	Yunnan, China
		BM-4831	NC014470	310-530	3	Rhinolophus blasii (bat)	Strandja Nature Park, Bulgaria
		BtKY72	KY352407	309-530	3	Rhinolophus sp. (bat)	Kenya



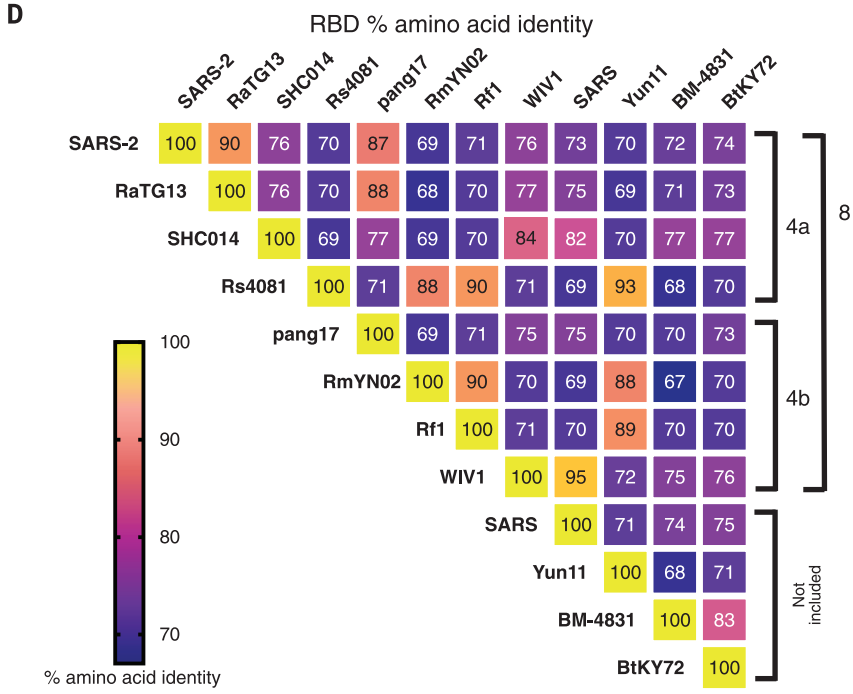


Fig. 1. Properties of RBDs chosen for this study. (**A**) Left: Structure of SARS-CoV-2 S trimer (PDB 6VXX) with one RBD (dashed circle) in an "up" position. Center and right: Sequence conservation of 12 RBDs calculated by the ConSurf Database (*49*) plotted on a surface representation of the RBD structure (PDB 7BZ5). Epitopes for representatives from defined classes of RBD-binding antibodies (classes 1 to 4) (*24*) are indicated by dashed lines. (**B**) Summary of properties of the viral strains from which the 12 sarbecovirus RBDs were derived. (**C**) Phylogenetic tree of human and selected other coronaviruses based on RBD protein sequences. Red shading indicates strains known to use ACE2 as a receptor. (**D**) Heat map showing percent amino acid sequence identities among 12 sarbecovirus RBDs.

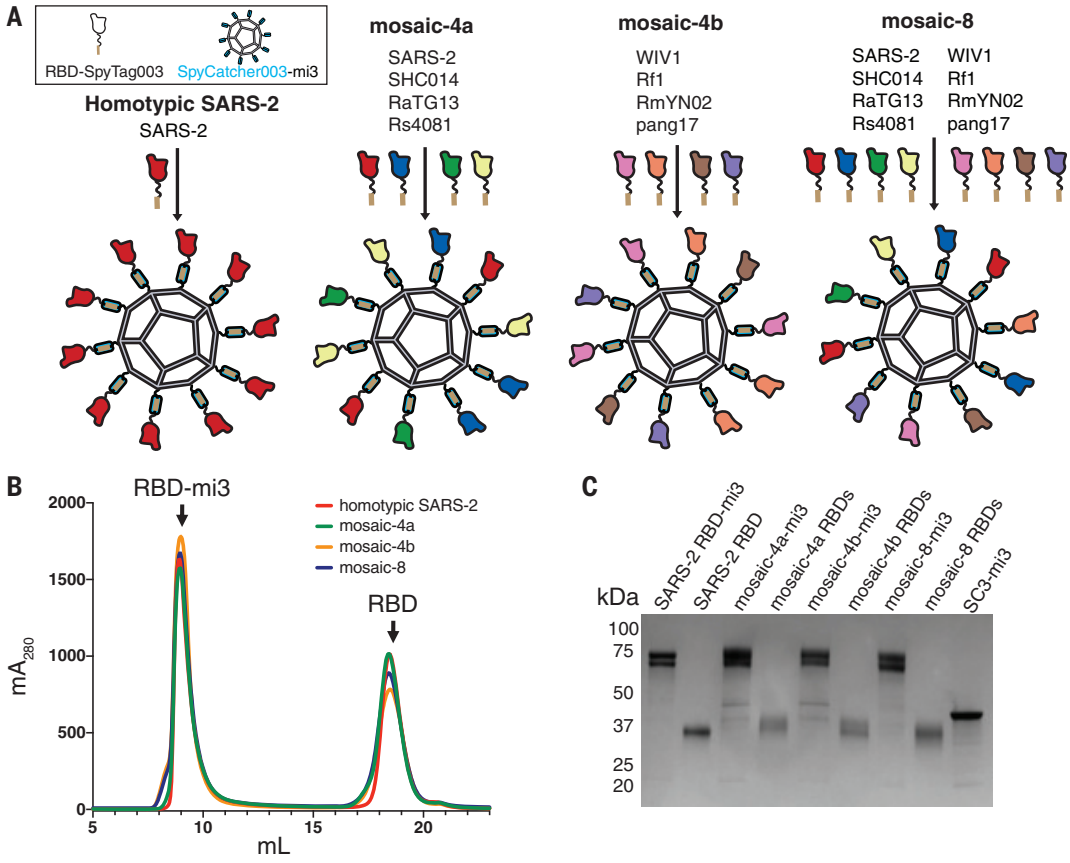


Fig. 2. Construction of RBD nanoparticles. (**A**) Left: SpyTagged RBDs were attached to SpyCatcher003-mi3 to make a homotypic particle and three mosaic particles. There are 60 potential coupling sites on mi3; only 10 are shown for clarity. (**B**) SEC profile showing separation of RBD nanoparticles and free RBD proteins. (**C**) Coomassie-stained SDS-PAGE of RBD-coupled nanoparticles, free RBD proteins, and uncoupled SpyCatcher003-mi3 particles (SC3-mi3).

In an experiment that demonstrated the advantages of mosaic versus homotypic SARS-2 nanoparticles, sera from mosaic-8–immunized mice bound SHC014 and WIV1 RBDs significantly better after priming than sera from homotypic SARS-2–immunized mice and retained better binding to SHC014 RBD after boosting (Fig. 3D). Thus, the potential increased avidity of the homotypic SARS-2 nanoparticle displaying only one type of RBD over the mosaic-8 nanoparticles did not confer increased breadth. Moreover, mosaic-8–immunized and boosted sera were more potent than sera from homotypic SARS-2–immunized animals in neutralizing SHC014 and WIV1 (Fig. 3D). Neutralization of the SHC014 and WIV1 pseudoviruses by mosaic-8 sera suggests that combining RBDs on a mosaic nanoparticle does not diminish the immune response against a particular RBD, as also suggested by ELISA binding of sera to Rs4081 and RaTG13 (fig. S3, A and B).

To further address whether RBD nanoparticles elicited antibodies that recognized totally mismatched strains and SARS-CoV-2 RBD mutants, we evaluated sera for binding to SARS, Yun11, BM-4831, and BtKY72 RBDs (Fig. 3, E and F); SARS-2 RBD mutants (fig. S3C); and MERS-CoV RBD (fig. S3D), as well as for neutralization in SARS pseudovirus assays (Fig. 3E). We found no reductions in SARS-2 RBD binding as a result of mutations [Y453F, the "Danish mink variant" (*45*), or a Q493K/Q498Y/P499T triple mutant (*46*)] (fig. S3C), no binding of any elicited sera to MERS-CoV RBD (fig. S3D), and higher and more cross-reactive antibody responses for mosaic immunizations than for homotypic SARS-2 immunizations (e.g., mosaic-8–primed and boosted animals showed significantly higher titers against SARS RBD than did sera from homotypic SARS-2–immunized mice) (Fig. 3E). After priming, sera from the homotypic SARS-2–immunized animals did not neutralize SARS, whereas the mosaic-4b and mosaic-8 sera were neutralizing (Fig. 3E), perhaps because these nanoparticles included WIV1 RBD, which is related by 95% amino acid identity to SARS RBD (Fig. 1D). After boosting, SARS-2 and mosaic-4a sera were also neutralizing, although titers were lower than for mosaic-8–immunized animals by a factor of ~4 (Fig. 3E). ELISA titers against other mismatched RBDs (Yun11, BM-4831, and BtKY72) were significantly higher for sera collected after mosaic-8 priming than for sera from homotypic SARS-2 priming, and heightened binding was retained after boosting (Fig. 3F). Thus, relative to homotypic SARS-2 nanoparticles, mosaic nanoparticles (particularly mosaic-8) induce higher antibody titers against mismatched RBDs, This is another finding that favors the co-display approach for inducing broader anti-coronavirus responses, especially after a single prime.

Using flow cytometry, we investigated the potential for cross-reactive recognition—specifically, whether B cell receptors on IgG^+ splenic B cells from RBD nanoparticle–boosted animals could simultaneously recognize RBDs from SARS-2 and Rs4081 (related by 70% sequence identity) (Fig. 1D and fig. S5). Whereas control animals were negative, all other groups showed B cells that recognized SARS-2 and Rs4081 RBDs simultaneously, suggesting the existence of antibodies that cross-react with both RBDs (fig. S5E).

To compare antibodies elicited by RBD nanoparticle immunization to antibodies elicited by SARS-CoV-2 infection, we repeated ELISAs against the RBD panel using IgGs from COVID-19 plasma donors (*47*) (Fig. 4). Most of the convalescent plasmas showed detectable binding to SARS-2 RBD (Fig. 4A). However, binding to other sarbecovirus RBDs (RaTG13, SHC014, WIV1, Rs4081, and BM-4831) was significantly weaker than binding to SARS 2 RBD, with many human plasma IgGs showing no binding above background (Fig. 4, B to G). In addition, although convalescent plasma IgGs neutralized SARS-CoV-2 pseudoviruses, they showed weak or no neutralization of SARS, SHC014, or WIV1 pseudoviruses (Fig. 4H). These results are consistent with little to no cross-reactive recognition of RBDs from zoonotic coronavirus strains resulting from SARS-CoV-2 infection in humans.

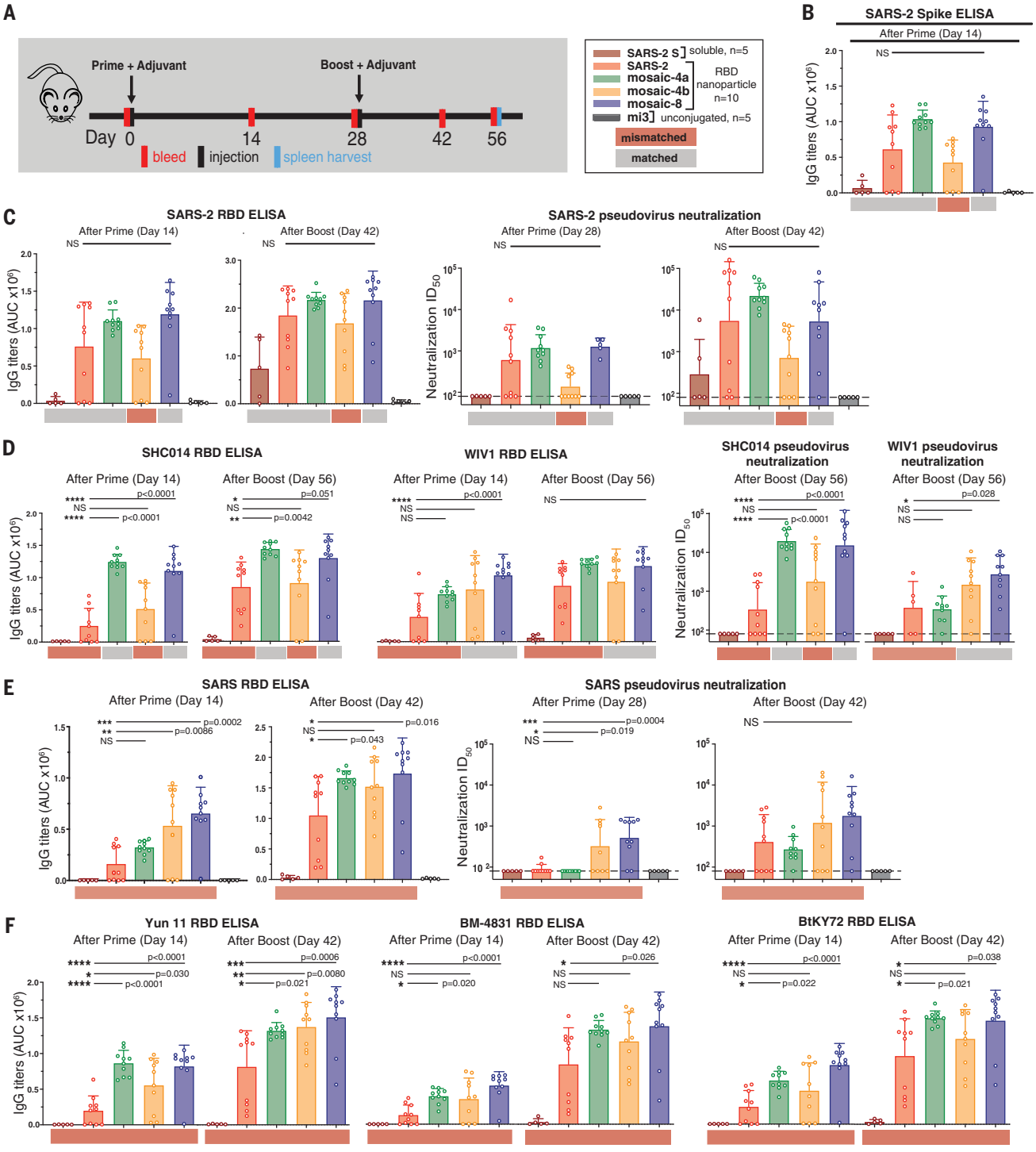
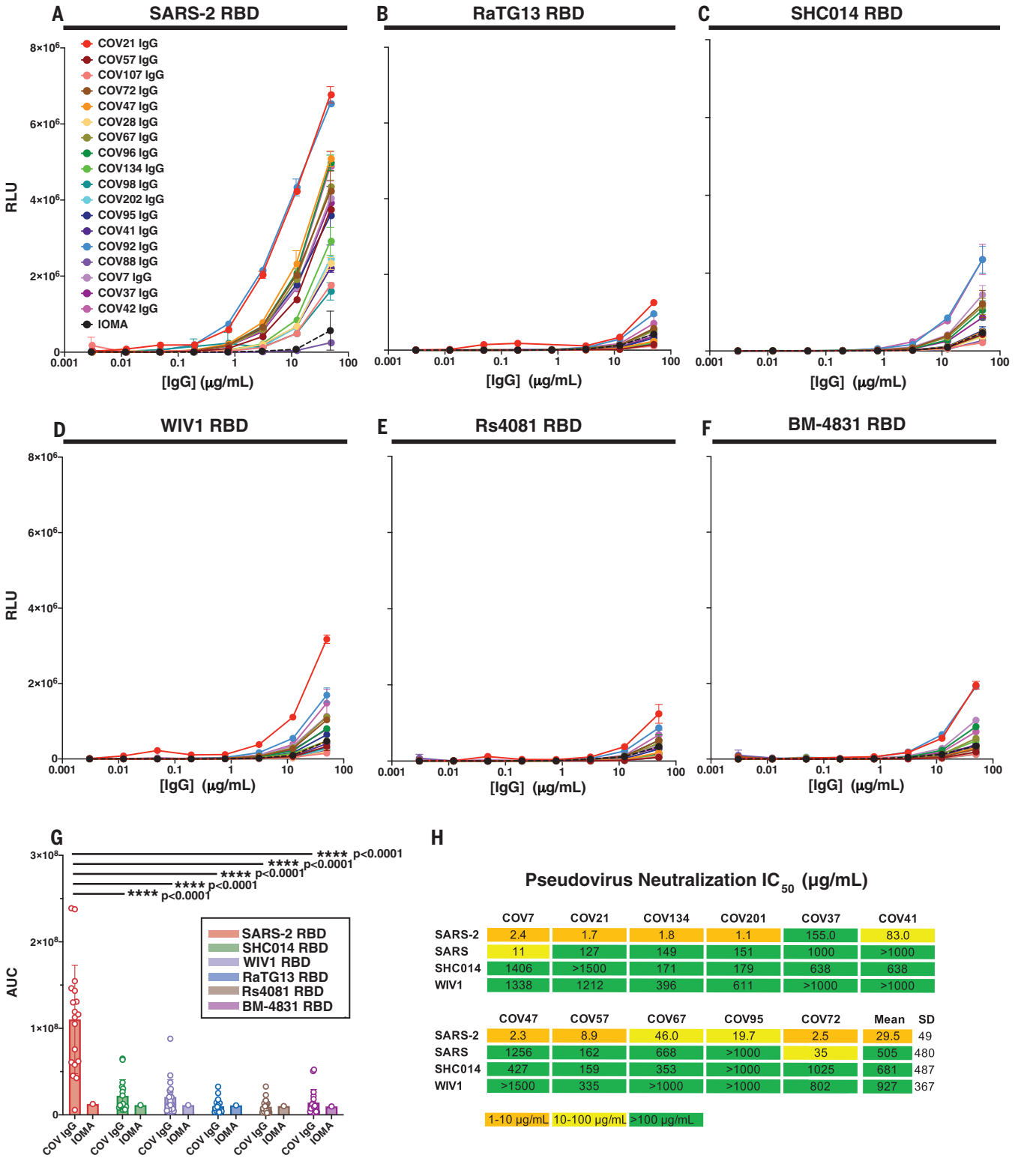


Fig. 3. RBD nanoparticles induce cross-reactive IgG responses in immunized mice. Red and gray rectangles below ELISA and neutralization data represent mismatched strains (red; RBD from that strain was not present on the immunized particle) or matched strains (gray; RBD was present on the immunized particle). (**A**) Left: Immunization schedule; adjuvant is AddaVax (Invivogen). Right: Key for immunizations; number of mice in each cohort is indicated. (**B** to **F**) Neutralization and/or binding data for serum IgGs for recognition of (B) SARS-2 spike trimer, (C) SARS-2 RBD and SARS-2 pseudovirus, (D) SHC014 and WIV1 RBDs and corresponding pseudoviruses, (E) SARS RBD and SARS pseudovirus, and (F) Yun11, BM-4831, and BtKY72 RBDs. Mice were immunized with soluble SARS-CoV-2 S trimer (SARS-2 S; brown bars) or the following nanoparticles: homotypic SARS-2 (red), mosaic-4a (green), mosaic-4b (yellow), mosaic-8 (blue), or unconjugated SpyCatcher003-mi3 (mi3; black). ELISA data from serum IgG responses to SARS-2 spike trimer (B) or RBDs [(C) to (F)] are shown as area under the curve (AUC). For (C) to (E), neutralization potencies are presented as half-maximal inhibitory dilutions (ID_{50} values) of sera against the pseudoviruses from the indicated coronavirus strains. Dashed horizontal lines correspond to the lowest dilution representing the limit of detection. Each dot represents serum from one animal, with means and SDs for vaccinated cohorts denoted by rectangles and horizontal lines, respectively. Significant differences between groups linked by horizontal lines are indicated by asterisks and *P* values. NS, not significant.



	COV7	COV21	COV134	COV201	COV37	COV41
SARS-2	2.4	1.7	1.8	1.1	155.0	83.0
SARS	11	127	149	151	1000	>1000
SHC014	1406	>1500	171	179	638	638
WIV1	1338	1212	396	611	>1000	>1000

	COV47	COV57	COV67	COV95	COV72	Mean	SD
SARS-2	2.3	8.9	46.0	19.7	2.5	29.5	49
SARS	1256	162	668	>1000	35	505	480
SHC014	427	159	353	>1000	1025	681	487
WIV1	>1500	335	>1000	>1000	802	927	367

Fig. 4. IgGs from convalescent COVID-19 plasma show little to no cross-reactive responses. (**A** to **F**) Plasma IgG (*18*, *24*) responses were evaluated by ELISA [data shown as binding curves with plasma names (*18*) listed] against RBDs from (A) SARS-2, (B) RaTG13, (C) SHC014, (D) WIV1, (E) Rs4081, and (F) BM-4831. Data points are plotted as means ± SD of duplicate measurements. IOMA, an anti–HIV-1 IgG (*50*), was used as a control. (**G**) ELISA results from (A) to (F), presented as AUC; each dot represents one plasma sample, with means ± SD shown as colored bars. Significant differences between groups linked by horizontal lines are indicated by asterisks and *P* values. (**H**) IC_{50} values for pseudotyped neutralization assays using IgGs from COV7, COV21, and COV72 plasmas (*18*) (evaluated at top concentrations of 1500 μg/ml) against the indicated strains. Mean = arithmetic mean IC_{50}.

Our results confirm that multimerization of RBDs on nanoparticles enhances immunogenicity relative to soluble antigen (*33*, *48*). We found that homotypic SARS-2 nanoparticle immunization produces IgG responses that bind zoonotic RBDs and neutralize heterologous coronaviruses after boosting. By contrast, soluble SARS-2 S immunization and natural infection with SARS-CoV-2 resulted in weak or no heterologous responses in plasmas. Co-display of SARS-2 RBD along with diverse RBDs on mosaic nanoparticles showed no disadvantages for eliciting neutralizing antibodies against SARS-CoV-2 relative to homotypic SARS-2 nanoparticles; therefore, mosaic nanoparticles may represent a candidate vaccine to protect against COVID-19. Furthermore, relative to homotypic SARS-2 RBD particles, the mosaic co-display strategy demonstrated advantages for eliciting neutralizing antibodies against zoonotic sarbecoviruses, thus potentially also providing protection against emerging coronaviruses with human spillover potential. Neutralization of matched and mismatched strains was observed after mosaic priming; hence, a single injection of a mosaic RBD nanoparticle might be sufficient in a vaccine. Because COVID-19 convalescent plasmas showed little to no recognition of coronavirus RBDs other than SARS-CoV-2, COVD-19–induced immunity in humans may not protect against another emergent coronavirus. However, the mosaic nanoparticles described here could be used as described or easily adapted so that they present RBDs from newly discovered zoonotic coronaviruses.

REFERENCES AND NOTES

1. P. Zhou *et al.*, *Nature* **579**, 270–273 (2020).
2. F. Wu *et al.*, *Nature* **579**, 265–269 (2020).
3. E. de Wit, N. van Doremalen, D. Falzarano, V. J. Munster, *Nat. Rev. Microbiol.* **14**, 523–534 (2016).
4. W. Li *et al.*, *Science* **310**, 676–679 (2005).
5. N. Wang *et al.*, *Virol. Sin.* **33**, 104–107 (2018).
6. V. D. Menachery *et al.*, *Nat. Med.* **21**, 1508–1513 (2015).
7. V. D. Menachery *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 3048–3053 (2016).
8. W. Li *et al.*, *Nature* **426**, 450–454 (2003).
9. M. Hoffmann *et al.*, *Cell* **181**, 271–280.e8 (2020).
10. V. S. Raj *et al.*, *Nature* **495**, 251–254 (2013).
11. R. Vlasak, W. Luytjes, W. Spaan, P. Palese, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4526–4529 (1988).
12. X. Huang *et al.*, *J. Virol.* **89**, 7202–7213 (2015).
13. T. S. Fung, D. X. Liu, *Annu. Rev. Microbiol.* **73**, 529–557 (2019).
14. P. J. M. Brouwer *et al.*, *Science* **369**, 643–650 (2020).
15. Y. Cao *et al.*, *Cell* **182**, 73–84.e16 (2020).
16. C. Kreer *et al.*, *Cell* **182**, 843–854.e12 (2020).
17. L. Liu *et al.*, *Nature* **584**, 450–456 (2020).
18. D. F. Robbiani *et al.*, *Nature* **584**, 437–442 (2020).
19. R. Shi *et al.*, *Nature* **584**, 120–124 (2020).
20. S. J. Zost *et al.*, *Nat. Med.* **26**, 1422–1427 (2020).
21. T. F. Rogers *et al.*, *Science* **369**, 956–963 (2020).
22. E. Seydoux *et al.*, *Immunity* **53**, 98–105.e5 (2020).
23. S. J. Zost *et al.*, *Nature* **584**, 443–449 (2020).
24. C. O. Barnes *et al.*, *Nature* **588**, 682–687 (2020).
25. D. Pinto *et al.*, *Nature* **583**, 290–295 (2020).
26. L. Piccoli *et al.*, *Cell* **183**, 1024–1042.e21 (2020).
27. J. López-Sagaseta, E. Malito, R. Rappuoli, M. J. Bottomley, *Comput. Struct. Biotechnol. J.* **14**, 58–68 (2015).
28. M. K. Slifka, I. J. Amanna, *Front. Immunol.* **10**, 956 (2019).
29. K. D. Brune, M. Howarth, *Front. Immunol.* **9**, 1432 (2018).
30. K. D. Brune *et al.*, *Sci. Rep.* **6**, 19234 (2016).
31. T. U. J. Bruun, A. C. Andersson, S. J. Draper, M. Howarth, *ACS Nano* **12**, 8855–8866 (2018).
32. B. Zakeri *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, E690–E697 (2012).
33. T. K. Tan *et al.*, bioRxiv 275701 [preprint]. 31 August 2020.
34. B. Zhang *et al.*, *Sci. Rep.* **10**, 18149 (2020).
35. B. Hu *et al.*, *PLOS Pathog.* **13**, e1006698 (2017).
36. R. Rahikainen *et al.*, *Angew. Chem. Int. Ed.* **60**, 321–330 (2021).
37. M. Kanekiyo *et al.*, *Nat. Immunol.* **20**, 362–372 (2019).
38. M. Letko, A. Marzi, V. Munster, *Nat. Microbiol.* **5**, 562–569 (2020).
39. T. T.-Y. Lam *et al.*, *Nature* **583**, 282–285 (2020).
40. K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, R. F. Garry, *Nat. Med.* **26**, 450–452 (2020).
41. H. Zhou *et al.*, *Curr. Biol.* **30**, 2196–2203.e3 (2020).
42. Y. Tao, S. Tong, *Microbiol. Resour. Announce.* **8**, e00548-19 (2019).
43. A. H. Keeble *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 26523–26533 (2019).
44. F. Schmidt *et al.*, *J. Exp. Med.* **217**, e20201181 (2020).
45. L. van Dorp *et al.*, bioRxiv 384743 [preprint]. 16 November 2020.
46. S. R. Leist *et al.*, *Cell* **183**, 1070–1085.e12 (2020).
47. C. O. Barnes *et al.*, *Cell* **182**, 828–842.e16 (2020).
48. A. C. Walls *et al.*, *Cell* **183**, 1367–1382.e17 (2020).
49. M. Landau *et al.*, *Nucleic Acids Res.* **33**, W299–W302 (2005).
50. H. B. Gristick *et al.*, *Nat. Struct. Mol. Biol.* **23**, 906–915 (2016).

ACKNOWLEDGMENTS

We thank K. Brune (Genie Biotech) for advice about mi3 production; J. Bloom (Fred Hutchinson) and P. Bieniasz (Rockefeller University) for neutralization assay reagents; J. Vielmetter and Caltech's Beckman Institute Protein Expression Center for protein production; A. Flyak for help with flow cytometry; M. Murphy for figures; COVID-19 plasma donors, B. Coller, S. Schlesinger, and the Rockefeller University Hospital Clinical Research Support Office and nursing staff; and A. Flyak and A. DeLaitsch for critical reading of the manuscript. **Funding:** Supported by NIH grant P01-AI138938-S1 (P.J.B. and M.C.N.), the Caltech Merkin Institute for Translational Research (P.J.B.), a George Mason University Fast Grant (P.J.B.), and the Medical Research Council (MR/P001351/1) (M.H.) (this UK-funded award is part of the EDCTP2 program supported by the European Union). M.C.N. is a Howard Hughes Medical Institute Investigator. **Author contributions:** A.A.C., C.O.B., and P.J.B. conceived and designed experiments; A.A.C., P.N.P.G., Y.E.L., P.R.H., S.O., and L.M.K. performed experiments; H.-J.W. generated and validated SpyCatcher003-mi3; M.H. supervised the generation and validation of SpyCatcher003-mi3; and A.A.C., J.R.K., A.P.W., C.O.B., M.C.N., and P.J.B. analyzed data and wrote the paper with contributions from other authors. **Competing interests:** M.H. is an inventor on a patent on SpyTag/SpyCatcher (EP2534484) and a patent application on SpyTag003:SpyCatcher003 (UK Intellectual Property Office 1706430.4), as well as a SpyBiotech cofounder, shareholder, and consultant. P.J.B. and A.A.C. are inventors on a provisional application from the California Institute of Technology that covers the mosaic nanoparticles described in this work. **Data and materials availability:** All data are available in the main text or the supplementary materials. Materials are available upon request to the corresponding author with a signed material transfer agreement. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/371/6530/735/suppl/DC1
Materials and Methods
Figs. S1 to S5
References (*51*–*58*)
MDAR Reproducibility Checklist

12 November 2020; accepted 7 January 2021
Published online 12 January 2021
10.1126/science.abf6840

CORONAVIRUS

Immunological characteristics govern the transition of COVID-19 to endemicity

Jennie S. Lavine[1]*, Ottar N. Bjornstad[2], Rustom Antia[1]

We are currently faced with the question of how the severity of infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) may change in the years ahead. Our analysis of immunological and epidemiological data on endemic human coronaviruses (HCoVs) shows that infection-blocking immunity wanes rapidly but that disease-reducing immunity is long-lived. Our model, incorporating these components of immunity, recapitulates both the current severity of SARS-CoV-2 infection and the benign nature of HCoVs, suggesting that once the endemic phase is reached and primary exposure is in childhood, SARS-CoV-2 may be no more virulent than the common cold. We predict a different outcome for an emergent coronavirus that causes severe disease in children. These results reinforce the importance of behavioral containment during pandemic vaccine rollout, while prompting us to evaluate scenarios for continuing vaccination in the endemic phase.

Humans have regularly been threatened by emerging pathogens that kill a substantial fraction of all people born. Recent decades have seen multiple challenges from acute virus infections, including severe acute respiratory syndrome (SARS), Middle East respiratory syndrome (MERS), Hendra, Nipah, and Ebola. Fortunately, all were locally contained. When containment is not immediately successful, as is likely for the novel betacoronavirus severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (*1*, *2*), we need to understand and plan for the transition to endemicity and continued circulation, with possible changes in disease severity

[1]Department of Biology, Emory University, Atlanta, GA 30322, USA. [2]Department of Biology and Center for Infectious Disease Dynamics, The Pennsylvania State University, University Park, PA 16802, USA.
*Corresponding author. Email: jslavin@emory.edu

owing to virus evolution and buildup of host immunity and resistance.

SARS-CoV-2 is an emerging virus that causes COVID-19. The virus has a high basic reproductive number (R_0) and is transmissible during the asymptomatic phase of infection, both of which make it hard to control (*3*). However, there are six other coronaviruses with known human chains of transmission, which may provide clues to future scenarios for the current pandemic. There are four human coronaviruses (HCoVs) that circulate endemically around the globe; these cause only mild symptoms and are not a considerable public health burden (*4*). Two other HCoV strains, SARS-CoV-1 and MERS-CoV, emerged in recent decades and have higher case fatality ratios (CFRs) and higher infection fatality ratios (IFRs) than COVID-19 but were contained and thus never spread widely (*5*, *6*).

We propose a model to explore the potential changes in both transmission and disease severity of emerging HCoVs through the transition to endemicity. We focus on SARS-CoV-2 and discuss how the conclusions would differ for emerging coronaviruses more akin to SARS-CoV-1 and MERS-CoV. Our hypothesis is that all HCoVs elicit immunity with similar characteristics, and the current acute public health problem is a consequence of epidemic emergence into an immunologically naïve population in which older age groups with no previous exposure are most vulnerable to severe disease. We use our estimates of immunological and epidemiological parameters for endemic HCoVs to develop a quantitative model for endemic transmission of a virus with SARS-CoV-2–like characteristics, including the age dependence of severity. Our model explicitly considers three separate measures for immune efficacy that wane at different rates (fig. S1).

Building on ideas from the vaccine modeling literature, we suggest that immunity may provide protection in three ways (*7*). In its most robust form, sterilizing immunity can prevent a pathogen from replicating, thereby rendering the host refractory to reinfection. We term this property immune efficacy with respect to susceptibility (IE_S). If immunity does not prevent reinfection, it may still attenuate pathology due to reinfection (IE_P) and/or reduce transmissibility or infectiousness (IE_I). Indeed, experimental reexposure studies on endemic HCoVs provide evidence that the three immune efficacies do not wane at the same rate (*8*, *9*). Callow *et al.*'s experimental study (*8*) shows that reinfection is possible within one year (relatively short IE_S); however, upon reinfection, symptoms are mild (high IE_P) and the virus is cleared more quickly (moderate IE_I). Details on the derivation of the model can be found in section 2 of the supplementary materials (SM).

We reanalyze a detailed dataset that estimates age-specific seroprevalence on the basis of both immunoglobulin M (IgM; acute response) and IgG (long-term memory) against all four circulating HCoVs in children and adults (*10*) to estimate parameter ranges for transmission and waning of immunity (Fig. 1A). The rapid rise in both IgM and IgG seroprevalence indicates that primary infection with all four endemic HCoV strains happens early in life, and our analysis of these data gives us an estimate for the mean age of primary infection (MAPI) between 3.4 and 5.1 years, with almost everyone infected by age 15 (see SM section 1 for details). The absence of detectable IgM titers in any individual over the age of 15 years suggests that reinfection of adults causes a recall response, indicating that while HCoV-specific immunity may wane, it is not lost. Whether immunity would wane to naïve levels in the absence of high pathogen circulation remains an open question.

For most people to be infected so early in life—younger even than measles in the prevaccine era—the attack rate must exceed transmission from primary infections alone. The model shows that a high attack rate can arise from a combination of high transmissibility from primary infections (i.e., high R_0), waning of sterilizing immunity, and substantial transmission from reinfections in older individuals. The rapid waning of sterilizing immunity is also reported in experimental HCoV infections of humans, which showed that reinfection is possible 1 year after an earlier infection, albeit

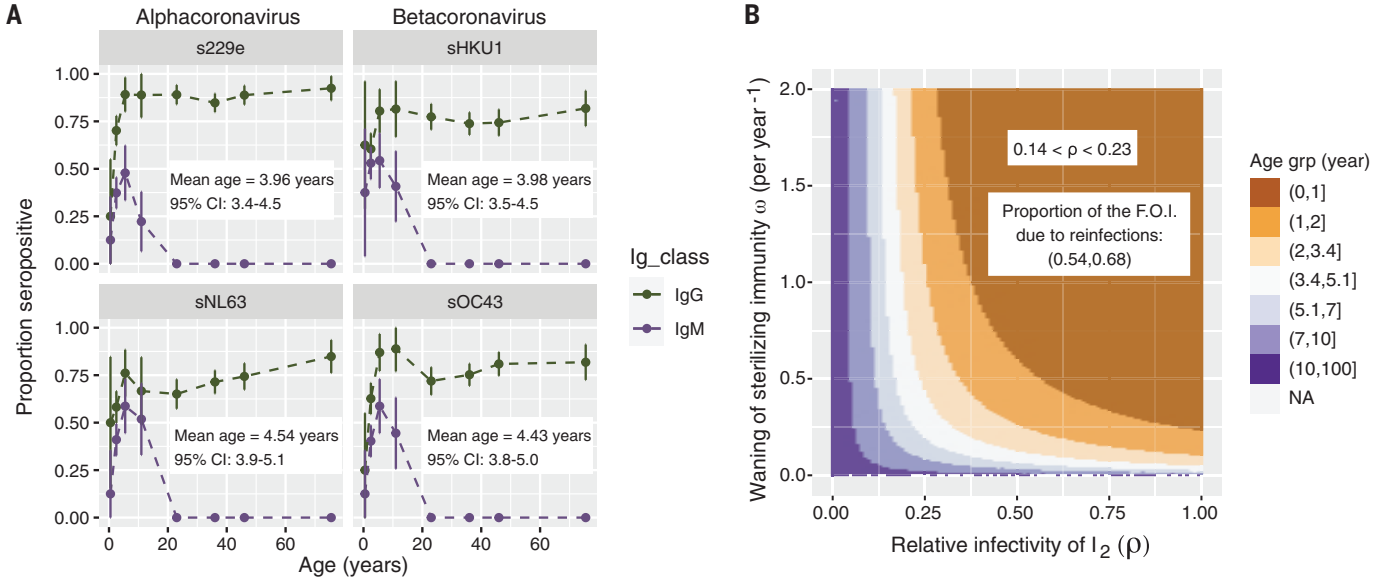


Fig. 1. A low mean age of primary infection suggests that partially transmissible reinfections are common. (**A**) Mean proportion seropositive for IgG (green, top lines) and IgM (purple, bottom lines) against the four endemic HCoV strains [dots connected by dashed lines; vertical lines represent the 95% confidence interval (CI); data from (*10*)]. The mean age of primary infection (MAPI) based on IgM data with 95% CI is shown in the inset of each panel (see SM for details). (**B**) MAPI as a function of waning of sterilizing immunity ω (*y* axis) and transmissibility of reinfections ρ (*x* axis). The MAPI was calculated from the equilibrium dynamics of the model shown in fig. S1 and supplementary equations 3 to 9 with a plausible basic reproductive number (R_0 = 5), $0 < \omega < 2$, and $0 < \rho < 1$. See SM section 2.1 for details. The inset shows the plausible combinations of values of ρ and ω consistent with the MAPI for HCoVs estimated in (A). F.O.I., force of infection; I_2, reinfection. [See fig. S1 for parallel figures calculated at extreme plausible values for R_0 (i.e., R_0 = 2 and 10).]

with milder symptoms (IE_P) and a shorter duration (IE_I) (*8*). Figure 1B shows the plausible combinations of waning immunity and transmission from reinfected individuals that are required to produce the MAPI observed in Fig. 1A, based on steady-state infection levels (see SM section 2.1 for details). Table 1 shows the ranges of the parameters used in our simulations.

At the beginning of an outbreak, the age distribution of cases mirrors that of the population (Fig. 2A). However, once the demographics of infection reaches a steady state, our model predicts that primary cases occur almost entirely in babies and young children, who, in the case of COVID-19, experience a low CFR and a concomitantly low IFR. Reinfections in older individuals are predicted to be common during the endemic phase and to contribute to transmission, but in this steady-state population, older individuals, who would be at risk for severe disease from a primary infection, have acquired disease-reducing immunity after infection during childhood. The top panel of Fig. 3B illustrates how the overall IFR for SARS-CoV-2 drops drastically, eventually falling below that of seasonal influenza (~0.001) once the endemic steady-state is reached.

The time it takes to complete the shift in IFR as endemicity develops depends on both transmission (R_0) and loss of immunity [waning of sterilizing immunity (ω) and transmissibility of reinfections (ρ)], as shown in Fig. 2B and fig. S4. The transition from epidemic to endemic dynamics is associated with a shift in the age distribution of primary infections to lower age groups (Fig. 2A). This transition may take anywhere from a few years to a few decades, depending on how quickly the pathogen spreads. The rate of spread, measured by R_0, is determined by a combination of viral properties and the frequency of social contacts and may therefore be reduced by social distancing. The top panel in Fig. 2A shows the effect of reducing R_0 to 2, whereas the middle and bottom panels show the dynamics for higher R_0, which are more akin to those of SARS-CoV-2 in the absence of control measures. If transmission is high, the model predicts a high case load and high death rate in earlier years following emergence (Fig. 2 and fig. S5). We see that, as might be expected, longer-lasting sterilizing immunity slows down the transition to endemicity (Fig. 2B). These results are robust to a more biologically realistic distribution for the duration of sterilizing immunity and the possibility that the generation of protective immunity requires more than one infection (see SM section 3 and figs. S5 to S9).

Slowing down the epidemic through social distancing measures that reduce R_0 to close to 1 flattens the curve, thus delaying infections and preventing most deaths from happening early on, affording critical time for the development of an effective vaccine (fig. S10). If vaccine-induced IE_S and IE_P immunity is similar to that induced by HCoV infections, the vaccine may usher in the endemic regime more quickly. The model code (see the acknowledgments) provides a flexible scaffolding for studying alternative vaccination scenarios. Notably, the model predicts that once the endemic state is reached, mass vaccination may no longer be necessary to save lives (see SM section 4 and fig. S11).

We can extend our predictions to two other potentially emerging coronavirus infections, SARS and MERS. Our model predicts that in the endemic state, the IFR of a circulating HCoV depends primarily on the severity of childhood infections. In the case of SARS-CoV-1, which is more pathogenic than SARS-CoV-2, we still expect a low disease burden in the endemic phase, because SARS-CoV-1, like SARS-CoV-2, has a low IFR in young people (Fig. 3). However, data suggest that not all emerging HCoVs follow this optimistic pattern; the overall IFR of an endemic MERS-like virus would not decrease during the transition to endemicity, as seen in Fig. 3B, and this is because disease severity (and IFR) is high in children, the age group expected to experience the bulk of primary cases during the endemic phase. In the endemic phase, a vaccination program against MERS would therefore be necessary to avoid excess mortality (fig. S11).

The key result from our model framework that explicitly recognizes that functional immunity to reinfection, disease, and shedding are different is that, in contrast with infections that are severe in childhood, SARS-CoV-2 could join the ranks of mild, cold-causing endemic

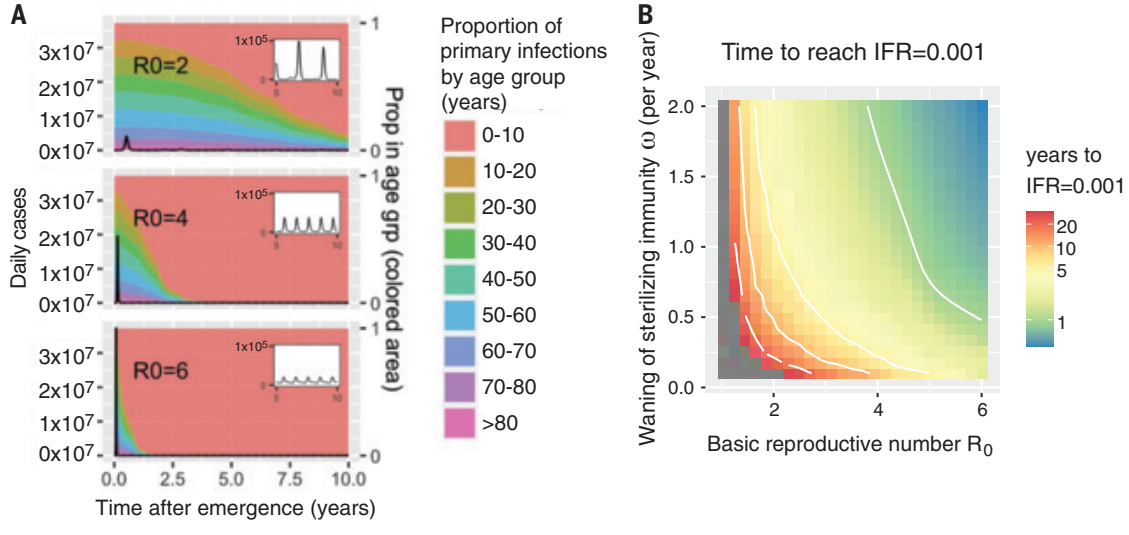


Fig. 2. The time scale of the transition from epidemic to endemic dynamics for emerging coronaviruses depends on R_0 and the rate of immune waning. Transition from epidemic to endemic dynamics for emerging HCoVs, simulated from an extension of the model presented in fig. S1 that includes age structure. Demographic characteristics (age distribution, birth, and age-specific death rates) are taken from the United States, and seasonality is incorporated via a sinusoidal forcing function (see SM section 2.2). Weak social distancing is approximated by $R_0 = 2$. (See figs. S9 to S11 for strong social distancing results, $R_0 < 1.5$.) (**A**) Daily number of new infections (black line; calculations in SM section 2.3). An initial peak is followed by a low-incidence endemic state (years 5 to 10 shown in the inset). A higher R_0 results in a larger and faster initial epidemic and a more rapid transition to endemic dynamics. The proportion of primary cases in different age groups changes over time (plotted in different colors), and the transition from epidemic to endemic dynamics results in primary cases being restricted to younger age groups. Parameters for simulations: $\omega = 1$ and $\rho = 0.7$. (**B**) Time for the average IFR (6-month moving average) to fall to 0.001, which is the IFR associated with seasonal influenza. Gray areas represent simulations where the IFR did not reach 0.001 within 30 years. The time to IFR = 0.001 decreases as the transmissibility (R_0) increases and the duration of sterilizing immunity becomes shorter. Results are shown for $\rho = 0.7$. See SM section 2.3 and figs. S4 to S7 for sensitivity analyses and model specifications.

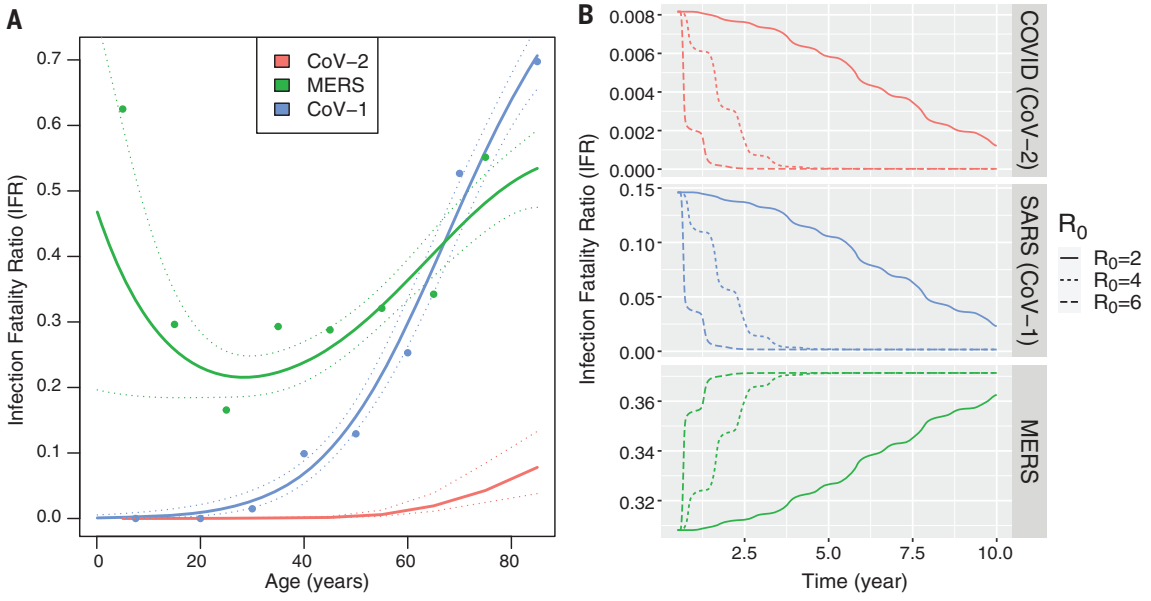


Fig. 3. The overall infection fatality ratio (IFR) of emerging coronaviruses once they reach endemicity is strongly influenced by the IFR of young children in the initial epidemic. The age dependence of the IFR determines how the overall IFR changes during the transition from epidemic to endemic dynamics for emerging HCoVs. (**A**) Age dependence of the IFRs for the three emerging HCoVs. Primary infections with MERS-CoV and SARS-CoV-1 are consistently symptomatic, and the IFR and CFR are therefore assumed to be the same. SARS-CoV-1 and SARS-CoV-2 have J-shaped profiles, with a monotonic increase in IFR with age. The age-specific IFR for MERS-CoV is U-shaped, with high mortality in both the young and old age groups. Details of the statistical smoothing are described in SM section 6. (**B**) The overall IFR changes during the transition to endemic dynamics. These calculations assume that deaths due to reinfections are negligible. We relax this assumption to allow for a slower buildup of immunity and possible death due to secondary infection in figs. S5 to S9 and show that the qualitative results do not change.

Table 1. Characteristics of coronavirus-immunity interactions and relevant parameter ranges.

Characteristic and symbol	Estimates from literature	Value (range)	Citations
Primary infectious period ($1/\gamma$)	≥5.6 days ~10 days	9 days	(*8*) (*30*)
Primary transmissibility [$R_0 = \beta/(\gamma + \mu)$]	4 to 9	2 to 10	(*31*)
Secondary transmissibility (relative to primary, ρ)	0.35 0.04 to 0.97	0 to 1	(*8*) Fig. 2 and fig. S2
Duration of sterilizing immunity ($1/\omega$)	0.91 years 1.67 years 0.5 to 2 years	0.5 to 10 years	(*8*) and SM section 5 (*8*) and SM section 5 (*11*, *32*)
Relative pathology of reinfections	Mild	–	(*8*)
Age-specific IFR (primary infections)	SARS MERS COVID-19	See Fig. 3	(*33*) (*5*) (*34*)

HCoVs in the long run. A critical prediction is that the severity of emergent HCoVs once they reach endemicity depends only on the severity of infection in children (Fig. 3), because all available evidence suggests that immunity to HCoVs has short IE_S and moderate IE_I, leading to frequent reinfection throughout adulthood (*11*, *12*), but strong IE_P such that childhood infection provides protection from pathology upon reinfection in adulthood, as evidenced by the rarity of severe infections or detectable IgM titers in adults. Strain-specific virulence factors, such as the shared cellular receptor, angiotensin-converting enzyme 2 (ACE-2), to which SARS-CoV-1, SARS-CoV-2, and the endemic strain NL63 all bind (*13*–*16*), may affect the CFR during the emergence phase but have little impact on the severity of disease in the endemic phase. Because the four endemic HCoVs have been globally circulating for a long time and almost everyone is infected at a young age, we cannot ascertain how much pathology would result from a primary or even a secondary case of any of these in an elderly or otherwise vulnerable person.

The key insights come from how our model explicitly incorporates different components of immunological protection with respect to susceptibility, pathology, and infectivity (IE_S, IE_P, and IE_I, respectively) and their different rates of waning. In our analysis, we hypothesized that these components of immunity for SARS-CoV-2 are comparable to those of endemic HCoVs, and this needs to be determined. Additionally, during the transition to endemicity, we need to consider how the immune efficacies depend on primary and secondary infections across ages (*17*) and how responses differ between vaccination and natural infection.

Longitudinal analysis of SARS patients provides an opportunity to measure the durability of immune memory in the absence of reexposure. The only long-term study we know of that follows SARS-CoV-1–specific antibodies suggests that they wane faster than antibodies to other live viruses and vaccines such as measles, mumps, rubella, and smallpox (*18*) and fall below the threshold of detection in 6 years (*19*). In contrast to antibody responses, memory T cells persist for much longer periods (*19*, *20*) and confer protection in animal model systems (*21*).

We further consider the effects of strain variation both for natural infection and vaccination. Strain variation and antibody escape may occur in endemic strains (*22*); however, the fact that symptoms are mild suggests that immunity induced by previously seen strains is nonetheless strong enough to prevent severe disease. Indeed, among HCoVs, frequent reinfections appear to boost immunity against related strains (*12*). However, the effect of strain variation may differ for vaccine-induced immunity, especially in light of the narrower epitope repertoire of many currently authorized vaccines.

If frequent boosting of immunity by ongoing virus circulation is required to maintain protection from pathology, then it may be best

for the vaccine to mimic natural immunity insofar as preventing pathology without blocking ongoing virus circulation. Preliminary results suggest the adenovirus-based vaccine is better at preventing severe than mild or asymptomatic infections (*23*), and it will be important to collect similar data for the other vaccines. Should the vaccine cause a major reduction in transmission, it might be important to consider strategies that target delivery to older individuals for whom infection can cause higher morbidity and mortality, while allowing natural immunity and transmission to be maintained in younger individuals. During the transition to endemicity, primary SARS-CoV-2 infections will frequently occur in older individuals, and we need to determine whether immunity induced by infection or vaccination in adulthood is similar to that produced by natural infections in childhood. Thus far, there have been few reinfections reported with SARS-CoV-2, and disease severity has varied (*24*); the only population-level study of reinfection that we are aware of estimates a low rate of reinfection in the first 6 months after primary infection and mild disease upon reinfection (*25*), but further analysis and monitoring are vital.

The findings presented here suggest that using symptoms as a surveillance tool to curb the spread of SARS-CoV-2 will become more difficult, as milder reinfections increasingly contribute to chains of transmission and population-level attack rates. In addition, infection or vaccination may protect against disease but not provide the type of transmission-blocking immunity that allows for shielding (*26*) or the generation of long-term herd immunity (*2*).

The details of the change in overall IFR through the transient period will be affected by a wide array of factors, such as age-specific human contact rates (*27*) and susceptibility to infection (*28*) as well as improvement in treatment protocols, hospital capacity, and virus evolution. The qualitative result of mild disease in the endemic phase is robust to these complexities, but quantitative predictions for the transient phase will depend on a careful consideration of these realities and how they interact with the dynamics of infection and components of immunity (*29*).

The changes in the IFR over time predicted by the model have implications for vaccination strategy against current and future emerging HCoVs. Social distancing and an effective vaccine are critical for control during a virgin epidemic and the transition out of it, but once we enter the endemic phase, mass vaccination may no longer be necessary. The necessity for continual vaccination will depend on the age-dependence of the IFR. If primary infections of children are mild (as for SARS-CoV-1 and SARS-CoV-2), continued vaccination may not be needed as primary cases recede to mild childhood sniffles. If, on the other hand, primary infection in children is severe (as for MERS), then vaccination of children will need to be continued.

From an ecological and evolutionary perspective, our study opens the door to questions regarding the within-host and between-host dynamics of human immunity and pathogen populations in the face of immune efficacies with different kinetics. It also opens the question of how these immune efficacies interplay with strain cross-immunity, which is likely relevant within the alpha- and betacoronaviruses. Considering data and model predictions from emergence through endemicity of HCoVs revealed a framework for understanding immunity and vaccination that may apply to a variety of infections, such as respiratory syncytial virus and seasonal influenza, which share similar age distributions and immune responses.

REFERENCES AND NOTES

1. S. Cobey, *Science* **368**, 713–714 (2020).
2. H. E. Randolph, L. B. Barreiro, *Immunity* **52**, 737–741 (2020).
3. C. Fraser, S. Riley, R. M. Anderson, N. M. Ferguson, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 6146–6151 (2004).
4. S. Su *et al*., *Trends Microbiol.* **24**, 490–502 (2016).
5. M. Salamatbakhsh, K. Mobaraki, S. Sadeghimohammadi, J. Ahmadzadeh, *BMC Public Health* **19**, 1523 (2019).
6. S. Ruan, *Lancet Infect. Dis.* **20**, 630–631 (2020).
7. M. E. Halloran, I. M. Longini, C. J. Struchiner, *Design and Analysis of Vaccine Studies* (Statistics for Biology and Health, Springer, 2010).
8. K. A. Callow, H. F. Parry, M. Sergeant, D. A. J. Tyrrell, *Epidemiol. Infect.* **105**, 435–446 (1990).
9. A. F. Bradburne, M. L. Bynoe, D. A. Tyrrell, *BMJ* **3**, 767–769 (1967).
10. W. Zhou, W. Wang, H. Wang, R. Lu, W. Tan, *BMC Infect. Dis.* **13**, 433 (2013).
11. M. Galanti, J. Shaman, *J. Infect. Dis.* jiaa392 (2020).
12. A. W. D. Edridge *et al*., *Nat. Med.* **26**, 1691–1693 (2020).
13. R. Yan *et al*., *Science* **367**, 1444–1448 (2020).
14. K. Kuba *et al*., *Nat. Med.* **11**, 875–879 (2005).
15. Y. Imai *et al*., *Nature* **436**, 112–116 (2005).
16. T. S. Fung, D. X. Liu, *Annu. Rev. Microbiol.* **73**, 529–557 (2019).
17. S. P. Weisberg *et al*., *Nat. Immunol.* **22**, 25–31 (2021).
18. A. Antia *et al*., *PLOS Biol.* **16**, e2006601 (2018).
19. F. Tang *et al*., *J. Immunol.* **186**, 7264–7268 (2011).
20. N. Le Bert *et al*., *Nature* **584**, 457–462 (2020).
21. R. Channappanavar, C. Fett, J. Zhao, D. K. Meyerholz, S. Perlman, *J. Virol.* **88**, 11034–11044 (2014).
22. R. Eguia *et al*., bioRxiv 2020.12.17.423313 [Preprint]. 18 December 2020. https://doi.org/10.1101/2020.12.17.423313.
23. M. Voysey *et al*., *Lancet* S0140-6736(20)32661-1 (2020).
24. European Centre for Disease Prevention and Control, "Reinfection with SARS-CoV: Considerations for public health response" (ECDC, 2020); www.ecdc.europa.eu/sites/default/files/documents/Re-infection-and-viral-shedding-threat-assessment-brief.pdf.
25. L. J. Abu-Raddad *et al*., *Clin. Infect. Dis.* 10.1093/cid/ciaa1846 (2020).
26. J. S. Weitz *et al*., *Nat. Med.* **26**, 849–854 (2020).
27. J. Wallinga, P. Teunis, M. Kretzschmar, *Am. J. Epidemiol.* **164**, 936–944 (2006).
28. J. Zhang *et al*., *Science* **368**, 1481–1486 (2020).
29. C. M. Saad-Roy *et al*., *Science* **370**, 811–818 (2020).
30. X. He *et al*., *Nat. Med.* **26**, 672–675 (2020).
31. S. Sanche *et al*., *Emerg. Infect. Dis.* **26**, 1470–1477 (2020).
32. S. M. Kissler, C. Tedijanto, E. Goldstein, Y. H. Grad, M. Lipsitch, *Science* **368**, 860–868 (2020).
33. M. Chan-Yeung, R.-H. Xu, *Respirology* **8** (suppl.), S9–S14 (2003).
34. R. Verity *et al*., *Lancet Infect. Dis.* **20**, 669–677 (2020).
35. J. S. Lavine, COVID Immunity and Endemicity Code, Version 1.0.0, Zenodo (2020); https://doi.org/10.5281/zenodo.4390505.

ACKNOWLEDGMENTS

We gratefully acknowledge helpful conversations with H. Ahmed, V. Zarnitsyna, and K. Fuller. **Funding:** This work was funded by NIH grants U01 AI150747, U01 HL139483, and U01 AI144616. **Author contributions:** J.S.L.: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing – original draft, Writing – review and editing, and Visualization. R.A.: Conceptualization, Methodology, Writing, Visualization, Supervision, and Funding acquisition. O.N.B.: Conceptualization, Methodology, and Writing – review and editing. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The scripts and data used to perform the analysis and generate the figures in this paper are available on GitHub (https://github.com/JennieLavine/covid-immunity-endemicity) and archived in Zenodo (*35*). This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/371/6530/741/suppl/DC1
Supplementary Text
Figs. S1 to S13
References (*36*–*42*)
MDAR Reproducibility Checklist

4 September 2020; resubmitted 10 November 2020
Accepted 7 January 2021
Published online 12 January 2021
10.1126/science.abe6522

Sample Evaporator for Microplates

Porvair Sciences announces Ultravap Mistral—an automation-ready sample evaporator that offers throughput advantages to laboratories looking to optimize and accelerate sample preparation. The Mistral directly and consistently delivers heated gas up to 80°C in each microplate well or tube, facilitating speedy, convenient evaporation of most common chromatography solvents, including dichloromethane, methanol, acetonitrile, hexane, and water. The option for straight or spiral needles allows users to choose between faster drydown (spiral) and better final drying in V-well plates (straight). Highly intuitive software and up to 15 easy-to-use, stored multistep evaporation programs enable even occasional users to gain the full benefits of this unit. For regular users, the Mistral offers the versatility of fully flexible programming, for example, providing for the ideal rate of evaporation for each solvent type.

Porvair Sciences
For info: +44-(0)-1978-666222
www.microplates.com/blowdown-evaporator-ultravap-mistral

Cell-Culture Analyzer

Nova Biomedical announces the addition of a Sample Retain Collector (SRC) for the BioProfile FLEX2 cell-culture analyzer. The FLEX2 automated analyzer offers comprehensive analysis of up to 16 key parameters, including pH, gases, metabolites, osmolality, cell density, and cell viability. A single FLEX2 with the SRC and the previously introduced Online Autosampler (OLS) module provides automated sampling and analysis of these fundamental cell-culture chemistries from as many as 10 bioreactors—and storage by the SRC in as fast as 1 h. This automation package saves hours of time spent on manual sampling, analysis, sample storage, and after-hours cell culture monitoring. The SRC automatically collects cell-culture samples from the FLEX2 OLS and stores them in a refrigerated environment to fulfill regulatory requirements for long-term sample retains, also enabling further offline testing. The SRC allows user-selectable retained sample volumes from 200 µL to 50 mL at a storage temperature of 4°C.

Nova Biomedical
For info: 800-822-0911
www.novabiomedical.com

UV Light Box

CS Medical is pleased to announce its distribution agreement with AirClean Systems to offer the AirClean UV Light Box. The growing demand to decontaminate N95 respirators has made shortwave UV light—used for years to decontaminate surfaces—an alternative to other chemical-based methods. Designed to protect the user from exposure to potentially harmful shortwave light energy, the AirClean UV Light Box is available in two widths and performs decontamination of N95 respirators in a total cycle time of 60 min, 30 min per side. Decontamination will make the masks reusable up to five times, helping alleviate the PPE shortage that is common in so many health care facilities and other industries across the nation right now.

CS Medical
For info: 919-255-9472
www.csmedicalllc.com/products/uv-light-box

Microplate Sealing for ADME Screening

BioChromato reports that pharmaceutical companies undertaking absorption, distribution, metabolism, and excretion (ADME) studies are benefiting from integrating RAPID Easy Piercing Seals (EPS) into their screening protocols. ADME scientists commonly use 96- or 384-well microplates to store large numbers of samples for screening. Sample-contamination issues can often arise in ADME studies, as common HPLC solvents such as acetonitrile, water, and dimethyl sulfoxide can extract siloxane out of the silicon-based adhesives used in the microplate seal. RAPID EPS seals use a synthetic-rubber adhesive to create a high-integrity, airtight microplate seal, preventing contamination of ADME samples analyzed by HPLC. In addition, the seals leave no particulate material when pierced by an HPLC autosampler, further safeguarding samples from contamination and eliminating damage to or clogging of your autosampler. They are proven to offer dependable microplate sealing over a working temperature range of −80°C to 80°C.

BioChromato
For info: +81-(0)-466-23-8382
biochromato.com/plate-and-seals/eps

Motorized Optical Stand

The OpenStand modular platform allows for easy customization. This fully configurable, motorized optical stand, when combined with a range of readily available optics, light sources, and accessories, creates a complete, customizable optical microscope. It is ideal for optogenetics, physiology, electrophysiology, neuroscience, industrial, and general imaging applications. Its modular approach allows maximum interchangeability and flexibility, enabling users to image a wide range of samples for virtually any life science and industrial application, and offering the largest imaging space available. It features a cost-effective, custom development platform that can be set up quickly as a fast-track to a prototype instrument. OpenStand lets you select only the components you need for your specific application, resulting in significant savings while giving you the flexibility to expand and add additional components if requirements change.

Prior Scientific
For info: 781-878-8442
www.prior.com

Stem-Cell Culture Substrate

FUJIFILM Irvine Scientific announces cellnest, a recombinant peptide attachment substrate that provides optimal adhesion and proliferation of stem cells in chemically defined, animal component–free conditions. Attachment substrates mimic the extracellular matrix, a complex, dynamic environment in which cells reside in vivo, and allow for adhesion, expansion, and potential differentiation of stem cells. Unlike animal-derived components, which can introduce unpredictability in results, cellnest delivers consistent results and can smooth the regulatory path to commercialization. It is compatible with any adherent cell type that binds to the Arg-Gly-Asp (RGD) domain, an amino-acid sequence within the extracellular matrix protein fibronectin that mediates cell attachment. cellnest is an ideal companion product to our PRIME-XV portfolio of xeno-free, chemically defined media for stem-cell culture and is well suited for the attachment and growth of mesenchymal stem cells.

FUJIFILM Irvine Scientific
For info: 800-577-6097
www.irvinesci.com/products/1063967-cellnest-recombinant-peptide

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

Stanford MEDICINE

DEPARTMENT OF CARDIOTHORACIC SURGERY

The Stanford Department of Cardiothoracic Surgery seeks to recruit **a scholar with an interest in studying cardiothoracic surgical diseases and their therapies**. Areas of investigation would include but not be limited to: atherosclerosis, genetics, molecular and cellular biology, metabolism/myocardial protection, transplant immunology, hematology/blood-prosthesis surface interface, neuroprotection, biomechanics, engineering, device development, thoracic cancer biology, cardiac development, regenerative medicine, translational therapies, and other areas.

The ideal candidate will possess PhD, MD/PhD, or other advanced academic degrees, have completed rigorous training, exemplify innovation, and demonstrate the potential for a successful track record of publishing. Appropriate resources will be made available to help the applicant develop their research program. The applicant will be expected to evolve a funded research effort focused in their area of expertise and to participate in mentorship and teaching activities in the Department of Cardiothoracic Surgery. Surgeon-scientists are also encouraged to apply and will have appropriate clinical responsibilities and resources provided.

Depending upon experience and focus, the applicant will be appointed in the University Tenure Line, the Non-Tenure Research Line, or the Medical Center Line at the Assistant, Associate, or Professor rank. The predominant criterion for appointment in the University Tenure Line is a major commitment to scholarship and teaching. The predominant criterion for appointment in the Non-Tenure Research Line is evidence of high-level performance as a researcher for whose special knowledge a programmatic need exists. The predominant criterion for appointment in the Medical Center Line is a major commitment to demonstrate excellence in the overall mix of scholarship, clinical care, and teaching.

Stanford is an equal employment opportunity and affirmative action employer. All qualified applicants will receive consideration for employment without regard to race, color, religion, sex, sexual orientation, gender identity, national origin, disability, veteran status, or any other characteristic protected by law. Stanford is committed to increasing the diversity of its faculty. It welcomes nominations of and applications from women and members of minority groups, veterans, and individuals with disabilities, as well as others who would bring additional dimensions to the university's research, teaching and clinical missions.

Applicants should send their curriculum vitae and a very brief letter outlining their interests electronically to Ms. Corrine Sanchez (**corrine.sanchez@stanford.edu**) or via regular mail to:
Sean Wu, MD, PhD
Search Committee Chair
c/o Corrine Sanchez
CVRB, Falk Bldg., Mail Code 5407
300 Pasteur Drive
Stanford, CA 94305-5407

KENNESAW STATE UNIVERSITY

CHAIR OF THE DEPARTMENT OF ECOLOGY, EVOLUTION, AND ORGANISMAL BIOLOGY

The College of Science and Mathematics at Kennesaw State University invites applications for **Chair of the Department of Ecology, Evolution, and Organismal Biology (EEOB)**, beginning **July 1, 2021**. The Chair is a member of the academic leadership team in the College of Science and Mathematics and is expected to advocate for EEOB with an open, and participatory administrative style.

The successful candidate must possess an earned doctorate (or equivalent) in Organismal Biology (or closely related field), and a distinguished record that merits appointment as a full professor with tenure in the EEOB department. Preference will be given to those candidates who can show: evidence of effective leadership; exceptional communication and interpersonal skills; an ability to work with faculty, staff, and students from diverse backgrounds; a record of scholarship that includes peer-reviewed publications and extramural funding; a record of excellence in teaching; and a demonstrated commitment to promoting inclusive excellence in teaching and research. For a full description of this position, application deadlines, and application procedures, visit **https://facultyaffairs.kennesaw.edu/faculty_openings.php.**

Kennesaw State University is an Equal Opportunity/Affirmative Action employer and does not discriminate based on age, color, disability, national origin, race, religion, sex, sexual orientation, and/or veteran status. Women, minorities, individuals with disabilities and veterans are encouraged to apply. Georgia is an Open Records state.

KENNESAW STATE UNIVERSITY

CHAIR OF THE DEPARTMENT OF MOLECULAR AND CELLULAR BIOLOGY AT KENNESAW STATE UNIVERSITY

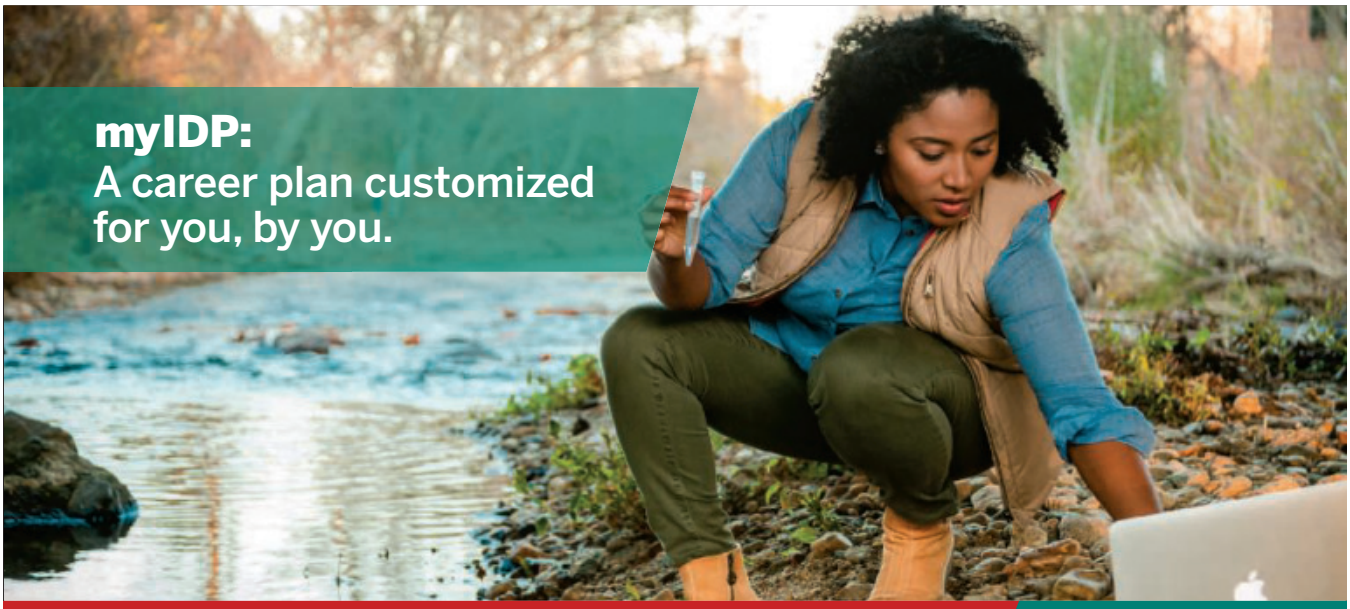
The College of Science and Mathematics at Kennesaw State University seeks a visionary and effective leader to serve as **Chair of the Department of Molecular and Cellular Biology**. This twelve-month position begins on **July 1, 2021**, and the successful applicant must meet the requirements of a full professor. The ideal candidate will have a distinguished record in a field that interfaces with the Department's research programs, a strong history of extramural funding, records of excellence in teaching and professional service, and a demonstrated commitment to shared governance, collaborative decision-making and the incorporation of inclusive excellence principles in teaching, research, and administrative operations in the department.

The successful candidate will be expected to maintain an extramurally funded research program in molecular, cellular, developmental or microbial biology. Primary responsibilities of the Chair include recruiting and managing faculty and staff, managing department budget and resources, providing oversight of academic programs, and implementing a plan designed to strengthen KSU as an R2 institution. For a full description of this position, application deadlines, and application procedures, visit **https://facultyaffairs.kennesaw.edu/faculty_openings.php**. Informal inquiries can be made to Dr. Sherry Ni (**sni@kennesaw.edu**).

Kennesaw State University is an Equal Opportunity/Affirmative Action employer and does not discriminate based on age, color, disability, national origin, race, religion, sex, sexual orientation, and/or veteran status. Georgia is an Open Records state.

Science Careers online @sciencecareers.org

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 In partnership with:

FASEB Federation of American Societies for Experimental Biology
UMASS University of Massachusetts Medical School
UCSF University of California San Francisco
UIC
BURROUGHS WELLCOME FUND

Science Careers helps you advance your career. Learn how!

- Register for a free online account on **ScienceCareers.org**.
- Search hundreds of job postings and find your perfect job.
- Sign up to receive e-mail alerts about job postings that match your criteria.
- Upload your resume into our database and connect with employers.
- Watch one of our many webinars on different career topics such as job searching, networking, and more.

Visit **ScienceCcareers.org** today — all resources are free

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

SCIENCECAREERS.ORG

Neuroscience Postdoctoral Programme

Linköping University is one of Sweden's six large universities, currently enrolling 27,000 students. The Centre for Systems Neurobiology involves some 50 independent research groups, from the Faculty of Medicine and the Faculty of Science and Engineering, as well as the University Hospital.

The Centre for Systems Neurobiology is now seeking Postdoctoral Fellows within several neuroscience research areas: Animal behavior/Behavioral genetics (Urban Friberg, Per Jensen, Dominic Wright), Psychiatric neuroscience and addiction (Markus Heilig), Neurodegenerative disease (Martin Hallbeck, Dan Lindholm, Walker Jackson), Neuroendocrinology/ neuroimmunology (Anders Blomqvist, David Engblom), Neurophysiology and clinical neuromodulation (Magnus Thordstein), Sensory and cognitive systems, and pain (Anders Fridberger, Kajsa Igelström, Håkan Olausson, Marcin Szczot, Susanna Walter).

For full consideration please apply as soon as possible, and no later than March 5.

LINKÖPING UNIVERSITY

For more details regarding the Centre, the different research labs involved in the programme, and to submit a letter-of- intent please visit: liu.se/en/research/center-for-systems-neurobiology. For information regarding the university and the region, please visit: liu.se, eastsweden.com.

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 **Ryoichi Fujiwara**

Balancing act

"Enjoy your life," the doctor told me. "Don't worry about work." I had recently moved to Germany for a visiting scholar position and was hoping to get a refill of my prescription sleep aids, which I had been taking for more than a year. The doctor spent nearly an hour with me, listening to my story. Then, to my surprise, he said he wasn't going to refill my prescription. Instead of relying on pills, he said, I needed to rethink my approach to life and work. I felt I was losing my life support—but I would soon learn that he was right.

I had spent the previous 5 years as an assistant professor in Tokyo. I worked overtime and on weekends—running a lab, writing grant proposals, and teaching. It was challenging but fun. I thought everything was going great.

Then, my body started to fall apart. I couldn't sleep at night. I never felt hungry, even when I skipped meals. I was already skinny, and after losing my appetite and 22 pounds, I was skin and bones. But the doctors couldn't identify a cause for my symptoms; they could only offer prescription sleep aids.

My sleep improved, but my appetite did not. Yet I was still optimistic. I would be fine, I thought. My body was sure to go back to normal at some point. Instead, my health continued to deteriorate. I even had to go to the emergency room for severe abdominal pain.

"I needed to rethink my approach to life and work."

I figured a change of scenery and work routine might help. I remembered that a researcher I had met at a conference was looking for someone to fill a yearlong visiting scholar position in Germany, which seemed like a good fit. I decided to temporarily leave my post in Japan—maintaining some teaching and advising responsibilities, which I managed remotely—and I moved to Germany.

It wasn't the cure. So, when the medication I had brought with me from Japan ran out, I went in for that fateful doctor's visit. Although I wasn't sure about his advice, I decided to give it a try. I stopped working overtime and began to spend more time outside. I even became a runner. After a few months, while training for my first marathon, I felt something I couldn't identify. Then I realized: I was hungry. I stopped running right there in the middle of the park, overcome with excitement and relief. I also noticed that I was sleeping at night without the sleep aids. My body was recovering.

As the visiting scholar position was ending, I thought about my career. I could return to my job in Japan, but I knew the toll that hectic life would take on me. I had no luck getting a faculty position elsewhere, but a professor in Arkansas offered me a research associate position. It would be a step down in some ways, but the decision was easy; maintaining a healthy work-life balance was more important. Plus, I had enjoyed working in the United States as a postdoc. Flying to Arkansas, I felt no frustration or regret, only excitement.

I've been in Arkansas for a few years now, and I'm doing really well. I have run 10 marathons and am currently training for an Ironman race, and I'm satisfied with my career. I was recently recruited for a nontenure, teaching-focused assistant professor role at my university. I felt a little hesitant about taking a faculty job again, but I can still make time for running, biking, and swimming. I also make sure to limit my working time to 8 hours a day, 40 hours a week. Sometimes it's tempting to work longer, but now I know I shouldn't.

I study and teach about pharmaceuticals, so it may seem ironic that giving them up was a turning point for me. Medications are important for many conditions, and no one should feel bad about taking them to maintain their physical or mental health. But in my case, I think the medicine was a Band-Aid for a deeper issue. I'll never know for sure what helped get my body back to normal—maybe becoming a runner, or maybe the German air and food. Regardless, I'm grateful for the doctor who helped me see that I should focus more on enjoying my life, and I am thankful that I have been able to adjust my career to do so. ■

Ryoichi Fujiwara is an assistant professor in the College of Pharmacy at the University of Arkansas for Medical Sciences in Little Rock. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

Science Webinars

AAAS

Science Webinars help you keep pace with emerging scientific fields!

Stay informed about scientific breakthroughs and discoveries.

Gain insights into current research from top scientists.

Take the opportunity to ask questions during live broadcasts.

Get alerts about upcoming free webinars.

Sign up at: webinar.sciencemag.org/stayinformed

ScienceImmunology.org

READY TO PUT THE SPOTLIGHT ON YOUR RESEARCH?

Submit your research:
cts.ScienceMag.org

Science Immunology
AAAS

Twitter: @SciImmunology
Facebook: @ScienceImmunology