Should deaf kids with hearing devices use sign language? p. 1138

Sundance films grapple with science themes p. 1142

A close-up view of parasite flagella pp. 1162 & 1163

Science

$15
14 MARCH 2025
science.org
AAAS

INTERSTELLAR DUST

Mapping dust properties in the Milky Way p. 1209

NOMIS & Science
YOUNG EXPLORER AWARD 2025

Research at the intersection of the social and life sciences

Unconventional. Interdisciplinary. Bold.

The NOMIS & *Science* Young Explorer Award recognizes and rewards early-career M.D., Ph.D., or M.D./Ph.D. scientists who perform research at the intersection of the social and life sciences. Essays written by these bold researchers on their recent work are judged for clarity, scientific quality, creativity, and demonstration of cross-disciplinary approaches to address fundamental questions.

A cash prize of up to USD 15,000 will be awarded to essay winners, and their engaging essays will be published in *Science*. Winners will also be invited to share their work and forward-looking perspective with leading scientists in their respective fields at an award ceremony.

Apply by May 15, 2025
at www.science.org/nomis

NOMIS FOUNDATION
Creating the Spark

Science
AAAS

CONTENTS

14 MARCH 2025 • VOLUME 387 • ISSUE 6739

1138

NEWS

IN BRIEF

IN DEPTH

FEATURES

INSIGHTS

BOOKS *ET AL.*

1142

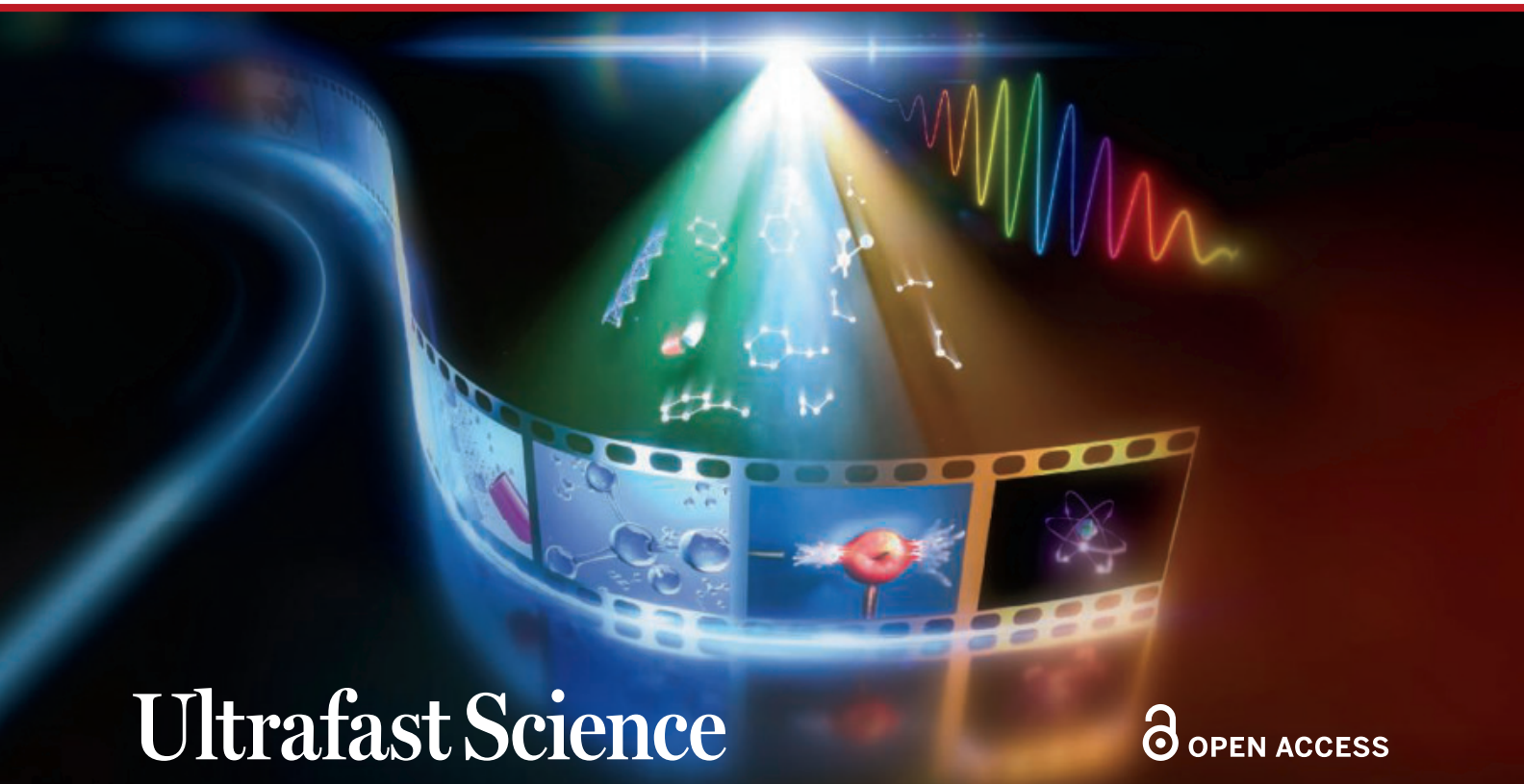
PERSPECTIVES

POLICY FORUM

LETTERS

CALL FOR PAPERS

SPJ
SCIENCE PARTNER JOURNALS



Ultrafast Science

OPEN ACCESS

Ultrafast Science is an online-only, Open Access journal published in affiliation with **Xi'an Institute of Optics and Precision Mechanics (XIOPM)** of CAS and distributed by the **American Association for the Advancement of Science (AAAS)**. The mission of the journal is to build a platform for ultrafast scientific research; present new theories, concepts, ideas, technologies and progress in this field; publish the latest research representing the forefront of the discipline and that of international concern; and promote academic exchange and development worldwide. *Ultrafast Science* publishs high-quality original research articles, comprehensive reviews, editorials, and perspectives which feature high novelty, significance and technical quality in cutting-edge and emerging topics in ultrafast science with broad interest from scientific communities.

Submit your research to *Ultrafast Science* today!

Learn more at **spj.science.org/ultrafastscience**

The Science Partner Journal (SPJ) program was established by the American Association for the Advancement of Science (AAAS), the nonprofit publisher of the *Science* family of journals. The SPJ program features high-quality, online-only, Open-Access publications produced in collaboration with international research institutions, foundations, funders and societies. Through these collaborations, AAAS furthers its mission to communicate science broadly and for the benefit of all people by providing top-tier international research organizations with the technology, visibility, and publishing expertise that AAAS is uniquely positioned to offer as the world's largest general science membership society.

Visit us at **spj.science.org**

@UltrafastSci **@SPJournals** **@SPJournals**

ARTICLE PROCESSING CHARGES WAIVED UNTIL JANUARY 2026

1134 & 1181
Russian military destruction of the Kakhovka Dam in Ukraine caused considerable impacts to downstream ecosystems.

PHOTO: IVAN ANTIPENKO

RESEARCH

IN BRIEF

RESEARCH ARTICLES

Microbiology

DEPARTMENTS

ON THE COVER

An optical image mosaic of the Milky Way galaxy shows a view looking toward the Galactic Center. The smooth band of starlight is occluded by dark clouds of interstellar dust, which absorbs and scatters background light (extinction), causing distant stars to appear fainter and redder. Variations in the properties of this dust extinction have been mapped in three dimensions using 130 million stellar spectra. See page 1209. *Credit: ESO/Serge Brunier, CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/)*

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. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $2865; Foreign postage extra: Air assist delivery: $135. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling available from backissues.science.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

Science AAAS

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

Executive Editor Valda Vinson

Editor, Research Jake S. Yeston **Editor, Insights** Lisa D. Chong **Deputy Executive Editor** Lauren Kmec

DEPUTY EDITORS Stella M. Hurtley (UK), Phillip D. Szuromi, Sacha Vignieri **SR. EDITORS** Caroline Ash (UK), Michael A. Funk, Angela Hessler, Di Jiang, Priscilla N. Kelly, Marc S. Lavine (Canada), Sarah Lempriere (UK), Mattia Maroso, Yevgeniya Nusinovich, Ian S. Osborne (UK), L. Bryan Ray, H. Jesse Smith, Keith T. Smith (UK), Jelena Stajic, Peter Stern (UK), Yury V. Suleymanov, Valerie B. Thompson, Brad Wible **ASSOCIATE EDITORS** Jack Huang, Sumin Jin, Bianca Lopez, Sarah Ross (UK), Madeleine Seale (UK), Corinne Simonti, Ekeoma Uzogara **SENIOR LETTERS EDITOR** Jennifer Sills **NEWSLETTER EDITOR** Christie Wilcox **RESEARCH & DATA ANALYST** Jessica L. Slater **LEAD CONTENT PRODUCTION EDITORS** Chris Filiatreau, Harry Jach **SR. CONTENT PRODUCTION EDITOR** Amelia Beyna **CONTENT PRODUCTION EDITORS** Anne Abraham, Robert French, Julia Haber-Katris, Nida Masiulis, Abigail Shashikanth, Suzanne M. White **SR. EDITORIAL MANAGER** Beverly Shields **SR. PROGRAM ASSOCIATE** Maryrose Madrid **EDITORIAL ASSOCIATES** Aneera Dobbins, Joi S. Granger, Lisa Johnson, Anita Wynn **SR. EDITORIAL COORDINATORS** Alexander Kief, Ronmel Navas, Jerry Richardson, Isabel Schnaidt, Alice Whaley (UK), Brian White **EDITORIAL COORDINATORS** Clair Goodhead (UK), Kat Kirkman, Samantha Price **ADMINISTRATIVE COORDINATOR** Karalee P. Rogers **ASI DIRECTOR, OPERATIONS** Janet Clements (UK) **ASI OFFICE MANAGER** Victoria Smith **ASI SR. OFFICE ADMINISTRATORS** Dawn Titheridge (UK), Jessica Waldock (UK) **COMMUNICATIONS DIRECTOR** Meagan Phelan **DEPUTY DIRECTOR** Matthew Wright **SENIOR WRITERS** Walter Beckwith, Joseph Cariz **WRITERS** Abigail Eisenstadt, Mahathi Ramaswamy **SENIOR COMMUNICATIONS ASSOCIATES** Zachary Graber, Sarah Woods **COMMUNICATIONS ASSOCIATES** Kiara Brooks, Haley Riley, Mackenzie Williams

News Editor Tim Appenzeller

NEWS MANAGING EDITOR John Travis **INTERNATIONAL EDITOR** David Malakoff **DEPUTY NEWS EDITORS** Rachel Bernstein, Shraddha Chakradhar, Martin Enserink, David Grimm, Eric Hand, Michael Price Kelly Servick, Matt Warren (Europe) **SR. CORRESPONDENTS** Daniel Clery (UK), Jon Cohen, Jeffrey Mervis **ASSOCIATE EDITORS** Jeffrey Brainard, Michael Greshko, Katie Langin **NEWS REPORTERS** Adrian Cho, Jennifer Couzin-Frankel, Phie Jacobs, Jocelyn Kaiser, Rodrigo Pérez Ortega (Mexico City), Robert F. Service, Erik Stokstad, Paul Voosen, Meredith Wadman **CONSULTING EDITOR** Elizabeth Culotta **CONTRIBUTING CORRESPONDENTS** Vaishnavi Chandrashekhar, Dan Charles, Warren Cornwall, Andrew Curry (Berlin), Ann Gibbons, Sam Kean, Kai Kupferschmidt (Berlin), Andrew Lawler, Mitch Leslie, Virginia Morell, Dennis Normile (Tokyo), Cathleen O'Grady, Elisabeth Pain (Careers), Charles Piller, Richard Stone (Senior Asia Correspondent), Gretchen Vogel (Berlin), Lizzie Wade (Mexico City) **INTERNS** Alexa Robles-Gil **COPY EDITORS** Julia Cole (Senior Copy Editor), Hannah Knighton, Cyra Master (Copy Chief) **ADMINISTRATIVE SUPPORT** Meagan Weiland

Creative Director Beth Rakouskas

DESIGN MANAGING EDITOR Chrystal Smith **GRAPHICS MANAGING EDITOR** Chris Bickel **PHOTOGRAPHY MANAGING EDITOR** Emily Petersen **MULTIMEDIA MANAGING PRODUCER** Kevin McLean **DIGITAL DIRECTOR** Kara Estelle-Powers **DESIGN EDITOR** Marcy Atarod **DESIGNER** Noelle Jessup **SENIOR SCIENTIFIC ILLUSTRATOR** Noelle Burgess **SCIENTIFIC ILLUSTRATORS** Austin Fisher, Kellie Holoski, Ashley Mastin **SENIOR GRAPHICS EDITOR** Monica Hersher **GRAPHICS EDITOR** Veronica Penney **SENIOR PHOTO EDITOR** Charles Borst **PHOTO EDITOR** Elizabeth Billman **SENIOR PODCAST PRODUCER** Sarah Crespi **SENIOR VIDEO PRODUCER** Meagan Cantwell **SOCIAL MEDIA STRATEGIST** Jessica Hubbard **SOCIAL MEDIA PRODUCER** Sabrina Jenkins **WEB DESIGNER** Jennie Pajerowski

Chief Executive Officer and Executive Publisher Sudip Parikh

Publisher, Science Family of Journals Bill Moran

DIRECTOR, BUSINESS SYSTEMS AND FINANCIAL ANALYSIS Randy Yi **DIRECTOR, BUSINESS OPERATIONS & ANALYSIS** Eric Knott **MANAGER, BUSINESS OPERATIONS** Jessica Tierney **MANAGER, BUSINESS ANALYSIS** Cory Lipman **BUSINESS ANALYSTS** Kurt Ennis, Maggie Clark, Isacco Fusi **BUSINESS OPERATIONS ADMINISTRATOR** Taylor Fisher **DIGITAL SPECIALIST** Marissa Zuckerman **SENIOR PRODUCTION MANAGER** Jason Hillman **SENIOR MANAGER, PUBLISHING AND CONTENT SYSTEMS** Marcus Spiegler **CONTENT OPERATIONS MANAGER** Rebecca Doshi **SENIOR CONTENT & PUBLISHING SYSTEMS SPECIALIST** Jacob Hedrick **SENIOR PRODUCTION SPECIALIST** Kristin Wowk **PRODUCTION SPECIALISTS** Kelsey Cartelli, Audrey Diggs **SPECIAL PROJECTS ASSOCIATE** Shantel Agnew

MARKETING DIRECTOR Sharice Collins **ASSOCIATE DIRECTOR, MARKETING** Justin Sawyers **GLOBAL MARKETING MANAGER** Allison Pritchard **ASSOCIATE DIRECTOR, MARKETING SYSTEMS & STRATEGY** Aimee Aponte **SENIOR MARKETING MANAGER** Shawana Arnold **MARKETING MANAGER** Ashley Evans **MARKETING ASSOCIATES** Hugues Beaulieu, Ashley Hylton, Lorena Chirinos Rodriguez, Jenna Voris **MARKETING ASSISTANT** Courtney Ford **SENIOR DESIGNER** Kim Huynh

DIRECTOR AND SENIOR EDITOR, CUSTOM PUBLISHING Erika Gebel Berg **ADVERTISING PRODUCTION OPERATIONS MANAGER** Deborah Tompkins **DESIGNER, CUSTOM PUBLISHING** Jeremy Huntsinger **SR. TRAFFIC ASSOCIATE** Christine Hall

DIRECTOR, PRODUCT MANAGEMENT Kris Bishop **PRODUCT DEVELOPMENT MANAGER** Scott Chernoff **PUBLISHING PLATFORM MANAGER** Jessica Loayza **ASSOCIATE DIRECTOR, PUBLISHING INTELLIGENCE** Rasmus Andersen **SR. PRODUCT ASSOCIATE** Robert Koepke **PRODUCT ASSOCIATES** Caroline Breul, Anne Mason

ASSOCIATE DIRECTOR, INSTITUTIONAL LICENSING MARKETING Kess Knight **ASSOCIATE DIRECTOR, INSTITUTIONAL LICENSING SALES** Ryan Rexroth **INSTITUTIONAL LICENSING MANAGER** Nazim Mohammedi, Claudia Paulsen-Young **SENIOR MANAGER, INSTITUTIONAL LICENSING OPERATIONS** Judy Lillibridge **MANAGER, RENEWAL & RETENTION** Lana Guz **SYSTEMS & OPERATIONS ANALYST** Ben Teincuff **FULFILLMENT ANALYST** Aminta Reyes

ASSOCIATE DIRECTOR, INTERNATIONAL Roger Goncalves **ASSOCIATE DIRECTOR, US ADVERTISING** Stephanie O'Connor **US MID WEST, MID ATLANTIC AND SOUTH EAST SALES MANAGER** Chris Hoag **DIRECTOR, OUTREACH AND STRATEGIC PARTNERSHIPS, ASIA** Shoupeng Liu **SALES REP, ROW** Sarah Lelarge **SALES ADMIN ASSISTANT, ROW** Victoria Glasbey **DIRECTOR OF GLOBAL COLLABORATION AND ACADEMIC PUBLISHING RELATIONS, ASIA** Xiaoying Chu **ASSOCIATE DIRECTOR, INTERNATIONAL COLLABORATION** Grace Yao **SALES MANAGER** Danny Zhao **MARKETING MANAGER** Kilo Lan **ASCA CORPORATION, JAPAN** Rie Rambelli (Tokyo), Miyuki Tani (Osaka)

DIRECTOR, COPYRIGHT, LICENSING AND SPECIAL PROJECTS Emilie David **RIGHTS AND PERMISSIONS ASSOCIATE** Elizabeth Sandler **LICENSING ASSOCIATE** Virginia Warren **RIGHTS AND LICENSING COORDINATOR** Dana James **CONTRACT SUPPORT SPECIALIST** Michael Wheeler

EDITORIAL
science_editors@aaas.org

NEWS
science_news@aaas.org

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

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

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

MEDIA CONTACTS
scipak@aaas.org

PRODUCT ADVERTISING & CUSTOM PUBLISHING
advertising.science.org
science_advertising@aaas.org

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

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

MEMBERSHIP AND INDIVIDUAL SUBSCRIPTIONS
science.org/subscriptions

MEMBER BENEFITS
aaas.org/membership/benefits

INSTITUTIONAL SALES AND SITE LICENSES
science.org/librarian

AAAS BOARD OF DIRECTORS
CHAIR Gilda A. Barabino
PRESIDENT Keith R. Yamamoto
PRESIDENT-ELECT Willie E. May
TREASURER Carolyn N. Ainslie
CHIEF EXECUTIVE OFFICER Sudip Parikh
BOARD Cynthia M. Beall, Janine Austin Clayton, Kaye Husbands Fealing, Kathleen Hall Jamieson, Jane Maienschein, Robert B. Millard, Babak Parviz, William D. Provine, Juan S. Ramírez Lugo, Susan M. Rosenberg, Vassiliki Betty Smocovitis

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

EDITORIAL

Uphold US-Canada science

A partnership can be demanding, and as with any couple, can have good days and bad. The United States–Canada relationship is most definitely having a bad one. It's difficult to fully comprehend all the dimensions of the current threats to one of the world's strongest, longest, and multifaceted alliances. From contemptuous musings on annexation to a tariff war that could wreak economic havoc on both sides of the border, the insults and aggravations are stoking uncertainty about a relationship that has flourished for decades. This includes a strongly intertwined connection between Canadian and American science—one that must continue in these challenging times.

The number one partner for Canadian science is by far the United States. For the past 5 years, 27% of all Canadian scientific publications were coauthored with American colleagues (according to a Canadian bibliometric database and the Web of Science). And the reverse is true as well. Canadian scientists are prominent international partners of American scientists in published research. Long-standing major programs between the two countries include joint research projects on the Great Lakes, the Arctic, space, health (including global public health), climate monitoring, artificial intelligence (AI), subatomic physics, and data sharing. Despite the uncertainty around tariffs, active partnerships have recently been reconfirmed and even extended between federal funding organizations in both countries. These include interactions between the US National Science Foundation and the Natural Sciences and Engineering Research Council of Canada as well as Canada's Social Science and Humanities Research Council. Such efforts are also strong at the regional level. For instance, research between Massachusetts and Québec focuses on climate change, biotechnology, and transportation, an alliance rooted in enduring cultural links.

> "...research between these neighbors at the moment is strong and should not be squandered..."

The exchange of trainees is a key component of a partnership that should persist. For decades, graduate students in Canada have continued training in the United States as postdoctoral fellows, and some have chosen to stay and forge fruitful collaborations with scientists in Canada. These have proven to be robust links, many of which are still active and prosperous today for both countries. American fellows coming to Canada to pursue their studies are not as numerous but are particularly interested in AI, quantum computing, clean energy, and environmental studies as well as the life sciences. Considering the current situation, it may be tempting for Canada to use the opportunity to lure both younger and well-established Canadian scientists back to Canada. Indeed, Canada is already receiving inquiries in that regard. But such efforts must not take precedence over drawing on the North American partnership more vigorously than before.

On both sides of the border, additional collaboration should focus on building capacity to advise elected officials and high-level policy-makers on scientific issues. Going further, the International Network for Governmental Science Advice (INGSA) and its 130 member countries, of which I am chair, aim to take on this challenge globally with three chapters in the Global South (Kuala Lumpur, Malaysia; Buenos Aires, Argentina; and Port Louis, Mauritius) as well as new European (Oxford, United Kingdom) and North American (Montreal, Canada) chapters that will be inaugurated over the next 2 years. A major objective is to increase the ability to offer advice not only at the national level but also to subregional and local officials who often must make critical decisions under emergency conditions.

Strengthening science diplomacy is more urgent than ever in North America and around the world. The American Association for the Advancement of Science (AAAS, the publisher of *Science*) and the United Kingdom's Royal Society have just released an updated framework on this topic as did the European Commission. In Québec, the Fonds de recherche du Québec launched a program this year to create new chairs in science diplomacy that will cultivate a network of experts across scientific disciplines throughout the province. The intent is to leverage the network to establish strong international science and policy partnerships.

Canada now has a new prime minister in place, and with the stability of US-Canada relations at stake, scientific partnerships should be upheld by the leaders of both nations. The current period is demanding, but research between these neighbors at the moment is strong and should not be squandered so that together, they can apply excellence in science to face major challenges for our societies—not just in North America but around the world.

–Rémi Quirion

Rémi Quirion
is the chief scientist of Québec, Canada; chief executive officer of Fonds de recherche du Québec; and president of the International Network for Governmental Science Advice, Auckland, New Zealand.
remi.quirion@frq.gouv.qc.ca

10.1126/science.adx2966

NEWS

IN BRIEF

Edited by
Jeffrey Brainard

Geneticist Francis Collins (center), a former director of the U.S. National Institutes of Health who retired from the agency last month, spoke at a pro-science rally in Washington, D.C.

PHOTO: KENT NISHIMURA/REUTERS

TRUMP TRACKER

Scientists rally against research cuts as new ones hit

PROTESTING FOR SCIENCE Thousands of researchers and their supporters demonstrated in front of the Lincoln Memorial in Washington, D.C., on 7 March against what they consider antiscience actions by President Donald Trump's administration. About 30 such protests, collectively called Stand Up for Science, were held that day across the United States, with additional ones in other countries. Participants demanded reversals of cuts in federally funded research projects, reinstatement of fired federal employees, preservation of diversity and inclusion in science, and ending government censorship of research.

VACCINE HESITANCY RESEARCH AXED The U.S. National Institutes of Health (NIH) this week began terminating dozens of research grants for studying why some people hesitate to receive vaccines and evaluating strategies that could encourage vaccine uptake. The cuts appear to be part of efforts to defund research not aligned with policies backed by Trump and Department of Health and Human Services (HHS) Secretary Robert F. Kennedy Jr., a noted vaccine skeptic. The vaccines in the canceled studies are for diseases such as COVID-19, chickenpox, human papillomavirus, and mpox. An NIH letter sent to these investigators, which *Science* has seen, says their award "no longer effectuates agency priorities." In addition, a source with direct knowledge told *Science* that NIH last week asked its institutes to list projects involving messenger RNA vaccines, which some vaccine skeptics think are unsafe because they believe, without evidence, that the vaccines could modify DNA or cause various health issues.

COLUMBIA GRANTS CUT The administration sent a shock wave through academe by announcing on 7 March that it had terminated $400 million in grants and contracts to Columbia University for its alleged lax response to anti-Israel protests on campus, an unprecedented sanction. NIH said on social media that this sum includes $250 million in more than 400 research awards from the agency—a sum representing more than one-third of total NIH funding received by Columbia in 2024. A list of the canceled grants was not immediately made public. The administration, which also alleged "persistent harassment of Jewish students" at Columbia, cited its authority to sanction recipients of federal funds for not preventing discrimination. On 10 March, the Department of Education says it warned 59 other universities they could lose grants for a similar reason.

SOME NIH FIRINGS BLOCKED At least 17 senior scientists at NIH facing termination last week had their terms extended 1 year. The reprieve followed an appeal from acting NIH Director Matthew Memoli to HHS. The employees are among a class of staff scientists, research fellows, and tenure-track investigators who make up the majority of NIH's nearly 4000 in-house scientists. NIH hired them under a statute, Title 42, that allows federal agencies to recruit talented scientists by paying salaries above federal norms, with their employment subject to periodic renewal. The Trump administration had halted the renewals. It is unclear whether others in this class who will soon reach the end of their term will be placed on leave without pay while NIH appeals for their retention.

INDIRECT COSTS PRESERVED The administration is unlikely to prevail if it seeks to overturn a U.S. federal judge's ruling last week that halts a proposed drastic reduction in overhead payments to universities receiving grants from NIH, say lawyers who have followed the case. In February, NIH set a flat rate of 15% for reimbursement of such indirect costs, which would deprive universities of billions of dollars for research buildings and shared lab equipment, for example. In a 5 March order blocking the cap's implementation, District Court Judge Angel Kelley ruled that the government ignored a congressional mandate not to change the current rates, now negotiated individually, as well as the rules that must be followed if NIH wants to change them.

QUIZZING SCIENTISTS The administration last week instructed federally funded researchers who work in other countries to disclose whether they work with communist governments and "combat Christian persecution," *The New York Times* reported. Some of the recipients are funded by the U.S. Centers for Disease Control and Prevention; some work on public health projects in Vietnam, for example. A questionnaire sent to the grantees ruled out spending tax dollars on "socialism" or "corrupt regimes that oppose free enterprise."

CENSUS, FOOD SAFETY ADVISERS AXED The Department of Commerce disbanded five outside panels that provided scientific and community advice to the U.S. Census Bureau and other federal statistical agencies, just as preparations are ramping up for the country's next decennial census, in 2030. The Trump administration also disbanded two long-standing scientific committees that for decades provided advice from academe and industry to the Department of Agriculture and other federal agencies about food safety, including microbial threats and meat and poultry inspections.

Acidity's toll on phytoplankton

CLIMATE SCIENCE | Ocean acidification levels expected by 2075 could reduce phytoplankton's ability to absorb atmospheric carbon dioxide (CO_2) by some 5 trillion kilograms per year, or 10%, researchers report. Scientists have long tried to gauge the effects of rising acidity—which stems from atmospheric CO_2 dissolved in water—on phytoplankton, marine microorganisms that grow by absorbing CO_2 and sunlight. The new estimate is based on the largest field study to date, conducted across vast stretches of the tropical and North Pacific oceans. The projected reduction could accelerate warming of the global climate, the research team reports this week in the *Proceedings of the National Academy of Sciences*.

Japan may end national council

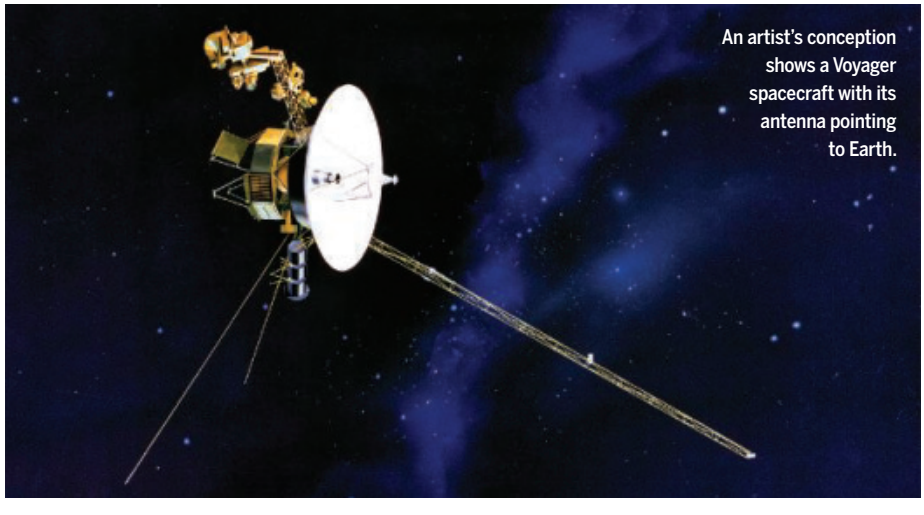
SCIENCE ACADEMIES | Japan's dominant Liberal Democratic Party (LDP) is out to crimp the independence of the Science Council of Japan, the country's national science academy, which represents 870,000 scholars from all academic disciplines. The LDP-led ruling coalition's cabinet last week approved a bill that would replace the council—an independent public entity that manages its own affairs—with a special corporation subject to increased government oversight. One point of conflict is the council's opposition to academics conducting research about military or dual-use technologies. Many researchers fiercely oppose the proposed change. A vote is expected within weeks.

More Voyager sensors turned off

SPACE EXPLORATION | NASA engineers say they switched off cosmic-ray–sensing telescopes aboard the Voyager 1 space probe last month and will shut off another instrument, measuring low-energy charged particles, on its twin, Voyager 2, later in March. The moves are part of a continuing effort to manage the gradually diminishing power supply of the probes, which were launched in 1977 and are now traveling in interstellar space, the only spacecraft ever to reach it. Both draw electricity from the heat of decaying plutonium. NASA plans to continue operating three of the original 10 science instruments on each spacecraft. The agency expects to shut off one more on each Voyager next year, hoping to prolong operation of the rest into the 2030s.

Russia's 'excellence' bid misfires

PUBLISHING | A program intended to bolster top Russian universities has by one measure backfired and spawned a heap of retractions, a study has found. Launched in 2013, the Russian University Excellence Initiative (RUEI) spent about $430 million over 8 years in a bid to help make scientists at 21 research-intensive universities competitive internationally. A major evaluation criterion was the number of articles authored by RUEI researchers that were indexed in the high-profile Web of Science and Scopus databases, which emphasize English-language content. The stakes were raised because authors of the indexed papers earned bonuses that could double their pay. An analysis, published online on 28 February in *Scientometrics*, identified 203 retracted papers from RUEI universities in the two databases—significantly more retractions than in a control group of non-RUEI Russian universities. The financial incentives to publish "may have encouraged unethical research practices and research misconduct," the study's authors conclude.



An artist's conception shows a Voyager spacecraft with its antenna pointing to Earth.

IMAGE: NASA/JPL

IN DEPTH

A researcher collects samples in a gentoo penguin colony in Antarctica.

INFLUENZA

H5N1 avian flu is spreading rapidly in Antarctica

Expedition finds the virus in 13 bird and seal species around the Antarctic Peninsula

By **Sofia Moutinho**

When the H5N1 avian influenza virus began killing seabirds on the Antarctic Peninsula 1 year ago, scientists wondered how fast the deadly pathogen would spread on the remote continent and how much damage it would do to its rich wildlife.

Now, they have some answers. In the past 6 weeks, the sailboat *Australis* has traveled along the shores of the peninsula, whose northern tip is just 650 kilometers from South America. Eight researchers dressed from head to toe in protective gear disembarked at 27 sites to swab animals and test carcasses. They found the virus in all but three locales, affecting a total of 13 bird and mammal species. "The virus has reached every corner [of the peninsula] and is infecting almost every animal species," says expedition leader Antonio Alcamí, a microbiologist with the Spanish National Research Council.

So far there is no evidence that the virus has spread beyond the peninsula, a tongue of land extending north from the continent that makes up less than 5% of its land mass. Surveillance elsewhere has been limited, however. Large parts of the mainland are rarely visited and most research stations there do not have the capability to test for the virus.

"The situation in Antarctica is a bit of a black box," says virus ecologist Michelle Wille of at the University of Melbourne, who was not involved in the expedition. "We're still very much in the learning phase, so this new data is really important for us to better understand what is possibly going on."

In October 2023, the highly pathogenic bird flu strain known as clade 2.3.4.4b—which has decimated poultry flocks and wild bird populations globally since 2021—jumped from South America to South Georgia and the South Sandwich Islands as well as the Falkland Islands in the subantarctic region. In February 2024, two skuas—large, gull-like predatory birds—found dead near the Argentine research station on the peninsula were the first confirmed cases on the continent itself. Testing at a few international research stations and a first expedition by the *Australis* team found another 58 infected animals, mainly in the northern part of the peninsula. But monitoring largely ended in March 2024, when the austral winter set in.

Since January, the *Australis* has been on a new Spanish-funded expedition. The ship, which has a lab that can detect viral DNA in samples, visited sites all along the peninsula's west coast and in the Weddell Sea, on the eastern side. Researchers collected cloacal and tracheal swabs from live animals and brain tissue from carcasses. Out of 846 samples, 188 tested positive, from nine bird and four seal species.

Most of the dead animals were skuas, likely because they feed on infected carcasses. Skua populations appeared to have dwindled at sites the team visited last year. Meagan Deward, a wildlife biologist at Federation University, says she, too, saw fewer skuas than usual while aboard a tourist vessel recently to collect samples. In the south of the peninsula, in contrast, the researchers saw evidence of an ongoing die-off: more skua carcasses and a wider variety of infected animal species, suggesting the virus has spread to these locations more recently.

One of the hardest hit areas was Armstrong Reef, a group of islets on the peninsula's west coast designated as an Important Bird and Biodiversity Area by BirdLife In-

PHOTO: ANTONIO ALCAMÍ

ternational because it supports a breeding colony of Adélie penguins and more than 500 pairs of Antarctic shags. Twenty-nine birds tested positive there, including some from those two species. Another severely affected location was Marguerite Bay, at the southern end of the peninsula, where 172 dead skuas were found.

Scientists fear wildlife in Antarctica is especially vulnerable to H5N1 because both birds and marine mammals breed there in dense colonies where multiple species intermingle, which helps the virus spread. Many species on the continent and the subantarctic region are not found anywhere else, making them more vulnerable to extinction. "One single island can hold 90% of the population of a species," says Marcela Uhart, a wildlife veterinarian at the University of California, Davis.

Still, the toll may be hard to gauge. "Unfortunately, we are not going to know the true impact for some of these species in Antarctica," Deward says. For most species, there are no reliable baseline population data. And assessing the mortality of marine animals is difficult, Alcamí notes. Seals, for example, can die at sea or on floating ice, leaving no carcasses behind. Last year, a local tourism operator reported seeing more than 100 dead crabeater seals on ice in the Weddell Sea; they weren't tested but likely died of H5N1, Alcamí says. This year, the *Australis* team found three carcasses on a nearby beach that tested positive. "The numbers we have are only part of the story. It is what we could see on a fast expedition," Alcamí says.

Antarctic penguins may be a bright spot. There were some suspected die-offs last year, but researchers have not seen abnormal mortality rates this season. That doesn't mean the birds are not infected. Using pumps that filter viruses from the air, the team found high concentrations of H5 in several colonies where penguins looked healthy. Alcamí speculates that they may have developed immunity as the virus swept through in the previous season.

Another wave of infections could hit the continent from the Indian Ocean. In October 2024, hundreds of king penguins and elephant seal pups were found dead from H5N1 in the subantarctic Crozet and Kerguelen islands, equidistant from South Africa and Australia in the southern Indian Ocean. The archipelagos could serve as a springboard to other parts of Antarctica, but also to Australia, now the world's only region free of H5N1 in the wild. "It's a massive range expansion of the virus," Wille says, "which is very, very concerning and shows that it has the capacity to spread more." ■

Sofia Moutinho is a science journalist in Rio de Janeiro.

TRUMP ADMINISTRATION

No sign that vaccine advisers are beholden to industry

RFK Jr. has claimed conflicts of interest must be rooted out of key CDC panel

By **Meredith Wadman**

In his maiden speech to department employees on 18 February, Health and Human Services Secretary Robert F. Kennedy Jr. made a vow about advisers to HHS agencies: "We will remove conflicts of interest on the committees and research partners whenever possible."

Since then, reports have surfaced that Kennedy, who has campaigned for decades against generally safe and effective vaccines, is preparing to remove experts from the Advisory Committee on Immunization Practices (ACIP), the body that recommends which vaccines people in the United States should take, and when. In an email sent to a potential candidate for the committee, physician and public health researcher Jeffrey Klausner of the University of Southern California wrote that Kennedy "is looking for highly credible individuals not conflicted with the vaccine industry."

In fact, the committee already has rigorous policies in place to address the possible impact of industry ties. And a *Science* investigation of its 13 physician members found minimal industry payments in recent years. The consulting, honoraria, speaking, travel, and "in-kind" research payments that eight members received from drugmakers averaged just over $4000 a year, nearly $3000 less than the average for all U.S. specialist physicians. For all but one member, these payments came before their terms started last year; data since they joined the committee are not yet available. Five others received no such payments.

"What RFK Jr. is doing has nothing to do with real undue influence by pharmaceutical companies," says former ACIP member Paul Offit, a vaccinologist at Children's Hospital of Philadelphia. "This is just him trying to stack the committee with people who are like-minded, so he can … make vaccines less available or less affordable."

Klausner says he knows nothing about Kennedy's reported plan to replace members, and notes that recruits might be added to the committee, which can have as many as 19 members; it currently has 15. The outgoing administration of former President Joe Biden in 2024 appointed 13 members to 4-year terms, an unusually large number for a single year. The HHS secretary can remove members at any time.

Made up of physicians, epidemiologists, and other public health experts with vaccine expertise, ACIP makes recommendations to the Centers for Disease Control and Prevention (CDC). Once accepted, they dictate which vaccines private insurers will pay for as well as those the U.S. government will provide free of charge to low-income children.

"ACIP is enormously influential in shaping our national vaccination program and it has been for decades," says Jason Schwartz, a vaccine policy specialist at Yale University who studies the committee. "That's why there is such concern about changes."

ACIP members, who annually file financial disclosures, are required to divest all vaccine manufacturer stock before they join, as are their family members; quit any service on vaccine company scientific advisory boards; and quit any consulting for vaccine companies. ACIP members may not serve as expert witnesses in vaccine litigation. They can't ac-

A 7-year-old receives the measles vaccine this month in Texas, where there's an outbreak.

PHOTO: JAN SONNENMAIR/GETTY IMAGES

cept travel or food from vaccine companies or hold patents or collect patent royalties on any vaccine product or process. Their family members can't work for vaccine companies.

There are exceptions: ACIP members can be part of vaccine clinical trials funded by industry, and the companies typically give their institutions grants to support their salaries for time spent on the trials. They can also serve on data safety monitoring boards that ensure trial participants are kept safe, and the vaccine companies pay for their time.

The rationale, Offit says, is that ACIP needs that kind of expertise to reach the most informed recommendations. But members must declare their involvement at ACIP meetings, which are public, and must remove themselves from deliberations about a vaccine and from votes on the vaccine, its potential competitors, and any product from the vaccinemaker.

The website Open Payments, where pharmaceutical companies are required by law to report payments to doctors, shows that eight ACIP members received payments from vaccinemakers from 2017–23. For seven of the members, all of whom joined the committee in 2024, these were personal remunerations for consulting, honoraria, speaking, travel, food, and lodging. The eighth joined ACIP in 2018 and in 2019 received a payment of $4662 for research-related "in-kind items and services" for work on a Sanofi influenza vaccine. The average payment to these eight members, $4106 a year, was slightly higher than the $3759 average for all U.S. doctors and less than the $6963 average for specialists. Vaccinemakers also paid several members' institutions up to $4.8 million over the 7 years, largely for clinical trial costs—a standard practice.

> **"If RFK Jr. is arguing that there is an unholy alliance … he should provide some direct evidence."**
>
> **Paul Offit**
> Children's Hospital of Philadelphia

"Where is the problem?" Offit asks. "If RFK Jr. is arguing that there is an unholy alliance between [ACIP] and these companies, he should provide some direct evidence."

Kennedy and HHS spokesperson Andrew Nixon did not respond to questions. Klausner said current measures "might be adequate, but given how contentious this area is it deserves further transparency."

As if in answer, CDC last week unveiled a tool to make it easier to see when ACIP members have declared a conflict and the ACIP votes they recused themselves from. Previously, watchdogs needed to attend a public meeting, listen to a lengthy recording, or wait months for written meeting minutes to post. The tool was released "in alignment with HHS Secretary Kennedy's commitment to radical transparency," a CDC spokesperson told *Science*.

Klausner praised the tool, but still thinks there's more that could be done. "While it may be hard to find people who are vaccine experts who do not have prior industry experience, I think it is possible and a worthwhile pursuit. There are many highly trained, objective doctors, scientists, and pharmacists who can review data and make unbiased recommendations for public health."

Klausner has a long track record in public health, previously working at the San Francisco Department of Public Health and the South African CDC. He has been an ardent proponent of vaccinating young people against the cancer-causing human papillomavirus, testifying in 2023 before the California legislature in favor of a bill that would have mandated HPV vaccination for middle school entry.

Kennedy by contrast has been a driving force behind hundreds of lawsuits by people who received Merck's HPV vaccine, Gardasil, who charge that the company didn't adequately warn them of injuries they say they sustained after receiving it; until his ascent to secretary of HHS, he was receiving payments for each client he referred to the law firm representing plaintiffs. According to his recent financial disclosure form, Kennedy received $856,559 in referral fees from the firm between 2016 and December 2024—fees he has said will now go to one of his sons.

As first reported by *The Wall Street Journal*, Klausner is helping Kennedy find new ACIP members. "I am a realist," he told *Science*. "Public health experts should not be sitting on the sidelines, outside the castle, yelling at the gates. I have an unusual opportunity to perhaps provide some advice, influence, and keep things on the rails."

Retired vaccinologist and former ACIP member Kathryn Edwards, who on 17 February received an email from Klausner that she shared with *Science*, is doubtful his approach will work. She also thinks interpreting any engagement with industry as conflict is wrongheaded. She made this clear in an emailed response to Klausner declining to have her name forwarded to Kennedy: "Vaccine studies are largely funded by pharmaceutical companies and the investigators are paid by these companies. This does not make them conflicted because the guidelines for participation and compensation are carefully monitored. This makes them experts on the vaccines that they have studied." ■

NEUROSCIENCE

Studies seek signs of consciousness before birth

Fetal and infant brains offer clues to when human experience begins

By **Kelly Servick**, *in New York City*

A recent meeting here on consciousness started from a relatively uncontroversial premise: A newly fertilized human egg isn't conscious, and a preschooler is, so consciousness must emerge somewhere in between. But the New York University (NYU)-sponsored gathering quickly veered into more unsettled territory.

At the Infant Consciousness Conference from 28 February to 1 March, researchers explored when and how consciousness might arise, and how to find out. They also considered tantalizing hints from recent brain imaging work that the capacity for consciousness could emerge before birth, toward the end of gestation.

"Fetal consciousness would have been a less central topic at a meeting like this a few years ago," says Claudia Passos-Ferreira, a bioethicist at NYU who co-organized the gathering. The conversation has implications for how best to care for premature infants, she says, and intersects with thorny issues such as abortion. "Whatever you claim about this, there are some moral implications."

How to define consciousness is itself the subject of debate. "Each of us might have a slightly different definition," neuroscientist Lorina Naci of Trinity College Dublin acknowledged at the meeting before describing how she views consciousness—as the capacity to have an experience or a subjective point of view. There's also vigorous debate about where consciousness arises in the brain and what types of neural activity define it. That makes it hard to agree on specific markers of consciousness in beings—such as babies—that can't talk about their experience.

Further complicating the picture, the nature of consciousness could be different for infants than adults, researchers noted at the meeting. And it may emerge gradually versus all at once, on different

timescales for different individuals.

Most do agree that certain features of the brain are crucial for consciousness. One is a set of connections between the thalamus—which receives and relays sensory and motor information from the body—and the cortex, where such information is further processed. Studies of fetal brains show the foundations of that link are not in place until about 24 weeks of development.

But whether those structures are sufficient for consciousness isn't clear, so investigators are finding creative ways to search for possible markers of consciousness in brain activity. One resource is the U.K.-based Developing Human Connectome Project, which has collected functional MRI (fMRI) data from hundreds of newborns, who managed to lie still enough in their swaddles by dozing through the noisy scanning process.

Naci searched those fMRI data for one purported sign of consciousness: a sort of alternating dance between a set of connected brain regions called the default mode network—active when a person's mind wanders or they think about themselves—and two others, the dorsal attention network and executive control network (ECN), which are more involved in processing the external world. When the default mode "is up and working hard … the other networks … are suppressed," Naci explained in her talk.

In work published in 2022, she and colleagues found all three networks, along with this alternating activity pattern, in newborns born after 37 weeks—and in premature infants who were born before then but scanned when they reached an equivalent age. But premature newborns scanned earlier, between roughly 32 and 37 weeks, showed no networks resembling the default mode or ECN and no reciprocal activity. That suggests these features emerge late in pregnancy.

In a study published last month, Naci and her colleagues found another, even earlier feature thought to support conscious awareness. Most premature newborns scanned before 37 weeks had signs of what's known as small-world architecture. This organizational pattern, consisting of dense neural connections between nearby brain areas and sparser links between more distant ones, has been shown to be disrupted during anesthesia and after brain injury.

Other insights have come from magnetoencephalography (MEG), which captures the magnetic fields created by neuronal activity. A team at the University of Tübingen used a MEG device designed for pregnant bellies to test how the fetus' brain reacts to disruptions in patterns of sounds. Previous experiments showed that like healthy adults, people with impaired consciousness due to brain damage can still register deviations in a short sequence of tones, such as the tone B at the end of the pattern AAAB. But their brains fail to react to longer timescale pattern deviations, such as the A at the end of the sequence AAAB AAAB AAAB AAAA. The bar for that type of response is "much higher" because it requires a memory process thought to depend on consciousness, explained Tübingen neuroscientist Joel Frohlich, who wasn't involved in the original work.

PHOTO: JILL LEHMANN/GETTY IMAGES

Clarifying when consciousness arises could help refine hospital care for premature infants.

Earlier studies had shown that even newborns can respond to these longer scale deviations. And in 2021 the Tübingen team reported that the brains of fetuses 35 weeks and older react to them as well.

Preliminary results from a survey Passos-Ferreira and colleagues conducted during the meeting suggests many attendees found the evidence for fetal consciousness compelling: Forty-seven percent of respondents said consciousness first develops in the "later prenatal," period, between about 24 weeks' gestation and birth. "Early postnatal: before 6 months old," garnered 13%, and 10% chose "early prenatal: before 24 weeks of gestation."

Topun Austin, a neonatologist at the University of Cambridge, voted for "later prenatal." But he thinks it's possible that the experience of birth—a transition from a warm soup of sedating hormones to the cold, bright world outside—could be a key step toward conscious experience.

Austin says the recent research has implications for infants born preterm. Babies born as early as 22 weeks can sometimes survive, living for months in the neonatal intensive care unit (NICU), he notes. All that time their developing brains are wiring up based on feedback from an environment that's very different from the uterus. "Giving them abnormal sensory inputs may affect their subsequent development," he says, which means it's important to limit excessive light and noise in NICUs.

Understanding when consciousness develops might also affect pain management in these tiny patients. Pain is typically defined as an unpleasant experience—which means it's likely to require "a degree of consciousness," Austin says. As recently as the 1980s, anesthesia wasn't routinely given to infants. But the modern practice of administering pain relief may have its own consequences, he says. "We don't know what giving loads of morphine to these developing opiate receptors [in the newborn brain] does" in the long term. If research could help pin down levels of consciousness in an infant at a given age, doctors might better tailor pain treatment to provide only what's necessary, he says.

Inevitably, the possibility of prebirth awareness led to discussions at the meeting about abortion. The vast majority of abortions in the United States occur before 8 weeks, noted pediatric neurologist William Graf at Connecticut Children's Medical Center—far earlier than most scientists believe pain or consciousness are likely. And consciousness is only one consideration potentially affecting the moral status of a fetus, some researchers argued. Still, claims about fetal pain have landed in recent legal debates, including the 2022 U.S. Supreme Court case that overturned the federal right to abortion. After the decision, two researchers published a paper describing how their previous pain research had been misinterpreted to claim that fetuses feel pain as early as 12 weeks after conception—before the formation of the cerebral cortex.

Any firm pronouncement that a fetus feels intense pain at X weeks is "doing a disservice to the science," Austin says, "because it's putting an absolute where there just isn't." That's one reason work like the recent fetal imaging studies is so important, he says. "I think you need to work your way down," sorting out what the accumulating clues might mean for consciousness, he says. "Then it's for society to decide what is acceptable." ■

"No swimming! Careful mines!" warns a sign posted in the Kakhovka Reservoir's dessicated bed.

ENVIRONMENTAL SCIENCE

In Ukraine, dam's destruction sets off a 'toxic time bomb'

Floods threaten to spread sediments laden with toxicants

By **Richard Stone**

In the Soviet Union, the Zaporozhets automobile was legend. Mass produced in the 1960s in the Ukrainian city of Zaporizhzhia, early models were known for steel frames as flimsy as cardboard and a gas tank under the front hood that could turn a fender bender into a carbecue. But the plebeian car did have a classy touch: nickel-plated bumpers, which had consequences that reverberate today. For decades, factories making Zaporozhets components, along with other industry and agriculture, poured effluents laden with nickel, cadmium, lead, and other heavy metals into the huge Kakhovka Reservoir nearby, where the toxicants settled into the lake-bottom sludge.

On 6 June 2023, those sediments triggered what Alexander Sukhodolov, a hydrodynamicist at the Leibniz Institute of Freshwater Ecology and Inland Fisheries, calls "a toxic time bomb." A 400-meter-wide section of the Kakhovka Dam collapsed, perhaps as a result of sabotage, sending 16.4 cubic kilometers of water—and tons of contaminated silt—surging down the Dnipro River in southern Ukraine. Scores of people died in the flood, up to 1 million lost access to drinking water, and irrigation canals for an important agricultural region turned into trickles. The deluge laid waste to ecosystems, leaving billions of mussels rotting in the desiccated lakebed and scrambling estuaries where the river empties into the Black Sea (*Science*, 5 January 2024, p. 18).

In the months since the disaster, "ecosystems have shown a remarkable resiliency," says Yuriy Kvach, a biologist at the National Academy of Sciences of Ukraine's Institute of Marine Biology. Habitats and species are reviving, he and colleagues reported in *Ecological Processes* in February. Now that the Dnipro is flowing freely, endangered sturgeon have reappeared, and meadows dotted with willows are taking hold on the dry lakebed.

But behind the vibrancy lurks a potent pollution threat, Sukhodolov and colleagues detail in a paper this week in *Science* (p. 1181). Before the dam breach, Kakhovka's reservoir accumulated a layer of loess silt up to 1.5 meters thick, amounting to as much as 1.7 cubic kilometers. The now exposed sediments contain about 83,000 tons of heavy metals, the researchers estimate. YouTube videos showing a brownish black crust on the lakebed are evidence of "the really huge accumulations of heavy metals," says ecologist Oleksandra Shumilova, lead author on the *Science* paper who's also at the Leibniz Institute.

The emerging danger is that those toxicants won't stay put, Shumilova says. Less than 1% of the sediment has been swept downstream so far, but seasonal floods from heavy rains or snowmelts will continue to wash the pollutants down the Dnipro and into the watershed around the former reservoir. In spring 2024, for instance, contaminated floodwaters inundated nearly 900 square kilometers, the *Science* authors report.

The war in Ukraine has stymied efforts to understand the scope of the threat. Much of the lower Dnipro wends along the front line, leaving scientists who venture there vulnerable to shelling or drone attacks. And the Ukrainian military forbids research cruises in the Black Sea, forcing scientists to rely mainly on remote sensing data to monitor water conditions. But some upstream areas, near Zaporizhzhia city, are safer to access. In 2023, Czech and Ukrainian nongovernmental organizations sampled the river at several sites, finding a witch's brew including heavy metals, polyaromatic hydrocarbons, and the insecticide DDT.

The "real danger" lies in the pollutants building up, or bioaccumulating, in the food web, Shumilova says. So far, however, scientists have little data on that. To fill the gap, Shumilova and Ukrainian colleagues are meeting next month to develop a research plan for tracking bioaccumulation, for instance by probing for heavy metals in the scat of local deer populations.

In the meantime, researchers and others are debating whether the Kakhovka Dam should be rebuilt. Some ecologists think that's a bad idea, noting that the willow meadows now thriving in the former lakebed are stabilizing sediments and sopping up heavy metals from the soil. Others argue rebuilding is necessary to restore irrigation networks and ensure drinking water supplies.

Sukhodolov cautions that the willows' powers of remediation won't eliminate the threat of heavy metals borne on seasonal floods. He expects Ukraine to rebuild the dam, and until the river is impounded again he and his co-authors propose building two 15-kilometer-long barriers along the Dnipro to curtail pollution releases.

As long as the war continues, such plans are pipe dreams. In the meantime, the specter of another threat haunts Ukraine: possible attacks on the Dnipro's five remaining dams or on the Dniester River cascade, a series of hydropower and flood control structures. "If more dams are targeted, the human toll and environmental damage could be cataclysmic," the *Science* authors warn.

"Water has been a weapon of war throughout history," says ecologist Carol Stepien of the Smithsonian Institution's National Museum of Natural History, who has worked in the Lower Dnipro Basin. Until the war ends, she says, "Ukraine will be in a rather precarious situation." ■

PHOTO: IVAN ANTIPENKO

MARINE SCIENCE

Oceans' trenches are home to 'incredible' diversity

In trio of studies, scientists explore life in the mysterious hadal zone

By **Phie Jacobs**

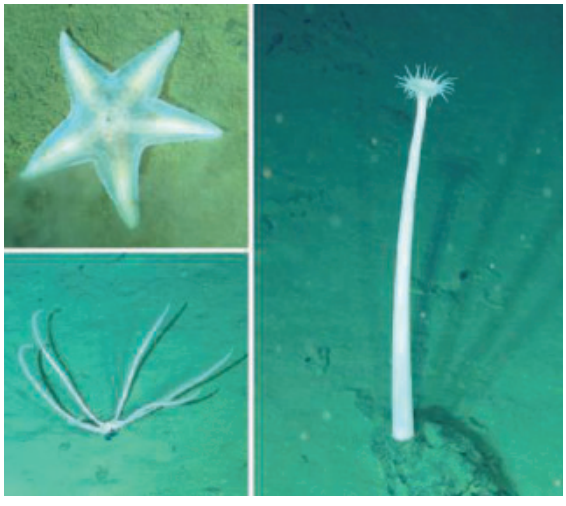
It was evening when the scientists began their journey into the abyss. As the Chinese submersible *Fendouzhe* plunged ever deeper, Weishu Zhao—an extremophile microbiologist at Shanghai Jiao Tong University—glimpsed bioluminescent creatures glowing green, yellow, and orange against the fathomless dark. Upon reaching the sea floor, more than 10,000 meters down, the team switched on the submersible lights to reveal "a profound and mysterious blue" filled with drifting plankton. "At that moment," Zhao says, "I knew immediately that the [deep ocean] must be a much more thriving habitat than we had imagined."

The Pacific Ocean's Mariana Trench, the deepest of the world's trenches, has long fascinated scientists and adventurers alike. One of the latest to explore it was movie director James Cameron, who in 2012 journeyed via cramped one-person submersible to a slot-shaped valley within the trench called the Challenger Deep, nearly 11,000 meters below sea level—the deepest point on Earth. Other expeditions have mapped the trench's bizarre topography. Now, in a trio of papers published last week in *Cell*, scientists are shining a spotlight on the weird and wonderful organisms that inhabit this mysterious environment.

"These studies significantly advance deep-sea biology," says Santiago Herrera, a molecular ecologist at Lehigh University who wasn't involved in the research. All three papers, part of the Mariana Trench Environment and Ecology Research project, shed light on one of the planet's most unexplored habitats: the hadal zone, which lies within trenches between 6000 and 11,000 meters down. There, crushing pressure, darkness, and frigid temperatures are hostile to life—and enough to thwart most expeditions.

To reach those depths, researchers used *Fendouzhe*, which Douglas Bartlett, a deep-sea microbiologist at the University of California San Diego, describes as "an engineering marvel." (The U.S. submersible *Alvin*, by comparison, can dive up to 6500 meters deep.) Capable of carrying three people to the very deepest parts of the ocean, the vessel is equipped with a pair of robotic arms and a sample basket, enabling it to collect hundreds of samples in a single dive. Between August and November 2021, *Fendouzhe* made dozens of dives in the Mariana Trench and neighboring regions, bringing back samples of microbe-filled sediment, fish, and tiny crustaceans called amphipods.

"An incredible level of diversity was uncovered," Bartlett says, referring to one new study that identified more than 7000 new microbial species in the Mariana Trench, 89% of them new to science. Study co-author Mo Han, a biologist at BGI Research, says the microbes have evolved strategies to survive despite darkness and a scarcity of nutrients. Genetic analysis revealed that some have small, highly efficient genomes specialized for a few vital functions, whereas others have larger, more versatile genomes equipping them to deal with changing environmental conditions. Some species have genes that enable them to consume difficult-to-degrade substances such as carbon monoxide. "Life finds more than one way," Han says.



At the bottom of the Mariana Trench, organisms have found ways to survive crushing pressures and scarce resources.

In a second study, researchers report that amphipods may have adapted to this extreme environment by forming a symbiotic relationship with deep-sea bacteria, making them one of the hadal zone's most abundant inhabitants, says study co-author and BGI researcher Shanshan Liu. Analysis of the crustaceans' gut contents revealed high levels of *Psychromonas* bacteria, which the team suspects may help produce a compound called trimethylamine N-oxide. Found in many deep-sea organisms, the substance maintains the balance of fluids in the body and helps protect against the damaging effects of high pressure.

A third study found evidence that fish living at depths of 3 kilometers or more all share a genetic mutation that allows their cells to more efficiently transcribe genes into essential proteins, helping them quickly respond to stress brought on by pressure, cold, and darkness. By studying the genomes of 11 species, the researchers were also able to determine when certain lineages first colonized the deep sea. Eels, for example, may have taken the plunge about 100 million years ago, avoiding the mass extinction event 65 million years ago that wiped out the dinosaurs and many marine organisms living in shallower water. Snailfish, meanwhile, might have ventured into deep-sea trenches about 20 million years ago, possibly coinciding with a period of tectonic upheaval. The deep sea may have served as an "ecological refuge" during environmental changes caused by dramatic temperature and oxygen fluctuations, says study co-author and BGI researcher Yue Song.

Besides revealing a thriving ecosystem in the hadal zone, the researchers also witnessed disturbing signs of human activity. During their dives, Zhao explains, the team came across plastic bags, beer bottles, and soda cans—and even a nearly intact laundry basket at the deepest part of another seafloor chasm, the Yap Trench. "This was deeply shocking to us," Zhao says. But they also discovered that deep-sea microbes seem to have a knack for breaking down pollutants and harnessing them for energy, she says, suggesting these organisms "might offer new solutions to current human environmental pollution problems."

In the future, the team plans to investigate other deep-sea trenches around the world—and possibly uncover clues to the origin of life on Earth. "Eighty percent of the global hadal zone is still a mystery," BGI researcher Liang Meng says, "and there could be even more extraordinary life forms to be discovered in this uncharted biosphere." ■

With reporting by Elizabeth Pennisi.

PHOTOS: (FROM TOP LEFT) SHANGHAI JIAO TONG UNIVERSITY (2); BGI GROUP

TRUMP ADMINISTRATION

NIH kills existing grants on transgender issues

Some termination letters cite "biological realities" to dismiss usefulness of such research

By **Sara Reardon**

Researchers across the United States and beyond who study transgender health are learning that their National Institutes of Health (NIH) grants have been abruptly canceled. After two killed grants were revealed last week, *Science* identified another 14 focused on transgender human health that were terminated on 28 February via NIH letters saying each project "no longer effectuates agency priorities."

NIH also last week placed seven employees on administrative leave who had worked in NIH's defunct Sexual and Gender Minority Research Office (SGMRO). They had been reassigned to other topics last year in an attempt to protect their jobs.

The moves appear part of President Donald Trump's administration's overall attempt to eliminate any funding or personnel connected to transgender issues from the federal government. In his speech to Congress last week, Trump even derided "$8 million for making mice transgender," which the White House the next day in a press release said referenced six "egregious" NIH grants, most of which involved giving rodents cross-sex hormones to study the effects on conditions such as cancer, heart disease, and asthma.

Some of the mouse research was meant to model humans receiving hormones as part of gender-affirming care. The White House didn't explicitly declare the grants had been canceled, but the president's Department of Government Efficiency (DOGE) later said it had killed a total of seven grants involving "transgender experiments on animals."

Brittany Charlton, an epidemiologist at Harvard University who studies sexual and gender inequity in medicine, says the grant cancellations will hurt a transgender community already frightened by the Trump administration's anti-LGBTQ actions. "These cuts are going to have catastrophic effects," she says. "They're not only impacting the LGBTQ community, but also scientific integrity and the broader community as well."

NIH funds more than 400 grants that mention "transgender," "gender minority," or "nonbinary" in their abstracts. The online newsletter Inside Medicine first reported that the National Institute on Aging had killed grants at Vanderbilt University and the University of Nevada, Las Vegas because of "transgender issues." But similar grants from at least 10 NIH institutes and totaling more than $8 million have also been killed, *Science* has determined by analyzing DOGE social media posts and the "savings" the organization lists on its website as examples of what it considers to be government waste. (A list of the grants is online at https://scim.ag/NIHtransgrants.)

Ivanka Savic-Berglund, a neurologist at the Karolinska Institute, received a termination letter last month. "I was mentally prepared," she says.

Antitransgender actions by President Donald Trump's administration

PHOTO: ANDREW THOMAS/NURPHOTO VIA AP

Her $120,174 grant from the National Institute of Biomedical Imaging and Bioengineering funded a project on gender dysphoria, when people identify as a different gender than they were assigned at birth. Volunteers with this gender disconnect could modify a 3D avatar to reflect how they perceive their body and the body they feel they should have. Savic-Berglund says the project had recruited nearly 40 participants and hoped to add more—but the cost savings for NIH may be minimal because her grant was set to expire at the end of March.

Sean Arayasirikul, a nonbinary medical sociologist at the University of California, Irvine, also learned that their $426,660 grant from the National Institute of Allergy and Infectious Diseases (NIAID), scheduled to run two more years, was terminated. The project studied how race- and gender-based stigma affect HIV risk and participation in HIV prevention efforts. NIAID's letter said "no modification of the project could align the project with agency priorities."

Science has also confirmed that NIH separated current grants into four categories based on Trump's executive orders, including those targeting "gender ideology" and LGBTQ+ issues. Projects that solely supported diversity, equity, or inclusion were to be terminated, whereas grants that partially support DEI activities or include some DEI language could be maintained with modifications and the rest could proceed as is. (*Nature* first reported the categorization.)

NIH did not respond to questions from *Science* about the grant terminations. A person with direct knowledge of the situation says the NIH Office of the Director, acting at the behest of the Department of Health and Human Services, instructed grants officers to send out the same termination letter except for a paragraph to be inserted depending on whether the reason was related to DEI, China, or transgender issues. At least 24 grants involving those issues were terminated, the source said, and many more terminations are coming.

Arayasirikul's letter does not mention the word "transgender." Instead, it says, "Research programs based on gender identity are often unscientific, have little identifiable return on investment, and do nothing to enhance the health of many Americans. Many such studies ignore, rather than seriously examine, biological realities."

SGMRO had coordinated internal and extramural research on transgender topics until its staff was reassigned before Trump's inauguration. "There was a strong concern that the office would become a target," an NIH employee with direct knowledge of the situation told *Science*. On 4 March, all seven of its former permanent employees received letters placing them on administrative leave. An email seen by *Science* stated the move was "not being done for any disciplinary purposes."

As for those receiving the NIH grant terminations, there's been an unexpected silver lining, Arayasirikul says. "It's forcing us to come together as a community," they say. "This is a moment that really allows us to stand together." ■

This fossil midface from Spain represents the earliest known human remains from Western Europe.

PHOTO: MARIA D. GUILLÉN/IPHES-CERCA

PALEOANTHROPOLOGY

Face to face with the first known Western European

At least 1.1 million years old, a fossil face suggests more than one type of early human inhabited Europe

By **Ann Gibbons**

On his first day digging in a Spanish cave in 2022, archaeology graduate student Edgar Téllez found something spectacular: a mud-covered facial bone with tooth roots intact. His colleague, archaeologist Rosa Huguet, took a look. "I was 95% sure we had found a human fossil, but I didn't dare say it was human remains," says Huguet, who is at the Research Centers of Catalonia's Catalan Institute for Human Paleoecology and Social Evolution at Rovira i Virgii University. She hesitated in part because it was embedded in sediments more than 1.1 million years old, older than any other human fossil known in Western Europe.

After cleaning and examining the fossil, the team of researchers was 100% certain. "The fossil represents the earliest human face of Western Europe," says paleoanthropologist Maria Martinón-Torres, director of Spain's National Human Evolution Research Centre (CENIEH) and a lead researcher of the Atapuerca Project, which has excavated many significant fossils in caves in the Atapuerca Mountains of northern Spain. The find, reported this week in *Nature*, confirms that a wide-ranging human ancestor, *Homo erectus*, whose remains are scattered across Africa and Asia, made it to Western Europe as well.

Although the ancestors of all humans arose more than 6 million years ago in Africa, the earliest hominin fossils outside of Africa are five stunning skulls that date to 1.8 million years at Dmanisi, in the Republic of Georgia. The Dmanisi fossils have been classified as early *H. erectus*, which arose in Africa 2 million years ago. But they look different enough from one another that some researchers think they represent at least two different species—suggesting more than one species of early *Homo* left Africa.

In Western Europe, the hominin fossil record was largely barren until 800,000 years ago. The only hint of an earlier human presence has been a single tooth and stone tools dating to 1.4 million years found in Spain and along the Mediterranean rim, and a 1.1-million-year-old hominin jawbone that Huguet found in the same Sima del Elefante site at Atapuerca in 2007.

The new fossil indicates that *H. erectus* or its close relative likely made at least some of those tools. The fragmentary midface shares key *H. erectus* traits, such as a primitive nose that doesn't project out as much as modern human noses. But it also shows some differences, such as a narrower midface than *H. erectus* in Asia or Africa, says paleoanthropologist José María Bermúdez de Castro of CENIEH. It only has two worn teeth, which makes it tough to pin down its identity, prompting the researchers to hedge and call it *H. "affinis" erectus*, Latin for affinity with a known species.

They do rule out that it belongs to a different human ancestor known as *H. antecessor*, fossils of which dating to about 860,000 years ago were found nearby. The midface of *H. antecessor* is much flatter and more closely resembles that of our species, *H. sapiens*, and its proteins suggest it was closely related to the common ancestor of our species, Neanderthals and Denisovans.

The date of 1.1 million to 1.4 million years old—during an epoch known as the early Pleistocene—is based on the known age for the combination of thousands of animal fossils found in the same layer of the cave. The researchers also relied on periodic shifts in Earth's magnetic field that can be detected in the sediment, and measurements of cosmogenic nuclides produced when sunlight reacted with minerals in the dirt. In the same layer, researchers also found the bone of a small mammal with cutmarks showing it had been defleshed with stone tools.

The presence of two types of early *Homo* at Atapuerca 200,000 to 300,000 years apart shows that more than one kind of hominin was already moving across Europe in the early Pleistocene, Martinón-Torres says. She thinks *H. erectus* or its close relative came first, but left Atapuerca or died out during a frigid period when glaciers advanced over Europe, about 900,000 years ago. Then, *H. antecessor* may have come from a source population in Eastern Europe or Africa, and spread to Western Europe.

Mirjana Roksandic, a paleoanthropologist at the University of Winnipeg who is not part of the team, says that scenario fits with evidence coming from archaeological sites in the Balkan Mountains. Fossils there reveal a refuge where diverse types of animals survived brutal cold periods, then spread east and west as glaciers retreated. Maybe these early hominins took refuge with those animals and followed them into the far reaches of Europe and Asia, she says.

Paleoanthropologist Katerina Harvati of the University of Tübingen says the new findings don't settle the question of where *H. antecessor* evolved. "Going forward it would be important to explore … the relationship of these populations with the early fossil humans of northern Africa, as [that] region represents an obvious, if not necessarily straightforward, potential place of origin for the Iberian groups," she says.

Regardless, Roksandic says the face is the long-sought fossil evidence that "gives us an idea of who moved, where they moved, and what was the result for human evolution in Europe." ■

IMPERFECT SOUNDS

Cochlear implants give deaf kids unprecedented access to sound. But insisting they avoid using sign language may be risky

By **Cathleen O'Grady**

Cason was 3 years old when doctors in South Carolina fitted him with cochlear implants—assistive devices for profoundly deaf people that send sound directly into the inner ear. Because the signal from the implants is not the same as biological hearing, implant users like Cason work hard to interpret what they hear through them. He soon started speech therapy, learning how to make sense of the sound coming from his implants.

It was at speech therapy that Cason's mother, Amanda Cooper, was advised not to use sign language with her son if she wanted him to learn to speak. The therapist "did not want me using it at all, in therapy or at home," Cooper says. But Cason struggled. His therapist would cover her mouth to encourage Cason to process speech without visual cues—an intensive method recommended by implant manufacturers and many specialists. "When you did that, he would break down," Cooper says, recalling the sessions. "He would cry." Although the exact reason for Cason's struggle is hard to know, his rare form of deafness, which disrupts how sound travels to the brain, may have distorted the sound from the implant; he may have also struggled because he received implants relatively late.

After a year of this, Cooper decided to stop Cason's speech therapy. Despite the therapist's warning not to sign, Cooper had started researching sign language and using very basic American Sign Language (ASL) at home to supplement Cason's communication; now, she made signing her focus. "That changed everything for him," Cooper says.

But Cason still suffered damage from being deprived of accessible language input at an early age. Now 12 years old, he reads below grade level and has a hard time expressing

This story uses both "Deaf" and "deaf" depending on what sources said was their preferred term.

himself. He can have an idea "in his head," Cooper says, but "not know how to get it out."

Cason's story underscores an ongoing, sometimes bitter debate about language access for the estimated 1000 deaf children who receive cochlear implants each year in the United States. Many providers—such as audiologists and speech therapists—as well as researchers maintain that a deaf child's best chance of learning spoken language after getting implants is to focus on speech exclusively, maximizing exposure to spoken language while a child's developing brain can best learn how to process cues from the surrounding environment.

But cochlear implants don't always deliver on their promise. Children may only get poor-quality sound through them—sometimes because the implants are not properly calibrated—or they may struggle to process the sound they hear. Kids have varying outcomes, for reasons that are not fully understood. For children whose cochlear implants don't give them good access to sound, focusing purely on speech puts their language development at risk, says Julia Hecht, a retired pediatrician and advocate for sign language. That can prompt a cascade of poor cognitive, educational, and social outcomes, she says. Kids deprived of language in their early years struggle with memory, numeracy, sequential processing, and other skills—and they may never attain full linguistic fluency.

It was in recognition of such risks that, in 2023, the American Academy of Pediatrics (AAP) published guidelines that for the first time recommended ASL or another signed language for parents to give deaf children early and unrestricted access to language—regardless of whether they have cochlear implants. Signed languages can create "immediate communication and bonding in the home," the guidelines say, and provide enough language to "prevent the risk of language deprivation."

Advocates of the speech-only approach aren't swayed, however. The American Cochlear Implant Alliance—a nonprofit advocacy group that includes clinicians, researchers, and implant manufacturers—published a response saying the recommendations were "inaccurate and biased." The group alleged that the AAP authors had excluded research that reports good speech outcomes in kids with no sign language exposure and glossed over the difficulty faced by hearing parents in learning a new language.

Both sides agree that early intervention is key. When children do not reach their cognitive, linguistic, and emotional potential because of language deprivation, "the stakes are the highest," Hecht says.

UNLIKE HEARING AIDS, which sit outside the ear and amplify sound, cochlear implants send sound directly into the inner ear via electrodes implanted into the cochlea. Roughly one in every 1000 babies born in the U.S. has severe to profound hearing loss, making them potential candidates for a cochlear implant. Various factors, including the anatomy of the inner ear, determine whether an implant will be viable, but fewer than half end up with the devices.

Kipp in conversation with his deaf mentor, Nobuo Hara (left), who taught Kipp and his family Australian Sign Language.

Guidelines for audiologists, pediatricians, and other medical experts emphasize the role of "parent choice" in the decision: A cochlear implant is recommended if parents want to prioritize spoken language for their child. Most deaf children are born to hearing parents, who don't know a signed language and who probably don't know a deaf person, says Ann Geers, a retired linguist who consults for Washington University in St. Louis and whose work has been influential in promoting the use of technology and speech therapy with deaf kids. Their goal for their child is "to live as typical a life as possible," she says.

Many doctors and other professionals still believe spoken language is "critical and necessary to function in society," says Wyatte Hall, a Deaf language deprivation researcher at the University of Rochester. But he and others in the deaf community feel this reflects ableism, a form of bias that views spoken language as intrinsically better than signing.

Children's progress with speech does not always live up to parents' hopes. A child with cochlear implants is "still deaf," Hall says, and may not have good enough access to sound to acquire fluent spoken language.

Because the signal from implants is imperfect, children need to learn to make sense of the input. That's why clinicians recommend cochlear implants as early as possible, while kids' brains are at their most flexible. In many countries, newborns are screened for deafness, and in the U.S. a child can receive cochlear implants as early as 9 months of age. A range of studies supports early intervention: The younger children are when implanted, the more likely they are to approach the spoken language skills of hearing peers.

Families should choose which languages to use with their child, says Jessica Messersmith, an audiologist at the University of South Dakota. But if their goal for a child with implants is spoken language, then "spoken language does need to be a primary input." Messersmith co-authored the 2019 American Academy of Audiology clinical practice guidelines, which say parents should be informed about the full range of language options, but that "the likelihood of a child gaining high benefit in the areas of speech perception, speech production, and spoken language increases when more emphasis is placed on listening and spoken language."

Many of the studies comparing how children fare when exposed only to speech or to speech and sign language are weak or flawed, a 2016 *Pediatrics* review concluded. But Messersmith and others point to an influential paper Geers and colleagues published the next year in *Pediatrics*. The study, done in 97 children with cochlear implants, found that those who had not been exposed to any kind of visual language—including ASL, as well as systems that mix signs with spoken language—had better speech outcomes than children with some visual language exposure.

It was not a controlled study, Geers emphasizes: To properly test the effects of different language choices, researchers would have had to randomly assign parents to use different language modalities with their children, which would be impossible in practice.

Hall and others had deeper reservations. They said that because the study grouped ASL with artificial communication systems that combine gestures and speech, it could say little about the effects of exposing children to a natural human signed language. They also argued the research couldn't show

CREDITS: (PHOTO) SARAH BROPHY; (ILLUSTRATION, OPPOSITE PAGE) MIRIAM MARTINCIC

that sign language actually caused the poorer outcomes. Parents could be more likely to use sign if their child struggles with speech, says Deborah Chen Pichler, a sign language linguist at Gallaudet University—so the causal arrow could point the other way: "Those children are probably signing because they weren't progressing well in spoken language."

That's not uncommon. A 2016 paper in *Otology and Neurotology* found that many early-implanted children had outcomes on a range of spoken language tests that were close to hearing children—but depending on the test, between 19% and 42% of these kids did not. Research has found that children are more likely to approach or match their hearing peers on speech if they are younger at implantation, have more intensive speech training, and have more educated parents. But it's not perfectly predictable which children will struggle, or why. "For all the children who are doing well with cochlear implants, there is this sort of trail of the children who are left behind," Hecht says.

And sometimes the technology itself falters. Implants need to be calibrated to ensure an implant user can hear a full range of frequencies at a comfortable volume, a step that sometimes goes awry. In one extreme case, news broke in 2022 that Adelaide, Australia's Women's and Children's Hospital had botched the "mapping" of many children's cochlear implants. So far, investigations have identified 141 children who were affected, as well as a further 59 cases found in a separate cluster at Townsville University Hospital in Douglas, Australia. To advocates of early signing, such cases are a cautionary tale.

WAITING TOO LONG to consider sign language can be risky, advocates say, even as parents are regularly advised to "wait and see" whether a child takes to speech therapy.

This was the case with Kipp, a boy who lives in Esperance, a remote town in Western Australia. His parents, Sarah Brophy and Damian Freeman, a nurse and paramedic, respectively, visited some 20 medical, hearing, and speech professionals about Kipp's hearing and language when his early language was not progressing and hearing test results were unclear. None encouraged them to use sign language, and about half warned against it, saying it could be harmful, Brophy says. An audiologist finally identified Kipp as profoundly deaf when he was nearly 3 years old, still steering his parents away from sign language, even though more than a year of speech therapy had been ineffective. By then, the family was drowning. Kipp had no way to communicate and was acting out in frustration. "The day was just filled with horrendous screaming and screeching. … I was at breaking point," Brophy says.

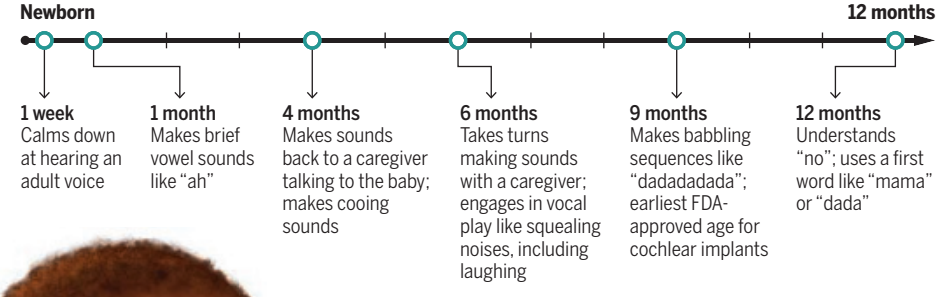
Shortly after Kipp's deafness was identified, Nobuo Hara, a deaf neighbor, learned about Kipp and visited the family. "He gave us a deaf awareness training in our kitchen and said, 'Whether you can afford to pay me or not, I want to teach your family Auslan [Australian Sign Language],'" Brophy says. Hara became the family's mentor, coming to their house three times a week to practice for an hour or two.

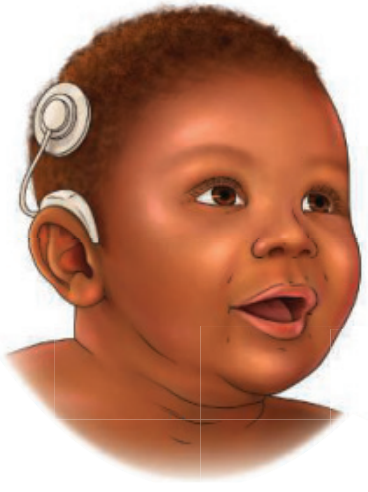
Within weeks, everything changed. Kipp picked up vocabulary rapidly, and within months, he was "storytelling and asking questions," Brophy says. "His whole world opened up and our family could now connect." At age 5, he has Auslan skills on par for his age. Kipp got his first cochlear implant at age 4 and the second at age 5, and Brophy and Freeman have leaned on Auslan to help him with spoken language: When he's confused by English, he asks questions in Auslan. "He is such a gentle soul—so empathetic and

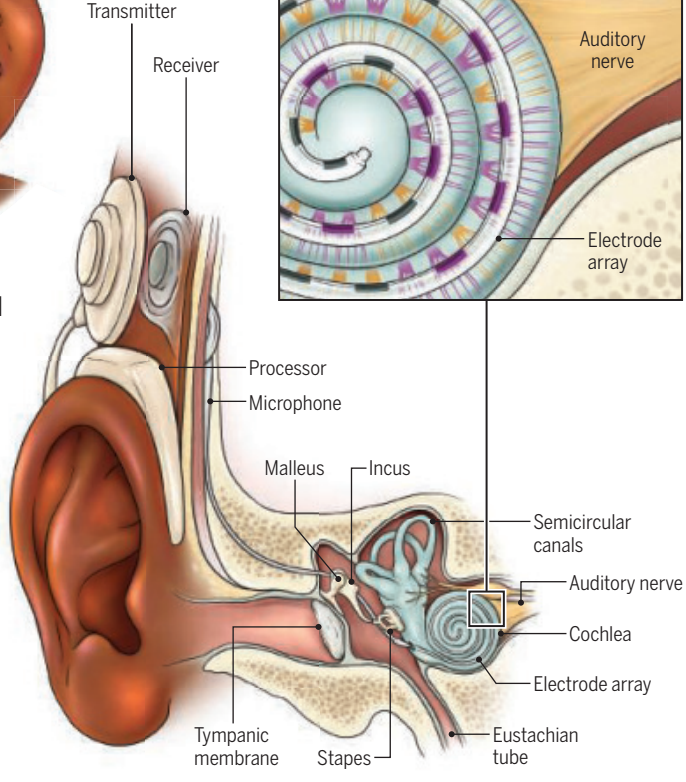
Early milestones

Linguistic skills begin to develop early in a newborn's life. Research finds that children with cochlear implants have greater success with spoken language if they are implanted at about 9 months, which is the earliest age for which the U.S. Food and Drug Administration (FDA) has approved these devices. But babies younger than this are already taking in language from their surroundings and building up to their first word.

Language before words







From sound to signal

A cochlear implant uses a sound processor that sits behind the ear. This turns audio waves from outside the ear into a digital signal sent to the transmitter coil that conveys that signal to the surgically implanted receiver. The receiver then sends electrical impulses into an electrode array inserted into the cochlea. This stimulates the auditory nerve to further transmit the information to the brain.

CREDITS: (GRAPHIC) A. MASTIN/*SCIENCE*; (DATA) C. BOWER *ET AL.*; *PEDIATRICS* **152:3** (2023)

patient and inquisitive," Freeman says, which makes it more painful to remember the profoundly frustrated and isolated toddler he once was: "I actually can't spend any time thinking about it. … It breaks you."

Kipp seems to have been young enough to catch up thus far despite having late access to language. But missing out on language exposure when the brain is at its most flexible can have lasting consequences, as it did for Cason. "People think language begins with the first word," says Carol Padden, a Deaf sign language linguist at the University of California San Diego (UCSD). "It doesn't." Hearing children begin cooing, turning their heads towards a voice, and making sounds back to their caregivers at just 4 months old. By 6 months, they engage in turn taking, making noises back and forth with other people in a conversationlike pattern. Children exposed to a signed language from birth hit the same milestones using their hands as articulators. "It's almost like you have to build the apparatus, and then you insert the words," Padden says.

The longer a child goes without language, the more severe the effects are, says Rachel Mayberry, a linguist at UCSD who studies the effects of language deprivation. Mayberry's work has found that language-deprived late sign language learners processed sign language using areas of the brain associated with vision, rather than language; and that they had anatomical differences in areas of the brain associated with language processing. Other research has found that language deprivation is linked to poverty, unemployment, and poor health, and that many children with late exposure never become truly fluent in any language. Late language learners may be able to hold a conversation but struggle to express themselves and to understand what other people are saying, Padden says: "They really don't have the kind of comprehension that we take for granted."

THE ANSWER, many researchers and advocates say, is to expose kids with implants to both speech and sign language. A growing body of research finds that raising deaf children "bimodal bilingual"—with both sign and speech—does no harm to their speech, and may benefit it, alongside the crucial benefit of guarding against language deprivation.

One worry is that hearing parents who learn sign language won't be fluent enough to help their kids' language outcomes. But deaf children whose parents are novice signers don't seem to be hampered: Naomi Caselli, who studies deaf education at Boston University, and her colleagues have found that deaf children whose hearing parents are learning ASL have vocabularies comparable to deaf children raised by fluent signers and that early exposure to ASL at school can help close any gap left by parents.

Yet parents often do not learn of this option. A 2023 paper in the *Journal of Speech, Language, and Hearing Research* surveyed 105 families with a deaf child who had cochlear implants or other assistive technology. The parents reported that 43% of pediatricians, 44% of otolaryntologists, 47% of audiologists, and 30% of speech-language pathologists had advised them to only use spoken language with their children. If doctors and other professionals advising parents give them no information about language deprivation, ASL, and bilingualism, "how we can we expect them to make a well-informed choice?" asks Sara Nović, a deaf author and advocate for access to sign language. Parents who have never met a deaf person may also have a skewed idea of what it means to be deaf.

Parents who do decide to sign with their kids often struggle to find the resources they need. Audiologists, speech and language therapists, and other practitioners are often housed under one roof, but practices seldom include ASL specialists or offer any support for ASL in-house, says Mallorie Evans, an audiologist who works in schools in the Los Angeles area. It's a systemic bias against ASL, she says, and sends a message to parents: "What you're telling those families is that you don't value that [service] enough to offer it."

For Karen Strache, who decided to raise her daughter Lorelai with English and ASL, the path has not been smooth. But she points to a recent crisis to explain why she insists on bilingualism: In December 2024, a chest infection spread through Lorelai's respiratory system and infected her cochlear implant site. In the hospital, facing blood draws, CT scans, and surgery, Lorelai—then 4 years old—couldn't wear her implants. Without ASL, she wouldn't have been able to understand what was happening to her. Having ASL, Strache says, "was huge."

For Strache and some others, the deaf community has made all the difference in being able to stick with signing. Strache lives in Eagle Creek, Alaska, where the deaf community is small, but locals make an effort to spend time signing with Lorelai at events, she says. Brophy and Freeman had a similar experience; even in remote Esperance, they have found ways to involve other deaf adults in Kipp's life. Tanya Lehmann, the mother of a deaf child in Adelaide, says her family has been met with kindness and understanding from the local deaf community, which runs counter to a widespread misperception that deaf adults are hostile to cochlear implants.

Advocates for sign language aren't arguing for it over spoken language, Hall says—they're pointing to the growing body of evidence that suggests kids can have both. But, "The key is to make sure this first language foundation is fully accessible and set in place," he says. "Because that's so critical to being a fully realized human being."

Cooper and her family can attest to that. They have moved to Florida for Cason to attend an ASL school, where he is flourishing. Cason loves swimming and math—and, like so many kids his age, is a big fan of the video game *Fortnite*. He reads every night for half an hour to try to improve at it. Cooper often has to rephrase questions for him to understand, and he still struggles to express his thoughts and feelings—but his expressive language continues to improve, Cooper says. In a video on social media, she asks him whether he likes being deaf. "Yes!" he signs emphatically. Why? "Signing." ■

Lorelai's parents, Christopher and Karen Strache (top), raised her with English and American Sign Language, which she used to understand what was happening during a hospital admission (bottom) that prevented her from wearing her cochlear implants.

PHOTOS: (TOP TO BOTTOM) ELANI ANN APERANS; KAREN STRACHE

REVIEW ROUNDUP

Science at Sundance 2025

Former student journalists revisit their efforts to expose an environmental cover-up in the 1990s. A late astronaut's life partner offers insights into a transformative period in the American space program. An attorney argues for prioritizing data over opinions in a critical health care fight. From a sensitive interrogation of disability and euthanasia, to an unflinching look at gun violence in America, to a tale of a library renovation that prompts reflection on the nature of knowledge creation, a number of films featured at this year's Sundance film festival probe themes of interest to scientists and engineers. Read on for reviews of six of the 2025 festival's offerings. **—Valerie Thompson**

Heightened Scrutiny

Reviewed by **Valerie Thompson**[1]

Sam Feder's urgent documentary, *Heightened Scrutiny*, follows American Civil Liberties Union attorney Chase Strangio as he prepares to argue before the US Supreme Court in December 2024 on behalf of three Tennessee transgender youths, their parents, and a Memphis-based physician, who maintain that the state's ban on gender-affirming hormone therapy and puberty blockers violates the equal protection clause of the 14th Amendment. The court's decision, expected in June 2025, will affect the future of transgender health care in America.

On phone calls with colleagues, alone in nondescript hotel rooms, huddled under an umbrella on a crowded European beach, and in countless media appearances, Strangio—who will be the first out transgender person to argue before the Supreme Court—spends

ILLUSTRATION: CHIARA VERCESI

months rehearsing his case, demonstrating how the same medical interventions that are prescribed for other conditions without state involvement only provoke alarm when prescribed to transgender youths and likening ban supporters' supposed concerns for children's welfare to the arguments once used to uphold the prohibition on interracial marriage. (The similarities to interracial marriage bans would find resonance during the proceedings with Justice Ketanji Brown Jackson, who observed that "…some of these questions about, sort of, who decides and the concerns and legislative prerogatives, et cetera, sound very familiar to me. They sound [like] the same kinds of arguments that were made back in the day, the '50s, '60s, with respect to racial classifications.")

Throughout the film and together with other attorneys, scholars, journalists, and activists, Strangio offers convincing evidence that public discourse rather than peer-reviewed research is the driving factor informing much of the country's legislation surrounding gender-affirming care. Political talking points unsupported by facts are invoked in public hearings, the film's participants argue, while mainstream media coverage emphasizing potential harm and regret over potential benefits and affirmation is cited in legislation. One piece, an opinion column published in *The New York Times* (*1*), was quoted repeatedly by the State of Idaho in defense of a law banning gender-affirming care for transgender youth just 6 days after publication, the film reveals. That opinions have supplanted evidence in public policy should be disturbing to all, if not surprising, but it is an issue of particular concern to scientists, many of whom have relevant expertise and data to add to this discourse and who may soon find it more difficult than ever to do so.

REFERENCES AND NOTES

1. P. Paul, "As kids, they thought they were trans. They no longer do," *New York Times*, 2 February 2024.

Heightened Scrutiny, *Sam Feder, director,* 2025, 89 minutes.

Life After

Reviewed by **Gabrielle Kardon[2]**

In 1983, disabled by cerebral palsy, suffering from chronic pain, and confined to a hospital in Riverside, California, 26-year-old Elizabeth Bouvia sought to end her life by refusing to eat. However, in a court case that attracted nationwide attention, Bouvia was denied this right. After participating in a number of high-profile interviews with the media, the last given in 1998, Bouvia disappeared from public view. Did she continue to fight, wondered filmmaker Reid Davenport nearly four decades after the court ruling. Is she still alive? If so, is she happy? In *Life After*, Davenport, who, like Bouvia, has cerebral palsy, seeks to learn more about Bouvia and to better understand society's attitudes toward individuals with disabilities and access to assisted suicide.

Disabled individuals often require extensive support, and many rely on resources

from underfunded, backlogged bureaucracies. When faced with overwhelming care costs, Michal Kaliszan—a computer scientist with spinal muscular atrophy—sought relief through a Canadian law that legalized physician-assisted suicide for individuals with serious illnesses and disabilities in 2021. "I didn't want to die, but I didn't want to live an undignified life," he tells Davenport, who invites viewers to consider whether better assistance might prevent disabled people from ending up in this impossible position.

Bouvia—who died in 2014, Davenport learns—is more inscrutable. Through interviews with her sisters, Teresa and Rebecca Castner, viewers learn that she was subjected to many painful surgeries throughout her childhood. "Her body was like a battlefield," Rebecca tells Davenport. Several times, Bouvia was institutionalized, and she wrote in her diaries that "she was a scared and forgotten soul trying hopelessly to maintain her dignity."

A central question is raised: What constitutes a good quality of life, and who makes this determination? "The only way you can justify the system in place right now is if you think that people who are disabled are not worth the same as someone with an able body," argues journalist Ash Kelly. Davenport likewise pushes back on narratives that frame euthanasia as a humane and dignified option for those with disabilities.

Winner of a Sundance US Documentary Special Jury Award, Davenport's film is not, he insists, about suicide, "it's about the phenomenon that leaves people desperate to find their place in the world that perpetually rejects them." *Life After* gives viewers a poignant and nuanced glimpse into this world.

Life After, *Reid Davenport, director*, 2025, 99 minutes.

How to Build a Library

Reviewed by **Valerie Thompson**[1]

The question of how knowledge is produced and by whom is one that is increasingly being confronted by the scientific community. This same question pervades Maia Lekow and Christopher King's *How to Build a Library*, which follows Shiro Koinange and Angela Wachuka as they lead a team attempting to reinvent a dilapidated Kenyan library system into a vibrant resource for modern-day citizens.

The women's task is a daunting one. The McMillan Memorial Library is an imposing structure in downtown Nairobi whose neoclassical facade includes a grand marble staircase and a pair of stately lion statues. But the building's interior fails to live up to its frontage. Scenes from inside reveal broken floorboards, cracked and water-stained walls, and vaulted rooms full of discarded furniture and other detritus. Most troubling to the pair is the library's collection, which is filled with outdated, colonialist perspectives—a legacy of the library's origins as a European-only institution established during British rule.

Confronting administrative hurdles and placating sensitive bureaucrats whose approvals are needed to ensure the project's success, the pair work tirelessly to fundraise, collect community feedback, and amass African-created works that will one day populate the updated library and two of its suburban branches. But opposition emerges from an unlikely source: the librarians who have long stewarded the institution's resources. Resentful of the new group's efforts to overhaul the library without deference to their expertise as library professionals, the librarians resist the team's move to abandon the Dewey decimal system. The debate that emerges about how racism and imperialism can be embedded in seemingly innocuous classification systems echoes similar discussions ongoing in the sciences.

Long-forgotten news stories and photographs unearthed from dusty archives during the renovation force the group to confront Kenya's colonial past, as does the revelation that the British Council is among the library renovation's biggest financial backers. A 2023 visit to the library from King Charles and Queen Camilla is juxtaposed with archival footage of a mid-20th-century visit from Queen Elizabeth II, and Koinange and Wachuka muse darkly about having to perform for the royal family like their predecessors did.

Cracks emerge in the women's friendship as the formidable work grinds on, but the pair's commitment to building a community resource that honors and celebrates African knowledge never waivers, and *How to Build a Library* succeeds in conveying the stakes and hard work that underlie their vision.

How to Build a Library, *Maia Lekow and Christopher King, directors*, Circle and Square Productions, 2025, 103 minutes.

Shiro Koinange promotes the Nairobi Book Festival in 2024.

PHOTO: STEVEN MAY/ALAMY STOCK PHOTO

SALLY

Reviewed by **Michael D. Shapiro**[3]

In the mid-1980s, astronaut Sally Ride was one of the most famous people in the United States, if not the world. She embraced her status as a role model and readily took professional risks to break barriers for women and girls. Other risks were too great, however, reveals *SALLY*, winner of the 2025 Alfred P. Sloan Feature Film Prize. Ride, who died in 2012, kept her 27-year same-sex relationship with scientist and education scholar Tam O'Shaughnessy a secret out of fear that going public would negate her accomplishments and derail her professional dreams.

As members of the astronaut class of 1978, Ride and five other women underwent the same grueling training as their male col-

Astronaut Sally Ride poses for a photo while training as a mission specialist for NASA's STS-7 spaceflight.

PHOTO: NASA

leagues. Still, the women were treated as novelties by NASA, facing ignorance and stereotypes fomented by the politics of exclusion. Some of Ride's male colleagues thought she had not earned her spot in the space program, but what they meant was that she did not earn it in a familiar way—her PhD in astrophysics and athletic achievements were seen as inferior to their military experience. Meanwhile, NASA engineers, nearly all of whom were men, cluelessly designed a toiletry kit for Ride containing 100 tampons for a 7-day mission and wondered whether that quantity was sufficient. The press corps was just as oblivious: One reporter asked Ride if she cried when something went wrong during training.

As challenging as it was to be a woman in the space program, Ride felt that being out as queer would be impossible, recalls O'Shaughnessy. When friends and family outside NASA came out or were outed, Ride saw their careers and reputations destroyed. Similarly, and tragically, she feared that disclosing her secret could transform her from a hero to a pariah.

Director Cristina Costantini uses archival video footage to document a transformative era in the American space program, but her present-day interviews with family and former colleagues in Ride's inner orbit are the highlights of the film. O'Shaughnessy's memories, in particular, bring unique context and humanity to Ride's story and reveal the pain and exhaustion she felt keeping the pair's romantic partnership hidden for so long.

SALLY is especially timely at a moment when efforts to embrace inclusion are under renewed assault by tired phobias repackaged as new grievances. Ride and other NASA pioneers whose abilities were questioned because of their differences capably contributed to many high-stakes space missions, and *SALLY* makes a forceful case that there is every reason to expect the same on Earth.

SALLY, *Cristina Costantini, director*, National Geographic, 2024, 103 minutes.

The Perfect Neighbor

Reviewed by **Paul L. Koch**[4]

Although violent crime in the United States has dropped substantially since 1990 (with a brief pandemic surge that is fast receding), the nation has an unusually high level of gun violence. As reported by CNN in 2024, the rate of firearm homicides in the US is 18 times that of other developed nations, which is accompanied by a much higher level of gun ownership (*1*). In *The Perfect Neighbor*, director Geeta Gandbhir grapples with one aspect of this crisis by uncovering the slow-moving catastrophe that led to the murder of Gandbhir's family friend Ajike Owens by a neighbor, Susan Lorincz. Along the way, the film obliquely examines the role that Florida's Stand Your Ground law may have played in the tragedy.

After moving to a tight-knit, multiracial, multigenerational community, Lorincz became increasingly angry and erratic when boisterous neighborhood children played in a yard adjacent to her property. Using almost entirely body cam footage, *The Perfect Neighbor* follows frustrated police as they respond to Lorincz's incessant complaints. The film focuses on her downward spiral and on neighbors who describe her increasingly irrational, aggressive, and often racist behavior to law enforcement, building a sense of foreboding. When Owens knocks on Lorincz's locked front door on 2 June 2023 to discuss yet another dustup between Lorincz and the kids, she is met with gunfire that kills her. Harrowing body cam footage captures the frantic police response and the anguish of Owens's neighbors and young children when they learn she has died.

Stand Your Ground laws allow a person to use deadly force if they believe their life is in danger. But interrogation room video footage raises troubling questions about Lorincz's fear of Owens and whether her actions might have been premeditated, with the intent to use Florida law as a shield. Eventually, Lorincz is charged with manslaughter and convicted of first-degree felony manslaughter. The film's closing credits play over scenes from her sentencing, where she receives 25 years in prison.

A growing number of studies, including two briefly referenced in *The Perfect Neighbor* (*2*, *3*), have revealed that Stand Your Ground laws increase gun violence and death. While it never explores this question directly—say,

In *Middletown*, former student documentarians reflect on a project that uncovered an environmental scandal.

PHOTO: SANDRA PHIPPS

through interviews with experts—this finely wrought film allows the participants and bystanders in one specific instance of gun violence to speak for themselves, and viewers must reach their own conclusions. What is not in question is the chaos and destruction that guns and our weak societal response to them produce. In *The Perfect Neighbor*, all those involved experience loss and sorrow.

REFERENCES AND NOTES

1. K. Fox *et al.*, "How US gun culture stacks up with the world," CNN, 15 February 2024; https://www.cnn.com/2021/11/26/world/us-gun-culture-world-comparison-intl-cmd/index.html.
2. M. D. Esposti *et al.*, *JAMA Netw. Open* **5**, e220077 (2022).
3. J. K. Roman, "Race, justifiable homicide, and Stand Your Ground laws: Analysis of FBI supplementary homicide report data" (The Urban Institute, 2013).

The Perfect Neighbor, *Geeta Gandbhir, director*, 2025, 96 minutes.

Middletown

Reviewed by **Robert S. Krauss**[5]

In 1991, Fred Isseks, a teacher at a high school in Middletown, New York, was tipped off that the nearby Wallkill landfill was a site of toxic waste dumping and set the students enrolled in his Electronic English class loose as novice journalists and filmmakers to uncover the story. More than 30 years later, veteran documentarians Jesse Moss and Amanda McBaine (*Boys State*, 2020) regather Isseks and four students to tell the story of this project in the new film *Middletown*. Moss and McBaine are aided immensely by access to the students' original videotapes, which contribute to a considerable portion of *Middletown*'s running time.

The Middletown students were a diverse group, and not all took school seriously. However, they each threw themselves into visiting the landfill site—at times trespassing and risking possible toxic exposure—and interviewing local officials, whose initial reactions ranged from patronizing to overtly hostile. Former landfill employees and a state wildlife pathologist proved more cooperative, and the students' perseverance slowly bore fruit. Over several years, a sordid story emerged, involving mob influence, corrupt officials, and a cover-up. In 1997, the class completed and released a film, *Garbage, Gangsters, and Greed*, helping to spur hearings organized by legendary US Congressman Maurice Hinchey, then a state assemblyman. Isseks's students learned how to navigate local government by attending public meetings and filing FOIA (Freedom of Information Act) requests, but they also learned hard lessons in what it means to challenge entrenched power. The dumping eventually stopped, but the site was capped rather than cleaned; it still exists and sits atop an aquifer.

Middletown begins as an *Erin Brockovich*–style tale but transitions to an older type of story, one about the lifelong benefits of an inspiring teacher. Two students discovered their vocation through the project. Rachel Raimist went to film school at the University of California, Los Angeles, and earned a PhD at the University of Minnesota, where a media center is named in her honor. She is currently a television director. Jeff Dutemple is an Emmy-nominated cinematographer. Most importantly, all the kids who took Electronic English embraced civic engagement and found, in Isseks's words, "civic courage." As we enter a political period where environmental concerns will again take a back seat, this lesson could not be more timely. I recommend *Middletown*, and if you have middle school– or high school–age kids, try to watch it with them. ■

Middletown, *Jesse Moss and Amanda McBaine, directors*, 2025, 111 minutes.

10.1126/science.adw4633

[1]The reviewer is the books and culture editor at *Science*, AAAS, Washington, DC, USA. Email: vthompso@aaas.org [2]The reviewer is at the Department of Human Genetics, University of Utah, Salt Lake City, UT, USA. Email: gkardon@genetics.utah.edu [3]The reviewer is at the School of Biological Sciences, University of Utah, Salt Lake City, UT, USA. Email: mike.shapiro@utah.edu [4]The reviewer is at the Department of Earth and Planetary Sciences, University of California, Santa Cruz, Santa Cruz, CA, USA. Email: plkoch@ucsc.edu [5]The reviewer is at the Department of Cell, Developmental, and Regenerative Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. Email: robert.krauss@mssm.edu

CANCER

Making sense of fat in cancer

A lipid chaperone enables sensing of an essential fatty acid to drive tumor growth

By **Estela Jacinto**

Dietary nutrients supply essential raw materials for cell growth and proliferation. Although a balanced diet promotes health, excess nutrients, such as carbohydrates and fats, can disrupt cellular processes, increasing the risk of diseases such as cancer. Determining how specific nutrients affect the cell growth machinery at the molecular level is vital to understanding the impacts of nutrition on human health. A key component of this machinery is the enzyme mechanistic target of rapamycin (mTOR). mTOR partners with other proteins to form two distinct complexes (mTORC1 and mTORC2) that regulate various aspects of anabolic metabolism (*1*). Yet, how mTORCs sense nutrients, particularly lipids or fats, remains largely unexplored. On page 1164 of this issue, Koundouros *et al.* (*2*) report the mechanism by which mTORC1 senses omega-6 linoleic acid (ω-6 LA)—a dietary long-chain polyunsaturated fatty acid (PUFA)—providing clues on how specific cancers depend on this essential fat for tumor growth.

Mammalian cells use PUFAs to form cell membranes, and they can be metabolized into other lipid derivatives that serve as fuel or signaling molecules. However, unlike most fats, both ω-6 LA and another PUFA, omega-3 alpha-linolenic acid (ω-3 ALA), are essential fatty acids, which means that humans cannot synthesize them. They must instead be acquired through dietary sources, such as fish, nuts, and vegetable oils. Notably, ω-6 LA is more abundant in the Western-style diet, whereas ω-3 ALA is enriched in the Mediterranean diet. Although some studies have highlighted the health benefits of PUFAs, the appropriate dietary amounts of these fats are controversial (*3*). This debate stems from a limited understanding of the intracellular targets of these lipids. Koundouros *et al.* help to address this knowledge gap by showing that ω-6 LA, but not ω-3 ALA or its derivatives, is a direct activator of mTORC1, thus linking ω-6 LA to the central growth machinery in cells. This strengthens the notion that excess fat, even the essential ω-6 LA, can have adverse effects by potentially overstimulating mTORC1 and consequently deregulating cell growth.

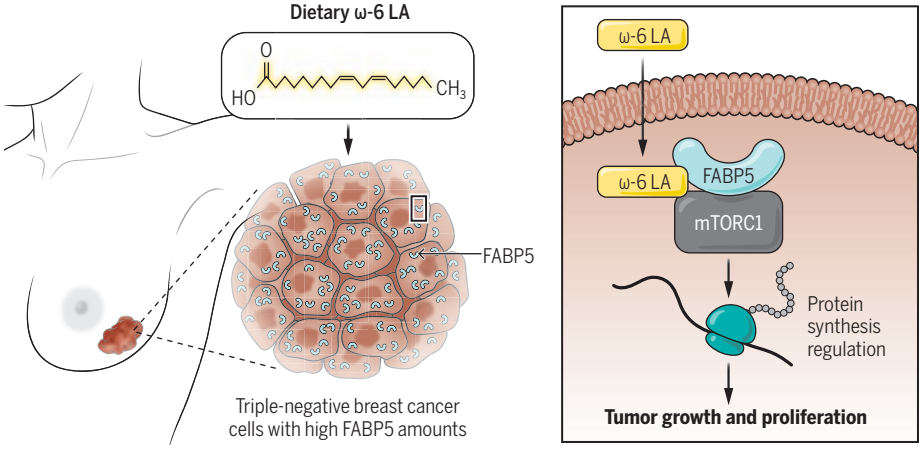
Koundouros *et al.* discovered that the effects of ω-6 LA on mTORC1 depend on amounts of the lipid chaperone fatty acid–binding protein 5 (FABP5). Like other FABPs, FABP5 binds to different types of fats and allows them—especially very hydrophobic long-chain fatty acids—to traverse the aqueous environment of the bloodstream and cells. FABP5 has garnered attention because of its involvement in several diseases and in lipid metabolism (*4*). Koundouros *et al.* observed that FABP5 acts as an adaptor to couple ω-6 LA availability to mTORC1 activation (see the figure). In the presence of ω-6 LA, FABP5 binds to the mTORC1 component called regulatory-associated protein of mTOR (Raptor). This interaction enhances mTORC1 complex assembly and substrate engagement. Activation of mTORC1 by ω-6 LA–FABP5 increases phosphorylation of specific substrates, including ribosomal protein S6 kinase (S6K) and eukaryotic translation initiation factor 4E–binding protein 1 (4E-BP1), which regulate protein synthesis (*5*).

The exact mechanism by which ω-6 LA–FABP5 activates mTORC1 is unclear, though it likely involves the Ras-associated guanosine triphosphatases (GTPases) (Rag) and Ras homolog enriched in brain (Rheb) GTPases, which also mediate mTORC1 activation in response to amino acid concentrations (*6*). The findings of Koundouros *et al.* suggest that mTORC1 integrates inputs from various nutrients to coordinate growth signals. Notably, ω-6 LA activates mTORC1 even in the absence of amino acids and growth factors, which implies that lipids can provide divergent regulatory signals for mTORC1 activation. Other lipids, such as phosphatidic acid, oleic acid, and palmitate, also activate mTORC1 (*7*), but they do not require FABP5 (*2*). Studies are needed to elucidate how the full range of different lipids are sensed by mTORC1. Determining how lipid-induced activation of mTORC1 and mTORC2 influences the dynamic trafficking of lipids used in membrane assembly and other anabolic processes will be a rich area for future research.

How might ω-6 LA specifically promote malignancy? Although FABP5 is ubiquitously expressed compared with other FABPs, its amounts are increased in several cancer types, including breast cancer (*8*). High FABP5 expression is associated with poor relapse-free survival in breast cancer patients. Koundouros *et al.* found that the amount of FABP5 is higher in so-called triple-negative breast cancer [which lacks the estrogen and progesterone receptors as well as human epidermal growth factor receptor 2 (HER2)] compared with receptor-positive breast cancer cells.

Sensing of dietary fat linked to cell proliferation in cancer

The polyunsaturated fatty acid omega-6 linoleic acid (ω-6 LA) induces the activation of mechanistic target of rapamycin complex 1 (mTORC1), a key regulator of anabolic growth and cell proliferation. The lipid chaperone fatty acid–binding protein 5 (FABP5), which has high affinity for ω-6 LA, directly binds mTORC1 to increase its activation, thus promoting cell growth and proliferation. FABP5 expression increases in triple-negative breast cancer cells, making these tumors highly sensitive to dietary ω-6 LA.



GRAPHIC: K. HOLOSKI/*SCIENCE*

Department of Biochemistry and Molecular Biology, Robert Wood Johnson Medical School, Rutgers University, Piscataway, NJ, USA. Email: jacintes@rwjms.rutgers.edu

Using cellular models, the authors demonstrated that the expression level and fatty acid–binding ability of FABP5 are important for ω-6 LA to activate mTORC1. Notably, mice implanted with triple-negative breast cancer cells had increased tumor growth when fed a diet with ω-6 LA but not one with ω-3 ALA, corroborating the notion that tumors with abundant FABP5 may have more appetite for dietary ω-6 LA. This raises the question as to how FABP5 expression changes during tumor initiation and progression. Increased FABP5 expression in cancer often correlates with increased expression of other regulators of lipid metabolism and cell proliferation (*9*), indicating coordinated regulation of these genes. Whether nontumor sources, such as adipose, contribute to the amount of FABP5 in the circulation—delivering lipids to tumors and other tissues—warrants further examination (*10*). Because ω-6 LA is only obtained through the diet, restricting its intake could be beneficial for patients with high amounts of FABP5. Investigating the role of excess dietary fat in reprogramming both cellular and whole-body metabolism that consequently deregulate gene expression should yield valuable insights into the mechanisms underlying cancer initiation and progression.

Targeting FABPs in cancer is currently under investigation (*11*). There are also clinical efforts targeting mTOR in cancer, although these have shown modest results because of toxicity (*12*). Combining mTOR and FABP5 inhibition, or blocking their interaction, together with dietary ω-6 LA restriction could be a therapeutic strategy for cancers dependent on the ω-6 LA–FABP5–mTORC1 axis for cell proliferation. A balanced intake of ω-6 LA and ω-3 ALA is already gaining traction in promoting healthy immunity (*3*). Future studies should explore how manipulating the amounts of these lipids rewires growth signals in different cell types to prevent cancer. ■

REFERENCES AND NOTES

1. A. Szwed *et al.*, *Physiol. Rev.* **101**, 1371 (2021).
2. N. Koundouros *et al.*, *Science* **387**, eadm9805 (2025).
3. I. Djuricic, P. C. Calder, *Nutrients* **13**, 2421 (2021).
4. B. Xu *et al.*, *Front. Cell Dev. Biol.* **10**, 857919 (2022).
5. S. Battaglioni *et al.*, *Cell* **185**, 1814 (2022).
6. T. D. Lama-Sherpa *et al.*, *Biochem. Soc. Trans.* **51**, 655 (2023).
7. D. Menon *et al.*, *J. Biol. Chem.* **292**, 6303 (2017).
8. W. G. Warren *et al.*, *Drug Discov. Today* **28**, 103628 (2023).
9. C. A. Powell *et al.*, *Oncotarget* **6**, 6373 (2015).
10. S. Guaita-Esteruelas *et al.*, *Oncologist* **22**, 1309 (2017).
11. W. G. Warren *et al.*, *Toxicol. Appl. Pharmacol.* **491**, 117079 (2024).
12. C. Magaway *et al.*, *Cells* **8**, 1584 (2019).

ACKNOWLEDGMENTS

E.J. acknowledges support from the National Institutes of Health, National Institute of General Medical Sciences (R01GM137493); the New Jersey Commission on Cancer Research (COCR22PRG009); and the Mizutani Foundation for Glycoscience.

10.1126/science.adw1956

PHOTO: THOMAS MARENT/MINDEN PICTURES

ECOLOGY

Reading palms to characterize islands

Rattan palms illuminate the drivers of biodiversity in tropical Asia

By **Anna J. Holmquist**[1] *and* **Rosemary G. Gillespie**[2]

The Asian tropics are among the world's most diverse regions, encompassing four distinct biodiversity hotspots. Although previous studies have highlighted the complex biogeographic patterns in the region, they have been constrained by low spatial resolution, scarcity of genetic data, and conceptual frameworks that fail to account for within-region speciation. Accordingly, understanding of the biotic assembly (i.e., the contribution of dispersal versus speciation to species diversity) of these regions remains limited. On page 1204 of this issue, Kuhnhäuser *et al.* (*1*) describe reconstructing the evolutionary history of one of the most species-rich Asian palm lineages, the rattans and relatives (Arecaceae, Calamoideae). They develop a biogeographic framework with which to view biodiversity dynamics in the Asian tropics, with the potential to inform conservation priorities.

Composed of fragments from the Sunda (Asian) and Sahul (Australian) continental shelves that merged over 150 million years ago, with substantial changes over the past 25 million years, the Asian tropics have an elaborate geological history (*2*). Extensive tectonic and volcanic activity resulted in over 20,000 islands in the Indo-Australasian Archipelago with wide-

Plant diversity in Asian tropical areas, such as the Danum Valley Conservation Area in Borneo, Malaysia, may be driven by distinct roles of individual regions.

ranging topographical complexity. Although some islands were variably connected to continental plates, others have been isolated since their origination, resulting in intricate patterns of diversification and a high concentration of endemic taxa.

Kuhnhäuser *et al.* combined molecular data from 360 rattan species with fossil data to investigate rattan palm diversity in the Asian tropics, using biogeographic stochastic mapping to infer speciation and dispersal histories. They present a compelling biogeographic framework that expands beyond the classic source-sink model, classifying areas into four groups: "radiators," in which diversity is both generated and distributed; "incubators," in which diversity is generated but not distributed; "corridors," characterized by high emigration; and "accumulators," in which diversity is acquired from outside of the region.

Previous efforts to relate island biogeographic theory and geological history have relied on the equilibrium theory of island biogeography (*3*), which describes how species diversity on islands can be predicted on the basis of a balance between immigration and extinction. Immigration decreases with increasing distance from a mainland source and extinction decreases with increasing island size. The analysis by Kuhnhäuser *et al.* confirmed a relationship between distance and island area but went beyond classical equilibrium theory by highlighting the importance of speciation in understanding assemblages in biodiversity hotspots. Although the equilibrium theory highlights the role of isolation in reducing immigration, a key focus is on the additional role of isolation in fostering biodiversity by providing the opportunity for speciation (*4*). Whether by distance, environment, or colonization, isolation can promote speciation by limiting the homogenizing effects of gene flow (*5*).

Likewise, although the equilibrium theory considers area in the context of its impact on population size and hence extinction rates, area also has a key role in speciation. Larger areas are generally correlated with landscape features that promote speciation, including habitat diversity, heterogeneity, and potential allopatric barriers (*6*). Diversification can be facilitated by larger areas that can accommodate a greater range of habitats, and by multiple scales of isolation within the area. Time also interacts with isolation to dictate the evolutionary outcome of immigration and extinction dynamics (*4*). The length of time over which islands have been available to receive colonizers affects whether an island can or cannot serve as a source. Furthermore, priority effects—in which the first species that colonizes an area has an advantage over subsequent colonizers—can dictate the ability of colonizers to settle, directing species formation in a predictable manner (*7*).

The role of different regions as radiators, incubators, corridors, or accumulators, as demonstrated by Kuhnhäuser *et al.*, can be contextualized by considering the paleogeography of the region. Borneo is noted as a sole radiator, important in both forming and distributing diversity. It is one of the oldest areas in the region and was connected by a land bridge to peninsular Asia until the sea level rise that occurred after the last glacial maximum. The long-standing existence of the island, in addition to its size, has contributed to the observed high rates of in situ speciation, whereas its connectivity to the Thai-Malay peninsula explains its role as a major source of biodiversity. New Guinea and Sulawesi were identified as incubators with high regional speciation and low emigration. New Guinea was submerged during the late Miocene (about 11 to 5 million years ago) to early Pliocene (about 5 to 4 million years ago) (*8*) and only reached its current form in the past few million years (*9*). Sulawesi was formed through a collision of multiple continental fragments 20 million years ago and existed as an isolated lowland archipelago until 1 to 2 million years ago (*10*). Both islands are thus young, meaning that neither had area available to colonize, and isolation prevented them from serving as major sources of biodiversity, whereas their size, topography, and isolation promoted extensive speciation.

The Philippine islands and Palawan were identified as accumulators; much of their land area emerged in the late Miocene (*11*) and they were never connected to surrounding areas. By contrast, areas defined as corridors (Thai-Malay Peninsula, Sumatra, Java, and Maluku) have been variably connected to surrounding areas, promoting emigration between connecting regions. Australia and India were also found to be accumulators, again potentially owing to modern rattans reaching them in the Miocene, (23 to 5 million years ago) whereas Indochina was identified as the origination of the clade, providing extensive time for speciation, and served primarily as an incubator.

The biogeographic framework presented by Kuhnhäuser *et al.* highlights the importance of considering speciation when studying biotic assemblage. Areas defined as "incubators" and "radiators" drive the region's endemism and could therefore be considered conservation priorities in rattans. By applying sophisticated modeling to a region-wide dataset, Kuhnhäuser *et al.* provide a powerful tool to study potential drivers of the extraordinarily rich biodiversity in tropical hotspots worldwide. ■

[1]Center for Comparative Genomics, San Francisco, CA, USA. [2]Department of Environmental Science, Policy and Management, University of California Berkeley, Berkeley, CA, USA. Email: gillespie@berkeley.edu

REFERENCES AND NOTES

1. B. G. Kuhnhäuser *et al.*, *Science* **387**, 1204 (2025).
2. R. Hall, *J. Asian Earth Sci.* **20**, 353 (2002).
3. R. H. MacArthur, E. O. Wilson, *Evolution* **17**, 373 (1963).
4. B. C. Emerson, R. G. Gillespie, *Trends Ecol. Evol.* **23**, 619 (2008).
5. L. Orsini *et al.*, *Mol. Ecol.* **22**, 5983 (2013).
6. J. B. Losos, C. E. Parent, "The speciation-area relationship" in *The Theory of Island Biogeography Revisited*, J. B. Losos, R. E. Ricklefs, Eds. (Princeton Univ. Press, 2010), pp. 415–438.
7. K. L. Shaw, R. G. Gillespie, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 7986 (2016).
8. D. P. Gold *et al.*, *Mar. Pet. Geol.* **86**, 1133 (2017).
9. K. C. Hill, R. Hall, Mesozoic-Cainozoic evolution of Australia's New Guinea margin in a West Pacific context. Defining Australia: The Australian Plate as Part of Planet Earth, Geological Society of America and Geological Society of Australia, joint publication, special paper, 1–43 (2002).
10. A. M. S. Nugraha, R. Hall, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **490**, 191 (2018).
11. M. de Bruyn *et al.*, *Syst. Biol.* **63**, 879 (2014).

10.1126/science.adw1635

GEOPHYSICS

How does Antarctic ice deform?

A deep-learning model infers large-scale dynamics of Antarctic ice shelves

By **Bryan Riel**

Mass loss from ice sheets that cover Earth presents the largest potential contribution to future sea level rise, yet uncertainties in predicting the magnitude and pace of this depletion hamper climate adaptation planning and coastal infrastructure decisions. Fast-flowing glaciers—moving masses of ice, snow, rock, and sediment—transition to floating ice shelves that extend from land to the ocean. These ice shelves exert buttressing forces on inland upstream ice, affecting ice sheet stability (*1*). A key factor governing the ice flow dynamics is the deformation of ice under stress. Elucidating rheological parameters at the scale of an entire ice shelf, which extends over hundreds of kilometers, has been a long-standing challenge in glaciology. On page 1219 of this issue, Wang *et al.* (*2*) report a physics-informed deep-learning model that can predict the deformation behavior of Antarctic ice shelves, revealing complexities of the process that extend beyond the traditional understanding.

School of Earth Sciences, Zhejiang University, Hangzhou, China. Email: briel@zju.edu.cn

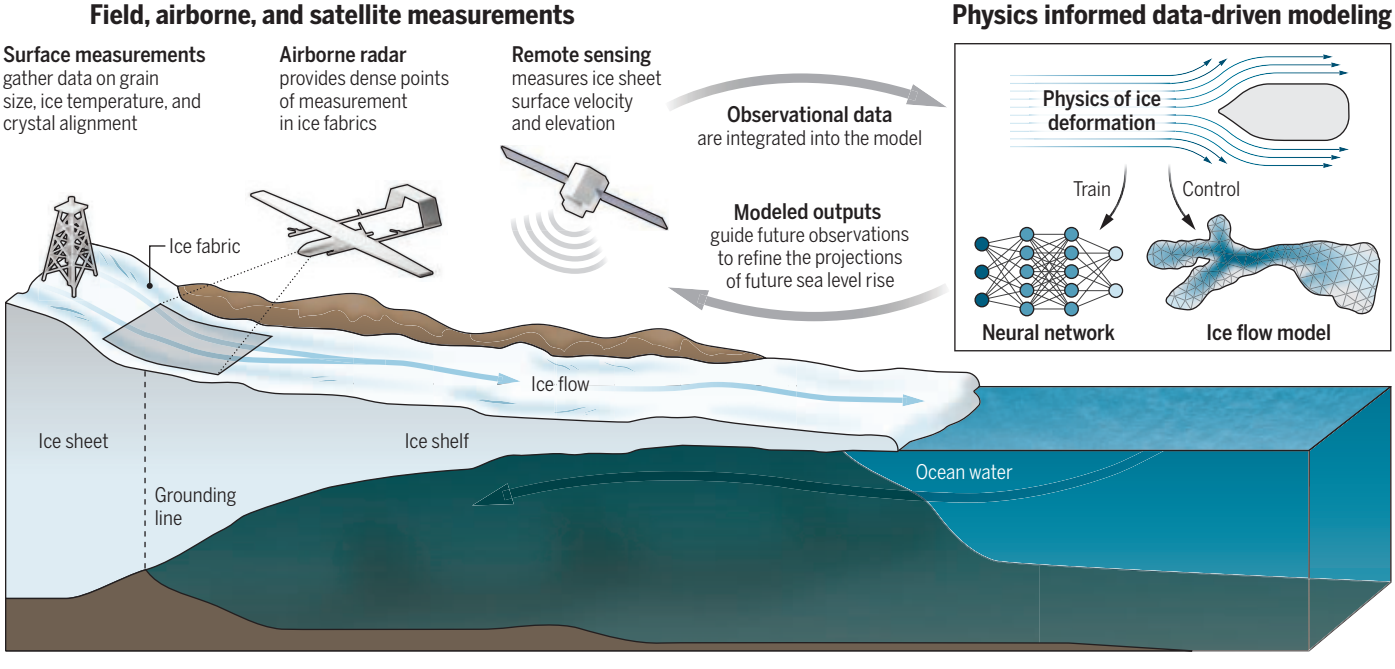
The relationship between stress and deformation of ice is crucial to understanding the dynamics of ice flow and predicting how much ice will contribute to rising sea levels. Ice shelves have been long recognized as ideal locations for understanding the rheology of ice because of their relatively simple flow physics and a large availability of high-quality velocity and thickness data. Traditionally, ice deformation has been described by Glen's flow law, which relates stress to deformation rate in a simple power-law formula (*3*). Although the coefficient and the stress exponent that make up the equation have been empirically obtained by laboratory-scale experiments (*4*), scaling these parameters for an entire ice shelf presents a grand challenge. Properties of an ice shelf tend to vary spatially, and direct measurement of stress exerted on ice is a difficult task. Moreover, the complex interplay between factors such as ice temperature, the average diameter of an ice crystal (grain size), and the alignment of ice crystals that constitute an ice shelf complicate the calculation (*5*).

The advent of satellite-based remote sensing has allowed continuous spatial and temporal measurements of ice velocity at a continental scale. Several studies have leveraged ice velocity data to estimate the deformation rate and stress of ice across large regions by assuming a specific flow direction and type of deformation (*6*, *7*). Although this technological advance has contributed to scaling the rheological parameters to an entire ice shelf, uncertainties and disagreements in calculations persist (*8*). Many traditional ice flow models assume a stress exponent of three in the Glen's flow law. However, laboratory experiments, field measurements, and observational analyses suggest that this exponent ranges between one and five, a value that depends on the sensitivity of the specific ice shelf to stress (*8*–*10*).

Wang *et al.* used a data-driven approach to infer the dynamic flow behaviors of the ice shelves in Antarctica. The physics-informed neural networks—a type of machine-learning model—are integrated with physical theories that govern a system, such as the ice flow equation, and observational data. Through layers of interconnected nodes that handle information, the model can reveal factors that govern ice flow be-

Understanding how ice sheets deform and flow at a large scale

Discerning ice flow dynamics at the scale of an entire ice shelf is critical for projecting ice sheet stability and future sea level rise. Integrating observational data, physical laws, and data-driven approaches can give more accurate predictions of ice flow and how this affects overall glacier movement.



GRAPHIC: N. BURGESS/*SCIENCE*

havior. Wang *et al*. reconstructed ice viscosity (resistance to deformation) across entire ice shelves and obtained more accurate estimates of the stress imposed on them. Comparisons of these predictions to those obtained from the traditional power-law relationship showed some intriguing evidence of spatial variation in the relationship between stress and deformation rate. In several densely packed locations (compression zones) of the ice shelves near the grounding line—the boundary between inland and floating ice sheets—the stress exponent was lower than the traditionally assumed value of three. This value has strong implications for estimating the stability of the grounding line (*8*). By contrast, the deformation rate of ice shelf areas that stretch toward the ocean (extension zones) obtained by Wang *et al*. deviate substantially from a simple power law. This suggests that spatial variations in ice temperature, grain size, crystal orientation, and defects within an ice shelf could affect the stability of ice.

A key factor that influences the spatial variation in ice flow is mechanical anisotropy within an ice shelf. The ensemble of ice crystal orientations, called the fabric, tends to have a crystal alignment that depends on the ice's flow history. Similar to grains in wood, the fabric possesses strong mechanical anisotropies and could promote shearing of ice in a specific direction. Despite its considerable effect on regions that experience shear deformation, the fabric is not well represented in flow models because field measurements of fabric are sparse at the scale of ice shelves. Wang *et al*. extended the neural network framework to analyze anisotropy in the viscosity of ice. The calculated variation in anisotropy aligned well with the suture zones, which are areas where ice masses with different properties merge together. This specific region is associated with marine ice development and affects how cracks spread in an ice sheet (*11*). These results are a valuable complement to in-field observations and airborne radar data used to simulate how ice fabric evolves at a large scale (see the figure). Further advancement in airborne technology could potentially provide dense points of measurements in ice fabrics (*12*). Combining computational models of fabric evolution and ice flow could also discern how ice fabrics develop and affect overall glacier movement (*13*).

Beyond predicting how ice shelves flow under stress, physics-informed neural networks are providing new ways to analyze data in other areas of geophysics such as earthquake science and atmospheric research. This deep-learning framework also offers several advantages compared with the traditional mathematical models. For example, it does not require prebuilt data grids that are usually necessary for numerical calculations. This benefit is particularly useful when only sparse measurements are available. At the same time, physics-informed neural networks also scale efficiently with large datasets. Nonetheless, these computational approaches require long training times and can experience reduced sensitivity to certain dynamic parameters. Wang *et al*. implemented multiple physics-informed neural network frameworks in parallel to speed up the computation. Independent models processed small subsets of ice shelf data, which were then seamlessly combined to calculate the overall ice flow. However, the sensitivity challenge needs further improvement. Combining these new technologies with traditional ice flow models, while also quantifying uncertainties in inputs and measurements, could improve the accuracy of model predictions (*14*).

As the ability to extract rheological parameters of ice from remote-sensing data improves, understanding of the cryosphere and its response to the rapidly changing climate could deepen. Wang *et al*. highlight the need of additional constraints to improve ice flow models that are obtained from direct observations, such as fabric properties and grain size. This would be particularly important for predicting the flow in vulnerable regions of ice sheets such as West Antarctica to refine the projections of future sea level rise. ■

REFERENCES AND NOTES

1. G. Gudmundsson, *Cryosphere* **7**, 647 (2013).
2. Y. Wang, C.-Y. Lai, D. J. Prior, C. Cowen-Breen, *Science* **387**, 1219 (2025).
3. J. W. Glen, *Proc. R. Soc. London Ser. A* **228**, 519 (1955).
4. D. Goldsby, D. L. Kohlstedt, *J. Geophys. Res.* **106**, 11017 (2001).
5. M. D. Behn, D. L. Goldsby, G. Hirth, *Cryosphere* **15**, 4589 (2021).
6. P. D. Bons *et al*., *Geophys. Res. Lett.* **45**, 6542 (2018).
7. J. D. Millstein, B. M. Minchew, S. S. Pegler, *Commun. Earth Environ.* **3**, 57 (2022).
8. M. Ranganathan, B. Minchew, *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2309788121 (2024).
9. C. M. Schohn, N. R. Iverson, L. K. Zoet, J. R. Fowler, N. Morgan-Witts, *Science* **387**, 182 (2025).
10. D. R. MacAyeal, *J. Geophys. Res.* **94**, 4071 (1989).
11. C. Borstad, D. McGrath, A. Pope, *Geophys. Res. Lett.* **44**, 4186 (2017).
12. T. Young *et al*., *J. Geophys. Res. Earth Surface* **126**, e2020JF006023 (2021).
13. D. A. Lilien *et al*., *J. Glaciol.* **69**, 2007 (2023).
14. B. Riel, B. Minchew, *J. Glaciol.* **69**, 1167 (2023).

ACKNOWLEDGMENTS

The author acknowledges support from the National Natural Science Foundation of China (42376230) and the Zhejiang University Global Partnership Fund.

10.1126/science.adw3158

DEVELOPMENTAL BIOLOGY

Tunneling through cardiac jelly

Membrane projections from muscle cells enable signaling in the developing mouse heart

By **José Luis de la Pompa**[1,2]

The heart is the first organ to become functional during embryonic development in vertebrates. The process begins with the formation of the heart tube, which consists of an inner endothelial lining (endocardium) and a layer of muscle cells (myocardium), separated by a thick extracellular matrix called the cardiac jelly. Subsequent development of the heart ventricles involves the formation of trabeculae—muscular ridges lined by the endocardium. How signals pass between the endocardium and myocardium across the cardiac jelly to ensure normal development is not fully understood. On page 1166 of this issue, Miao *et al*. (*1*) report that, in mice, the myocardium and endocardium of the early heart tube communicate directly through tiny membrane-enclosed projections called tunneling nanotube–like structures (TNTLs). Loss of TNTLs disrupts trabecular formation and thus ventricular development and embryonic viability.

In mice, development of the ventricles of the heart follows a clear timeline. At around 9.5 days of gestation, referred to as embryonic day 9.5 (E9.5), trabeculae develop in the heart tube. From E10.5 to E12.5, myocardial compaction refines the ventricles, and they continue to mature through E13.5 to E15.5. The development of trabeculae and maturation of ventricular chambers depend on Notch signaling. Notch is a highly conserved signaling pathway that involves physical contact between a Notch receptor on one cell and a Notch ligand [for example, jagged-1 (JAG1) or delta-like protein 4 (DLL4)] on another cell. Signaling between the endocardium and myocardium is crucial for trabeculation (*2*–*4*). The presence of DLL4 in the endocar-

[1]Intercellular Signaling in Cardiovascular Development and Disease Laboratory, Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid, Spain. [2]Ciber de Enfermedades Cardiovasculares, Instituto de Salud Carlos III, Madrid, Spain. Email: jlpompa@cnic.es

dium drives trabeculation, whereas JAG1 in the myocardium regulates compaction (*5*). One important question is how signals are exchanged when these tissues are separated in the heart tube by cardiac jelly.

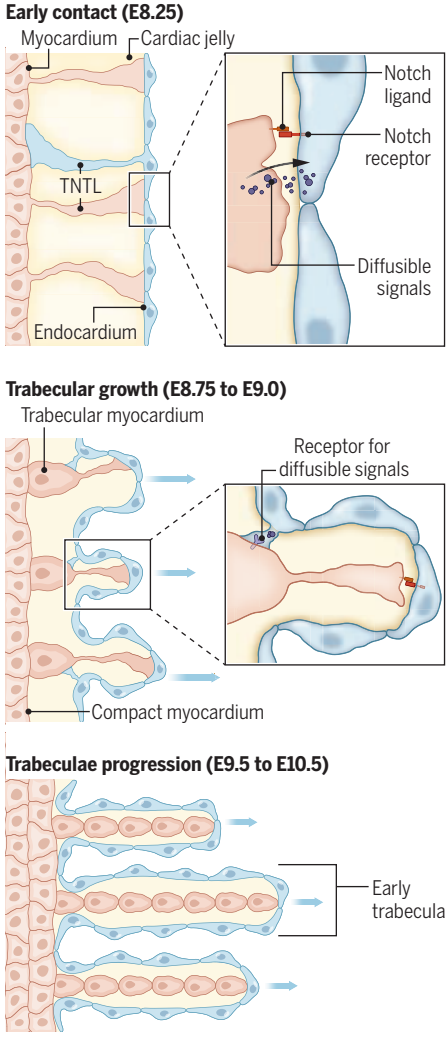
Miao *et al.* genetically engineered mice in which embryonic cardiomyocytes express a Notch ligand tagged with green fluorescent protein (GFP) and endothelial cells express a modified Notch receptor bearing an extracellular GFP-binding domain and an intracellular tetracycline transactivator (tTA). When the GFP-labeled ligand from cardiomyocytes binds to the receptor on endocardial cells, tTA is activated and moves into the endocardial cell nucleus. There, it switches on the expression of a β-galactosidase (LacZ) reporter gene, allowing the identification of endocardial cells in which Notch ligand-receptor interaction has occurred using X-gal staining, a simple method used to visualize gene activation. The authors observed this staining in endocardial cells of E8.5 embryos, in which the endocardium and myocardium are separated by the cardiac jelly.

Using high-resolution three-dimensional images Miao *et al.* observed nanotube-like protrusions extending from cardiomyocytes and making contact with endocardial cells at E8.75. The authors also identified noncontacting nanotubes that extended from cardiomyocytes or endocardial cells without reaching the opposite layer. To study the role of these nanotubes in trabeculation, the authors cultured cells from the ventricles of a different line of Notch reporter mice at E9.5. Endocardial cells only showed Notch reporter activity when contacted by TNTLs from cardiomyocytes, indicating that direct cell-cell interactions activate Notch in endocardial cells.

Further in vitro analysis of the TNTLs revealed that they are made of actin filaments. Administration of an actin inhibitor to cardiomyocytes and endocardial cells from E9.5 embryos reduced TNTL formation, which suggests that actin polymerization is crucial for TNTL development. TNTL formation also relied on the small guanosine triphosphatase (GTPase) cell division control protein 42 homolog (CDC42). Deleting the *Cdc42* gene in cardiomyocytes prevented TNTL formation. Three-dimensional imaging of hearts from wild-type embryos at E8.25 to E8.75 showed cardiomyocytes at the top of the forming trabeculae, with TNTLs connecting them to endocardial cells. At the same developmental stage, hearts from embryos with myocardial inactivation of *Cdc42* lacked trabeculae and TNTLs. Notably, myocardial *Cdc42* inactivation at later stages (E9.5) did not affect trabeculation, which suggests that the formation of TNTLs is critical during early trabeculation (E8.25 to E8.75), before endocardial cell-cardiomyocyte interactions at E9.5.

Long-distance contacts

At ~8.25 days of gestation [embryonic day 8.25 (E8.25)], cardiomyocytes in mice form tunneling nanotube–like structures (TNTLs) that reach through the cardiac jelly to contact endocardial cells. These TNTLs enable Notch ligands on the cardiomyocyte to interact with Notch receptors on the endocardial cell, initiating trabeculation. TNTLs might also transport diffusible signals and are required for the progression of trabeculation through E9.0. By E9.5, the two cell types make direct contact, particularly at the base of trabeculae. TNTLs reaching from endocardial cells to cardiomyocytes may also exist during trabeculation.



The TNTLs also facilitated the exchange of signaling molecules between cells. For example, fluorescently tagged versions of actin and JAG1 moved along TNTLs from cardiomyocytes to enter endocardial cells. Furthermore, the activity of several signaling pathways involved in heart development, including those that require the diffusible signals Neuregulin-1 and Angiopoietin, was reduced in the *Cdc42* mutant mice. This indicates that TNTLs improve the efficiency of multiple signaling interactions between the endocardium and myocardium (see the figure).

The molecular mechanisms and cellular processes that contribute to trabecular growth, patterning, and maturation are under intense investigation (*6*–*10*). The study by Miao *et al.* demonstrates that TNTLs in cardiomyocytes facilitate Notch signaling across the cardiac jelly, enabling the activation of Notch receptors on endocardial cells and initiation of trabeculation. Notably, Notch activation in endocardial cells triggers non–cell-autonomous responses in cardiomyocytes, promoting their proliferation and differentiation (*3*, *5*). Early in trabeculation, Notch activity is concentrated in endocardial cells at the base of trabeculae and gradually extends to the tip as trabeculation progresses (*3*). The contribution of TNTLs to this process remains to be studied, as does their potential role in delivering diffusible signals (for example, Neuregulin-1 and Angiopoietin) and the production of TNTLs by endocardial cells.

Whether TNTLs are formed at specific locations in response to signaling cues is also unknown. TNTLs may form in response to cell-cell adhesion cues such as integrins or cadherins, which could be tracked using fluorescent reporters. Mechanical inputs or hypoxia could also promote TNTL formation and could be assessed using techniques that mimic these stresses in cell cultures. The complexity and persistence of embryonic trabeculae are closely linked to adult heart function, including contractility and disease susceptibility. Atypical trabeculation is associated with cardiomyopathies such as left ventricular noncompaction and dilated cardiomyopathy (*11*–*14*). Therefore, understanding the fundamental mechanism of trabeculation and ventricular wall maturation is critical for both biological understanding and clinical advancement. ■

REFERENCES AND NOTES

1. L. Miao *et al.*, *Science* **387**, eadd3417 (2025).
2. K. F. Lee *et al.*, *Nature* **378**, 394 (1995).
3. J. Grego-Bessa *et al.*, *Dev. Cell* **12**, 415 (2007).
4. G. Del Monte-Nieto *et al.*, *Nature* **557**, 439 (2018).
5. G. D'Amato *et al.*, *Nat. Cell Biol.* **18**, 7 (2016).
6. J. Grego-Bessa *et al.*, *Circ. Res.* **133**, 927 (2023).
7. R. Priya *et al.*, *Nature* **588**, 130 (2020).
8. D. Passer, A. van de Vrugt, A. Atmanli, I. J. Domian, *Cell Rep.* **14**, 1662 (2016).
9. J. Li *et al.*, *Cell Rep.* **15**, 158 (2016).
10. V. Jiménez-Amilburu *et al.*, *Cell Rep.* **17**, 2687 (2016).
11. G. Luxán *et al.*, *Nat. Med.* **19**, 193 (2013).
12. M. Siguero-Álvarez *et al.* *Circulation* **147**, 47 (2023).
13. H. V. Meyer *et al.*, *Nature* **584**, 589 (2020).
14. K. A. McGurk *et al.*, *Nat. Cardiovasc. Res.* **3**, 1503 (2024).

ACKNOWLEDGMENTS

The author acknowledges support from Centro de investigación Biomédica en Red–Enfermedades Cardiovasculares (CIBERCV) grants PID2022-104776RB-100 and CB16/11/00399 (CIBER CV) from Spanish Ministry of Science grant MCIN/AEI/10.13039/501100011033 and by the La Caixa Research Health Foundation (ref. HR23-00084).

10.1126/science.adw1567

GRAPHIC: A. FISHER/*SCIENCE*

POLICY FORUM

ARTIFICIAL INTELLIGENCE

Large AI models are cultural and social technologies

Implications draw on the history of transformative information systems from the past

By **Henry Farrell[1], Alison Gopnik[2,3,4], Cosma Shalizi[4,5,6], James Evans[4,7]**

Debates about artificial intelligence (AI) tend to revolve around whether large models are intelligent, autonomous agents. Some AI researchers and commentators speculate that we are on the cusp of creating agents with artificial general intelligence (AGI), a prospect anticipated with both elation and anxiety. There have also been extensive conversations about cultural and social consequences of large models, orbiting around two foci: immediate effects of these systems as they are currently used, and hypothetical futures when these systems turn into AGI agents—perhaps even superintelligent AGI agents. But this discourse about large models as intelligent agents is fundamentally misconceived. Combining ideas from social and behavioral sciences with computer science can help us to understand AI systems more accurately. Large models should not be viewed primarily as intelligent agents but as a new kind of cultural and social technology, allowing humans to take advantage of information other humans have accumulated.

The new technology of large models combines important features of earlier technologies. Like pictures, writing, print, video, internet search, and other such technologies, large models allow people to access information that other people have created. Large models—currently language, vision, and multimodal—depend on the internet having made the products of these earlier technologies readily available in machine-readable form. But like economic markets, state bureaucracies, and other social technologies, these systems not only make information widely available, they allow it to be reorganized, transformed, and restructured in distinctive ways. Adopting Simon's terminology, large models are a new variant of the "artificial systems of human society" that process information to enable large-scale coordination [(*1*), p. 33].

Our central point here is not just that these technological innovations, like all other innovations, will have cultural and social consequences. Rather we argue that large models are themselves best understood as a particular type of cultural and social technology. They are analogous to such past technologies as writing, print, markets, bureaucracies, and representative democracies. Then we can ask the separate question about what the effects of these systems will be. New technologies that are not themselves cultural or social, such as steam and electricity, can have cultural effects. Genuinely new cultural technologies—Wikipedia, for example—may have limited effects. However, many past cultural and social technologies also had profound, transformative effects on societies, for good and ill, and this is likely to be true for large models.

These effects are markedly different from the consequences of other important general technologies such as steam or electricity. They are also different from what we might expect from hypothetical AGI. Reflecting on past cultural and social technologies and their impact will help us to understand the perils and promise of AI models better than worrying about superintelligent agents.

SOCIAL AND CULTURAL INSTITUTIONS

For as long as there have been humans, we have depended on culture. Beginning with language itself, human beings have had distinctive capacities to learn from the experiences of other humans, and these capacities are arguably the secret of human evolutionary success. Major technological changes in these capacities have led to dramatic social transformations. Spoken language was succeeded by pictures then by writing, print, film, and video. As more and more information became available across wider gulfs of space and time, new ways of accessing and organizing that information also developed, from libraries to newspapers to internet search. These developments have had profound effects on human thought and society, for better or worse. Eighteenth-century advances in print technology, for example, which allowed new ideas to quickly spread, played an important role in the Enlightenment and the French Revolution. A landmark transformation occurred around 2000 when nearly all the information from text, pictures, and moving images was converted into digital formats; it could be instantly transmitted and infinitely reproduced.

As long as there have been humans, we have also relied on social institutions to coordinate individual information-gathering and decision-making. These institutions can themselves be thought of as a kind of technology (*1*). In the modern era, markets, democracies, and bureaucracies have been particularly important. The economist Friedrich Hayek argued that the market's price mechanism generates dynamic summaries of enormously complex and otherwise unfathomable economic relations (*2*). Producers and buyers do not need to understand the complexities of production; all they need to know is the price, which compresses vast swathes of detail into a simplified but usable representation. Election mechanisms in democratic regimes focus distributed opinion toward collective legal and leadership decisions in a related way. The anthropologist Scott argued (*3*) that all states, democratic or otherwise, have managed complex societies by creating bureaucratic systems that categorize and systematize information. Markets, democracies, and bureaucracies have relied on mechanisms that generate lossy (incomplete, selective, and uninvertible) but useful representations well before the computer. Those representations both depend on and go beyond the knowledge and decisions of individual people. A price, an election result, or a measure such as gross domestic product (GDP) summarizes large amounts of individual knowledge, values, preferences, and actions. At the same time, these social technologies can also themselves shape individual knowledge and decision-making.

The abstract mechanisms of a market, state, or bureaucracy, like cultural media, can influence individual lives in crucial ways, sometimes for the worse. Central

[1]SNF Agora Institute and School of Advanced International Studies, Johns Hopkins University, Baltimore, MD, USA. [2]Department of Psychology, University of California, Berkeley, CA, USA. [3]Department of Philosophy, University of California, Berkeley, CA, USA. [4]Santa Fe Institute, Santa Fe, NM, USA. [5]Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA, USA. [6]Department of Machine Learning, Carnegie Mellon University, Pittsburgh, PA, USA. [7]Department of Sociology, University of Chicago, Chicago, IL, USA. Email: jevans@uchicago.edu

banks, for example, reduced the complexities of the financial economy down to a few key variables. This provided apparent financial stability but at the cost of allowing instabilities to build up in the housing market, which central banks paid little attention to, precipitating the 2008 global financial crisis (*4*). Similarly, markets may not represent "externalities" such as harmful carbon emissions. Integrating such information into prices through, for example, a carbon tax can help but requires state action.

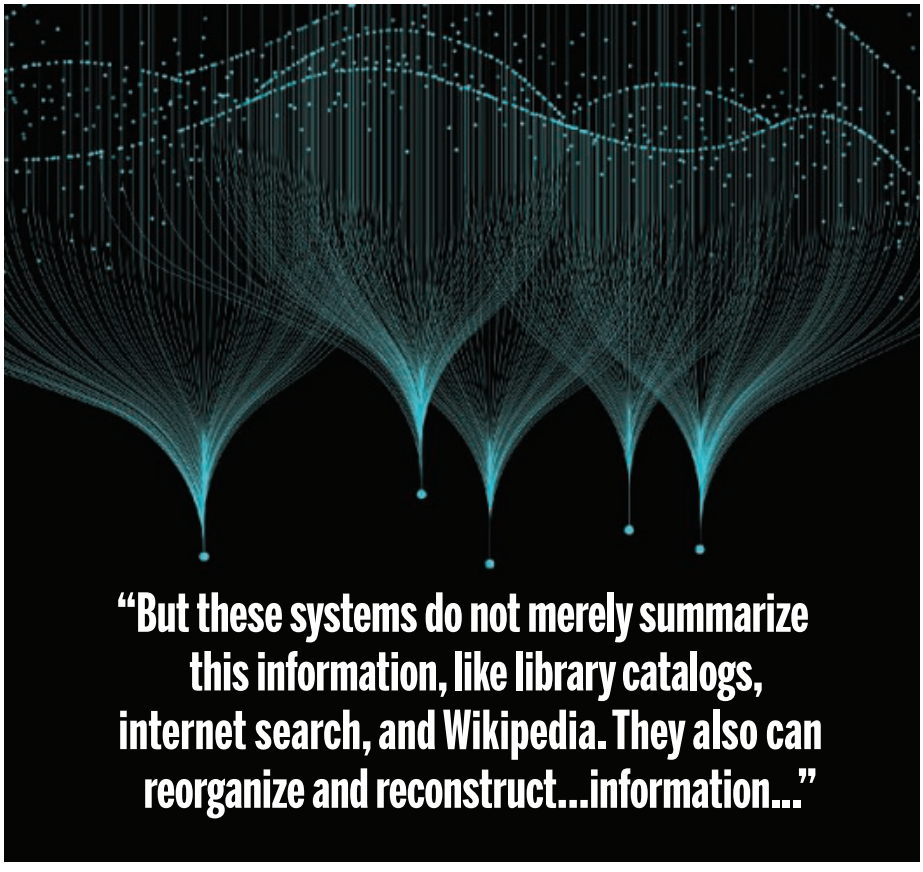
Humans rely extensively on these cultural and social technologies. These technologies are only possible, however, because humans have distinct capacities characteristic of intelligent agents. Humans, and other animals, can perceive and act on a changing external world, build new models of that world, revise those models as they accumulate more evidence, and then design new goals. Individual humans can create new beliefs and values and convey those beliefs and values to others through language or print. Cultural and social technologies transmit and organize those beliefs and values in powerful ways, but without those individual capacities, the cultural and social technologies would have no purchase. Without innovation, there would be no point to imitation (*5*).

Some AI systems—in robotics, for example—do attempt to instantiate similar truth-finding abilities. There is no reason, in principle, why an artificial system could not do so at some point in the future. Human brains do, after all. But at the moment, all such systems are far from these human capacities. We can debate how much to worry now about these potential future AI systems or how we might handle them if they emerge. But this is different from the question of the effects of large models at present and in the immediate future.

LARGE MODELS

Large models, unlike more agentive systems, have made notable and unexpected progress in the past few years, making them the focus of the current conversation about AI in general. This progress has led to claims that "scaling," simply taking the current designs and increasing the amount of data and computing power they use, will lead to AGI agents in the near future. But large models are fundamentally different from intelligent agents, and scaling will not change this. For example, "hallucinations" are an endemic problem in these systems because they have no conception of truth and falsity (although there are practical steps toward mitigation). They simply sample and generate text and images.

Rather than being intelligent agents, large models combine the features of cultural and social technologies in a new way. They generate summaries of unmanageably large and complex bodies of human-generated information. But these systems do not merely summarize this information, like library catalogs, internet search, and Wikipedia. They also can reorganize and reconstruct representations or "simulations" (*1*) of this information at scale and in new ways, like markets, states, and bureaucracies. Just as market prices are lossy representations of the underlying allocations and uses of resources, and government statistics and bureaucratic categories imperfectly represent the characteristics of underlying populations, so too are large models "lossy JPEGs" (*6*) of the data corpora on which they have been trained.



Because it is hard for humans to think clearly about large-scale cultural and social technologies, we have tended to think of them in terms of agents. Stories are a particularly powerful way to pass on information, and from fireside tales to novels to video games, they have done this by creating illustrative fictional agents, even though listeners know that those agents are not real. Chatbots are the successor to Hercules, Anansi, and Peter Rabbit. Similarly, it is easy to treat markets and states as though they were agents, and agencies or companies can even have a kind of legal personhood.

But behind their agent-like interfaces and anthropomorphic pretensions, large language models (LLM) and large multimodal models are statistical models that take enormous corpora of text produced by humans, break them down into particular words, and estimate the probability distribution of long word sequences. This is an imperfect representation of language but contains a surprisingly large amount of information about the patterns it summarizes. It allows the LLM to predict which words come next in a sequence and so generate human-like text. Large multimodal models do the same with audio, image, and video data. Large models not only abstract a very large body of human culture, they also allow a wide variety of new operations to be carried out on it. LLMs can be prompted to carry out complex transformations of the data on which they are trained. Simple arguments can be expressed in flowery metaphors, while ornate prose can be condensed into plain language. Similar techniques enable related models to generate new pictures, songs, and video in response to prompts.

ILLUSTRATION: ISTOCK/GETTY IMAGES PLUS

A body of cultural information that was previously too complex, large, and inchoate for large-scale operations has been rendered tractable.

In practice, the most recent versions of these systems depend not only on massive caches of text and images generated and curated by humans but also on human judgment and knowledge in other forms. In particular, the systems rely on reinforcement learning from human feedback (RLHF) or its variants: Tens of thousands of human employees provide ratings of model outputs. They also depend on prompt engineering: Humans must use both their background knowledge and ingenuity to extract useful information from the models. Even the newest "chain of thought" models regularly begin from dialogue with their human users.

The relatively simple though powerful algorithms that allow large models to extract statistical patterns from text are not really the key to the models' success. Instead, modern AI rests atop libraries, the internet, tens of thousands of human coders, and a growing international world of active users. Someone asking a bot for help writing a cover letter for a job application is really engaging in a technically mediated relationship with thousands of earlier job applicants and millions of other letter writers and RLHF workers.

CHALLENGES AND OPPORTUNITIES

The AI debate should focus on the challenges and opportunities that these new cultural and social technologies generate. We now have a technology that does for written and pictured culture what large-scale markets do for the economy, what large-scale bureaucracy does for society, and perhaps even comparable with what print once did for language. What happens next? Like past economic, organizational, and informational "general purpose technologies," these systems will have implications for productivity (*7*), complementing human work but also automating tasks that only humans could previously perform, and for distribution, affecting who gets what (*8*).

Yet they will also have wider and more profound cultural consequences. We do not yet know whether these consequences will be as great as those of earlier technologies such as print, markets, or bureaucracies, but thinking of them as cultural technologies increases rather than decreases their potential impact. These earlier technologies were central to the extensive social transformations of the 18th and 19th centuries, both as causes and effects. All of these technologies, like large models, supported the abstraction of information so that new kinds of operations could be carried out at scale. All provoked justified concerns about the spread of misinformation and bias, cultural homogenization or fragmentation, and shifts in the distribution of power and resources. The emergence of new communications media, including both print and television, was accompanied by reasonable worries that the new media would spread misinformation and strengthen malign cultural forces. Similarly, the categorization schemes that bureaucracies and markets deploy often embed oppressive assumptions.

At the same time, these technologies generated new possibilities for recombining information and coordinating actions among millions of people at a planetary scale. Emerging debates over the social, economic, and political consequences of LLMs continue deep-rooted historical worries and hopes about new cultural and social technologies. Orienting these debates requires both recognizing the commonalities between new arguments and old ones and carefully mapping the particulars of the new and evolving technologies.

Such mapping is among the central tasks of the social sciences, which emerged from the social, economic, and political upheavals of the Industrial Revolution and its aftermath. Social scientists' investigation of the consequences of these past technologies can help us think about less obvious social implications of AI, both negative and positive, and to consider ways that AI systems could be redesigned to increase the positive impacts and reduce the negative. As media, markets, and bureaucratic technologies expanded in the 19th and 20th centuries, they generated economic losers and winners, displacing whole categories of workers, from clerks and typists to "human computers." Today, there are obvious worries that large models and related technologies may displace "knowledge workers."

There are also less obvious questions. Will large models homogenize or fragment culture and society? Thinking about this in historical context can be particularly illuminating. Current concerns resemble 19th- and 20th-century disagreements over markets and bureaucracies. Weber worried (*9*) about the deadening homogenizing consequences of economic and bureaucratic "rationalization," whereas Mill (*10*) thought that market exchanges would expose participants to different forms of life and soften impulses to conflict ("doux commerce").

Large models are designed to work well—to faithfully reproduce the actual probabilities of sequences of text, images, and video—on average. They therefore have an intrinsic tendency to be most accurate in situations most commonly found in their training data and least accurate in situations that were rare in data or entirely new. This might lead large models to worsen the kind of homogenization that haunted Weber.

On the other hand, large models may allow us to design new ways to harvest the diversity of the cultural perspectives they summarize. Combining and balancing these perspectives may provide more sophisticated means of solving complex problems (*11*). One way to do this may be to build "society-like" ecologies in which different perspectives, encoded in different large models, debate each other and potentially cross-fertilize to create hybrid perspectives (*12*) or to identify gaps in the space of human expertise (*13*) that might usefully be bridged. Large models are surprisingly effective at abstracting subtle and nonobvious patterns in texts and images. This suggests that such technologies could be used to find patterns in text and images that crisscross the space of human knowledge and culture, including patterns invisible to any particular human. We may require new systems that diversify large model reflections and personas and produce the same distribution and diversity as do human societies.

Diversifying systems like this might be particularly important for scientific progress. Formal science itself depended on the emergence of the new cultural technologies of the 17th and 18th centuries, from coffee houses and rapid mail to journals and peer review. AI technologies have the potential to accelerate science further, but this will depend on imaginative ways of using and rethinking these technologies. By wiring together so many perspectives across text, audio, and images, large models may allow us to discover unprecedented connections between them for the benefit of science and society. These technologies have most commonly been trained to regurgitate routine information as helpful assistants. A more fundamental set of possibilities might open up if we deployed them as maps to explore formerly uncharted territory.

There are also less obvious and more interesting ways that new cultural and social technologies influence economic relationships. The development of cultural technologies leads to a fundamental economic tension between the people who produce information and the systems that distribute it. Neither group can exist without the other: A writer needs publishers as much as the publisher need writers. But their economic incentives push in opposite directions. The distributors will profit if they can access the producer's information cheaply, whereas the producers will profit if they can get their information distributed cheaply. This ten-

sion has always been a feature of new cultural technologies. The ease and efficiency of distributing information in digital form has already made this problem especially acute, as evidenced by the crisis in everything from local newspapers to academic journals. But the very speed, efficiency, and scope of large models, processing all the available information at once, combined with the centralized ownership of those models, makes these problems loom especially large. Concentrated power may make it easier for those who own the systems to skim the benefits of efficiency at the expense of others.

There are crucial technical questions: To what extent can the systematic imperfections of large models be remedied, and when are they better or worse than the imperfections of systems based around human knowledge workers? Those should not overshadow the crucial political questions: Which actors are capable of mobilizing around their interests, and how might they shape the resulting mix of technology and organizational capacities? Very often, commentators within the technology sector reduce these questions into a simple battle between machines and humans. Either the forces of progress will prevail against retrograde Luddite tendencies, or on the other hand, human beings will successfully resist the inhuman encroachment of artificial technology. Not only does this fail to appreciate the complexities of past distributional struggles, struggles that long predate the computer, it ignores the many different possible paths that future progress might take, each with its own mix of technological possibilities and choices (*8*).

In the case of earlier social and cultural technologies, a range of further institutions, including normative and regulatory institutions, emerged to temper their effects. These ranged from editors, peer review, and libel laws for print, to election law, deposit insurance, and the Securities and Exchange Commission for markets, democracies, and bureaucracies. These institutions had varied effectiveness and required continual revision. These countervailing forces did not emerge on their own, however, but resulted from concerted and sustained efforts by actors both within and outside the technologies themselves.

LOOKING FORWARD

The narrative of AGI, of large models as superintelligent agents, has been promoted both within the tech community and outside it, both by AI optimist "boomers" and more concerned "doomers." This narrative gets the nature of these models and their relation to past technological changes wrong. But more importantly, it actively distracts from the real problems and opportunities that these technologies pose and the lessons history can teach us about how to ensure that the benefits outweigh the costs.

Of course, as we note above, there may be hypothetical future AI systems that are more like intelligent agents, and we might debate how we should deal with these hypothetical systems, but LLMs are not such systems, any more than were library card catalogs or the internet. Like catalogs and the internet, large models are part of a long history of cultural and social technologies.

The social sciences have explored this history in detail, generating a distinct understanding of past technological upheavals. Bringing computer science and engineering into close cooperation with the social sciences will help us to understand this history and apply these lessons. Will large models lead to greater cultural homogeneity or greater fragmentation? Will they reinforce or undermine the social institutions of human discovery? As they reshape the political economy, who will win and lose? These and other urgent questions do not come into focus in debates that treat large models as analogs for human agents.

"...large models are themselves best understood as a particular type of cultural and social technology."

Changing the terms of debate would lead to better research. It would be far easier for social scientists and computer scientists to cooperate and combine their respective strengths if both understood that large models are no more—but also no less—than a new kind of cultural and social technology. Computer scientists could bring together their deep understanding of how these systems work with social scientists' comprehension of how other such large-scale systems have reshaped society, politics, and the economy in previous eras, elaborating existing research agendas and discovering new ones. This would help remedy past confusions in which computer scientists have adopted overly simplified notions of complex social phenomena (*14*) while social scientists have failed to understand the complex functioning of these new technologies.

It would move policy discussions over AI decisively away from simplistic battles between the existential fear of a machine takeover and the promise of a near-future paradise in which everyone will have a perfectly reliable and competent artificial assistant. The actual policy consequences of large models will surely be different. Like markets and bureaucracies, they will make some kinds of knowledge more visible and tractable than they were in the past, encouraging policy-makers to focus on the new things that they can measure and see at the expense of those less visible and more confusing. As a result, reflecting past cases of markets and media, power and influence will shift toward those who can fully deploy these technologies and away from those who cannot. AI weakens the position of those on whom it is used and who provide its data, strengthening AI experts and policy-makers (*14*).

Last, thinking in this way might reshape AI practice. Engineers and computer scientists are already aware of the problem of large model bias and are thinking about their relationship to ethics and justice. They should go further. How will these systems affect who gets what? What will their practical consequences be for societal polarization and integration? Can large models be developed to enhance human creativity rather than to dull it? Finding practical answers to such questions will require an understanding of social science as well as engineering. Shifting the debate about AI away from agents and toward cultural and social technologies is a crucial first step toward building that cross-disciplinary understanding (*15*). ■

REFERENCES AND NOTES

1. H. Simon, *The Sciences of the Artificial* (MIT Press, 1996).
2. F. A. von Hayek, *Am. Econ. Rev.* **35**, 519 (1945).
3. J. C. Scott, *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed* (Yale Univ. Press, 1998).
4. D. Davies, *The Unaccountability Machine* (Univ. Chicago Press, 2025).
5. E. Yiu, E. Kosoy, A. Gopnik, *Perspect. Psychol. Sci.* **19**, 874 (2024).
6. T. Chiang, *New Yorker* 9 (2023).
7. C. Goldin, L. Katz, *Q. J. Econ.* **113**, 693 (1998).
8. D. Acemoglu, S. Johnson, *Power and Progress: Our 1000 Year Struggle over Technology and Prosperity* (Hachette, 2023).
9. M. Weber, *Wissenschaft Als Beruf* (Duncker & Humblot, 1919).
10. J. S. Mill, *Principles of Political Economy* (Longmans and Green, 1920).
11. L. Hong, S. E. Page, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16385 (2004).
12. S. Lai *et al.*, *Proc. 41st Int. Conf. Mach. Learn.* **235**, 25892 (2024).
13. J. Sourati, J. A. Evans, *Nat. Hum. Behav.* **7**, 1682 (2023).
14. S. L. Blodgett, S. Barocas, H. Daumé, H. Wallach, *arXiv*:2005.14050 [cs.CL] (2020)
15. L. Brinkmann *et al.*, *Nat. Hum. Behav.* **7**, 1855 (2023).

ACKNOWLEDGMENTS

All authors contributed equally to this work. J.E. began a visiting researcher affiliation with Google after this manuscript was submitted.

10.1126/science.adt9819

PHOTO: GORODENKOFF/ISTOCK PHOTO

Scientists hoping to convert research into market-ready products can benefit from specialized training.

LETTERS

Edited by **Jennifer Sills**

Support PhDs building deep-tech ventures

PhDs in science, technology, engineering, and mathematics (STEM) increasingly pursue start-ups as a full-time career choice. Between 2012 and 2023, an average of 13% of recent Massachusetts Institute of Technology (MIT) PhD graduates worked at start-ups, and an average of 4% of graduates per year cofounded companies (*1*). The percentage of PhD founders backed by venture capital firms has also increased (*2*). However, educational programs in entrepreneurship often leave behind this growing cohort of would-be entrepreneurs (*3*).

Most entrepreneurship programs focus on "dorm-room" start-ups created by undergraduates or MBAs with software ideas that can be tested in short sprints. This model fails to address the needs of laboratory-based endeavors—referred to as deep-tech ventures—that transform scientific research into market-ready innovations. Intellectual property founded on breakthrough science requires more attention to technical, market, regulatory, and production risks.

To prepare PhDs to become founders of ventures based on their own research, universities and innovation ecosystem partners must establish dedicated training for and must remove barriers to the deep-tech venture pathway (*4*). Although some programs, such as the National Science Foundation Innovation Corps (I-Corps) program (*5*), address specific gaps in the traditional PhD experience, most existing programs do not scale to thousands of scientist-entrepreneurs, do not provide a full suite of support, or remain inaccessible until after graduation. An effective program must scale while supporting STEM PhDs through three distinct phases: initial venture curiosity, serious exploration of a laboratory-based project (during their program), and the creation of a so-called "spin-out" company.

Early-stage support should focus on deep-tech venture-building skills. For example, MIT's Engine Lab course (*3*) and Harvard's Lab to Market online course (*6*) ensure that students undertake projects to build entrepreneurial skills, network, and find role models. Beyond universities, venture firms, such as Flagship Pioneering, offer summer fellowships (*7*) for PhDs and postdocs to learn the venture-building process.

If PhD students want to commercialize a specific project, educational programming should help them evaluate specific market opportunities and assess commercialization risks. The Engine's Blueprint program (*8*) provides this guidance, and more broadly, the suite of facilities at programs such as the Cornell Tech Runway Startup Postdocs program (*9*) and University of California Berkeley Postdoctoral Entrepreneurship Program (*10*) offer essential infrastructure that supports translational activities.

After they graduate, PhDs need help building their companies. Fellowships, such as those offered by Activate (*11*) and Breakthrough Energy (*12*), give recent PhDs the freedom to pursue applied research with the goal of commercialization. These programs combine stipends, curriculum, research and development funding, networks, and low-cost access to experimental equipment and facilities—essential elements for scientist-entrepreneurs transitioning discoveries to market.

The future of US innovation depends on a robust academia-to-venture pipeline for STEM PhDs. Implementing a comprehensive support system for ambitious scientist-entrepreneurs can substantially increase the likelihood that scientific advances achieve real-world impact.

Theo Diamandis[1], Josué J. López[2,3], Fiona Murray[4]*
[1]Bain Capital, New York, NY, USA. [2]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]Fine Structure Ventures–F-Prime Capital, Cambridge, MA, USA. [4]Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, USA.
*Corresponding author. Email: fmurray@mit.edu

REFERENCES AND NOTES

1. Recent Alumni Survey, *MIT Institutional Research;* https://ir.mit.edu/projects/recent-alumni-survey/. To find the percentages cited, click the "Working – 2" tab. On the "Select Breakout" dropdown menu, choose "Highest Degree." See the "Doctorates" row for "This is a start-up company" and "I am a founding member of this company or organization."
2. P. Zhegin, "How may PhD startup founders build for the long term?" (peter zhegin | future platform, 2021); https://www.peterzhegin.com/how-may-ph-d-startup-founders-build-for-the-long-term/.
3. S. Brown, T. Mayor, M. Somers, "These 5 new courses track emerging business trends and skills" (MIT Sloan School of Management, 2023); https://mitsloan.mit.edu/ideas-made-to-matter/these-5-new-courses-track-emerging-business-trends-and-skills.
4. T. Diamandis *et al.*, *Technol. Innov.* 10.1080/19498241.2024.2438581 (2025).
5. US National Science Foundation, NSF I-Corps™; https://new.nsf.gov/funding/initiatives/i-corps.
6. Laboratory for Innovation Science at Harvard, Technology Entrepreneurship: Lab to Market; https://lish.harvard.edu/technology-commercialization-mooc.
7. Flagship Pioneering, Flagship Pioneering Fellowship; https://www.flagshippioneering.com/join/fellows.
8. The Engine, Blueprint: Turning Tough Tech breakthroughs into Tough Tech startups; https://engine.xyz/blueprint.
9. Cornell Tech, Runway Startup Postdocs; https://tech.cornell.edu/programs/phd/startup-postdocs/.
10. UC Berkeley, Berkeley Postdoctoral Entrepreneurship Program; https://bpep.berkeley.edu/home.
11. Activate, The Activate Fellowship; https://www.activate.org/the-fellowship.
12. Breakthrough Energy, Breakthrough Energy Fellows; https://www.breakthroughenergy.org/our-work/fellows/.

COMPETING INTERESTS

J.J.L. was an Activate Fellow from 2020 to 2022. F.M. is vice chair of the NATO Innovation Fund and nonexecutive director of Avathon.

10.1126/science.adt1709

The problem with AI dialogue at scale

In their Research Article "Durably reducing conspiracy beliefs through dialogues with AI" (13 September 2024, p. 1183), T. H. Costello *et al.* provide evidence that dialogue with generative artificial intelligence (AI) can reduce conspiracy beliefs. The paper's conclusions frame AI as a potentially scalable intervention that can save societies from misinformation and conspiracy theories (*1*, *2*). However, the applicability of this strategy to real-world problems remains unclear.

Volumes of scientific information and consistency of messaging are not enough to change entrenched conspiracy beliefs, especially among those likely to distrust scientific institutions (*3*). For example, flat-Earth theory resurged in the early 21st century even though consistent messages about Earth's sphericity had reached most adherents (*4*). During the pandemic, scientists observed that "persuasive words are not enough" (*5*) to break through deeply held beliefs.

In the 2000s, evidence from lab studies suggested that "framing" could combat climate change misinformation by making scientific evidence more relatable, relevant, and understandable to the public. However, meta-analyses later showed that the positive effects of framing were clearer in the lab than in the real world (*6*). In practice, AI dialogue would face similar challenges.

If AI dialogues were scaled up, they would have to interact with users who harbor beliefs for a variety of reasons. Messaging that is culturally appropriate and effective in one setting would likely be inappropriate and ineffective in another (*7*). Different strategies are required to combat misinformation in communities that resist global vaccine campaigns because they lack access to decent health care or have suffered historical injustices compared with those whose resistance to science is based in religious or political ideology.

We agree with Costello *et al.* that making people feel heard through personalized conversation is crucial for building trust—the foundation of effective communication (*8*). The focus on AI should not overshadow the vital role of a trusted messenger. During the pandemic, people turned to trusted sources, regardless of the accuracy of the information. AI dialogues may have the capacity to inform and engage conspiracy adherents, but the strategy's effectiveness would be undermined if used only for constant point-counterpoint debate. Because resistance to official science often has little to do with the science itself or the evidence provided (*9*, *10*), the solution to misinformation may be social policy rather than more information or tailored counterarguments to conspiracies.

Ehsan Nabavi*, Faranak Hardcastle, Joan Leach, Fabien Medvecky, Merryn McKinnon, Sujatha Raman, Dan Santos
Australian National Centre for the Public Awareness of Science, The Australian National University, Canberra, ACT, Australia.
*Corresponding author.
Email: ehsan.nabavi@anu.edu.au

REFERENCES AND NOTES

1. B. Bago, J.-F. Bonnefon, *Science* **385**, 1164 (2024).
2. H. H. Thorp, *Science* **385**, 1143 (2024).
3. B. T. Rutjens, B. Većkalov, *Curr. Opin. Psychol.* **46**, 101392 (2022).
4. C. Garwood, *Flat Earth: The History of an Infamous Idea* (Thomas Dunne Books, 2008).
5. H. H. Thorp, *Science* **368**, 1405 (2020).
6. E. Amsalem, A. Zoizner, *Br. J. Polit. Sci.* **52**, 221 (2022).
7. T. Sesan, *Technol. Soc.* **39**, 142 (2014).
8. C. A. Roos, T. Postmes, N. Koudenburg, *PLOS ONE* **18**, e0292865 (2023).
9. International Science Council (ISC), "Flipping the science model: A roadmap to science missions for sustainability" (ISC, 2023).
10. N. Ishmael-Perkins *et al.*, "The contextualization deficit: Reframing trust in science for multilateral policy" (The Centre for Science Futures, 2023).

10.1126/science.adu1526

Response

Our Research Article shows that evidence-based artificial intelligence (AI) dialogues reduce conspiracy beliefs much more than expected on the basis of previous work. Nabavi *et al.* agree with our conclusion that personalized conversations are crucial in addressing epistemically unfounded beliefs. However, they argue that this approach alone will not be enough to solve the misinformation challenge. Of course, we agree that no single solution, including AI chatbots, will address this challenge (*1*). We did not claim that AI dialogues alone could "save societies from misinformation and conspiracy theories" but rather that the strategy could serve as part of a toolkit to combat inaccurate beliefs.

Nabavi *et al.* highlight the limits of interventions that lack contextualization, given that people are not homogeneous. These concerns are consistent with our paper, in which we explicitly argued that human-AI dialogues were likely effective because users expressed their beliefs in their own words. This allowed the AI to leverage evidence most relevant to each individual conspiracist. The authors mention cultural variability, but psychological variation among individuals within the same culture far exceeds differences across cultures (*2*). Tailoring persuasive messaging based on cultural context (and other less predictive cues, such as demographic information, as in so-called microtargeting) (*3*) may be less effective than truly bespoke messaging based on a person's unique belief system.

Nonetheless, as we stated in our paper, AI dialogues will not persuade everyone, and learning how to get people to choose to engage with the AI is a key step for achieving large-scale impact. However, we are optimistic about this approach's potential reach, given that conspiracy believers often emphasize the importance of "doing your own research" and engaging in debate. This aligns will our real-world observations: Since publication, our DebunkBot.com website (which allows users to discuss their conspiracy beliefs with GPT-4) has attracted nearly 100,000 visitors.

Critically, the system-level impacts of widely available persuasive argumentation for any given epistemic perspective remain unknown. In such circumstances, truth could provide the edge—indeed, this is the principle on which both adversarial legal systems and the institution of science itself are based (*4*). Successful argumentation also serves as a credible signal corroborating the veracity (or, at least, the thoughtfulness) of one's claims and thus may not always require "trusted messengers" or aligned incentives between messenger and audience to be persuasive (*5*). Exposure to conflicting positions in conversation with one another may increase nuance and ward against extremism [as it did in previous work (*5*, *6*)]. These possibilities warrant further investigation rather than summary dismissal.

Our Research Article focuses on scientific questions rather than applications. We did not develop a ready-to-deploy intervention but rather showed that the field is mistaken about the obstinacy of conspiracy belief. Our results suggest that evidence-based argumentation approaches, which many had prematurely abandoned, can be relatively effective, a point with which Nabavi *et al.* seem to agree.

Thomas H. Costello[1,2]*, Gordon Pennycook[3], David G. Rand[2]
[1]Department of Psychology, American University, Washington, DC, USA. [2]Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]Department of Psychology, Cornell University, Ithaca, NY, USA.
*Corresponding author.
Email: thcostello1@gmail.com

REFERENCES AND NOTES

1. L. Bode, E. Vraga, *Bull. At. Sci.* **77**, 129 (2021).
2. P. H. P. Hanel, G. R. Maio, A. S. R. Manstead, *J. Pers. Soc. Psychol.* **116**, 541 (2019).
3. B. M. Tappin, C. Wittenberg, L. B. Hewitt, A. J. Berinsky, D. G. Rand, *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2216261120 (2023).
4. J. Habermas, *Reason and the Rationalization of Society*, vol. 1 of *The Theory of Communicative Action*, Transl. T. MacCarthy (Beacon Press, 2007).
5. H. Mercier, *Not Born Yesterday: The Science of Who We Trust and What We Believe* (Princeton Univ. Press, 2020).
6. M. H. Tessler *et al.*, *Science* **386**, eadq2852 (2024).

10.1126/science.adu6608

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

CHIRAL ELECTRONICS

Efficient circularly polarized OLED

There is considerable interest in engineering chirality into semiconductor materials, with the goal of achieving high levels of circularly polarized luminescence (CPL). Currently, efficient organic light-emitting diode (OLED) systems use light-emitting molecules spatially isolated in a host, resulting in low levels of CPL. Previous attempts to build uniform chiral supramolecular order have not been compatible with optimized OLED device architectures. Chowdhury *et al*. found a way to fabricate thin, uniform films with chiral supramolecular nanostructures based on triazatruxene molecules. The method is fully compatible with OLED fabrication and demonstrated strong green CPL. This work opens new avenues for development in organic electronics driven by circularly polarized light. —Yury Suleymanov *Science* p. 1175, 10.1126/science.adt3011

Confocal microscopy image of a patterned triazatruxene thin film in an OLED

PHOTO: SAMARPITA SEN AT THE UNIVERSITY OF CAMBRIDGE, RENDERED BY RITUPARNO CHOWDHURY. THE UNIVERSITY OF CAMBRIDGE

PALEONTOLOGY

No flashy colors for early mammals

In the past decade or so, the identification of melanosomes in some dinosaur feathers and skin has led to the suggestion that many bore bright and complex coloration. Less work has been done to explore the coloration of early mammals. Li *et al*. looked at melanosome patterns in multiple species of Mesozoic mammals and compared these with those seen in more than 100 species of extant mammals. There was little color variation, and the animals were entirely dark in color. The authors argue that this was likely due to the animals' nocturnal nature, and that brighter and more varied colors in mammals may have arisen after the extinction of dinosaurs at the end of the Cretaceous period. —Sacha Vignieri

Science p. 1193, 10.1126/science.ads9734

SUSTAINABILITY

Water limits mineral production

Mineral production is key to our global economy, with new technologies creating higher demand for many minerals. However, large amounts of water are required for mining and processing, which could limit mineral production in some locations. Islam *et al*. evaluated these constraints using published data on mineral production and water requirements for different minerals, coupled with regional water-carrying capacities from a hydrologic model. They found that mineral production exceeds water resources in many regions because of high production or low water availability. Coal, iron, copper, and gold showed some of the highest overconsumption, coal because of its high production rates and the metals because they require more water to process. Water requirements for mineral production are expected to increase in the future. —Bianca Lopez

Science p. 1214, 10.1126/science.adk5318

ORGANIC CHEMISTRY

A selective aryl cleaver

One of the defining characteristics of aryl compounds such as benzene is their greater stability compared with other arrangements of carbon-carbon double bonds. If one of the aryl bonds is broken, those remaining tend to be more reactive. Olivier *et al*.

report a reaction manifold that reverses this prevailing trend. A nitroarene reagent can be photoexcited to a state that preferentially reacts with arenes over olefins, likely through transient complexation, thereby enabling clean oxidative cleavage of just one aryl bond. —Jake S. Yeston

Science p. 1167, 10.1126/science.ads3955

ENVIRONMENTAL SCIENCE

Spoiled by war

Along with the enormous costs to human life and livelihoods, war creates environmental catastrophes. In the ongoing war in Ukraine, dams have been the targets of military attacks, creating floods and severely diminishing water resources for agricultural irrigation. Shumilova *et al*. measured and modeled some of the less well-recognized environmental impacts of dam destruction, focusing on the Kakhovka Dam in southern Ukraine. The destruction of the dam in June 2023 released pollutants that had accumulated in the reservoir sediments, creating a long-term source of contaminants that could be spread by future floods. Fast-growing riparian vegetation could help recover the floodplain of the former lake bed in the short term, but actions must be taken to address contamination in the region. —Bianca Lopez

Science p. 1181, 10.1126/science.adn8655

ATMOSPHERE

A new way

Nitrous oxide (N_2O) is an important greenhouse gas with an atmospheric abundance that is increasing faster than predicted. It is the chemical species primarily responsible for stratospheric ozone depletion. Leon-Palmero *et al*. describe a pathway that is absent from current greenhouse gas budgets in which sunlight drives abiotic N_2O formation in oxic fresh and marine surface waters. They found that production rates are proportional to radiation flux and estimate that this pathway creates more N_2O than biological production in surface waters, thus constituting a major source of N_2O emissions globally. —Jesse Smith

Science p. 1198, 10.1126/science.adq0302

AUTOIMMUNITY

Photosensitivity features Z-DNA

Autoimmune photosensitivity is a common symptom of type I interferon–driven skin diseases such as cutaneous lupus erythematosus (CLE). Mechanisms mediating this response have been linked to ultraviolet B (UVB) light–induced mitochondrial damage in keratinocytes. Klein *et al*. show that mitochondrial Z-DNA accumulates and is released by CLE keratinocytes upon UVB light exposure. This Z-DNA is stabilized by Z-DNA binding protein 1 (ZBP1) and potentiates the type I interferon response through the cGAS-STING signaling pathway. ZBP1 was also shown to accumulate in samples from patients with lupus-associated photosensitivity. —Christiana N. Fogg

Sci. Immunol. (2025) 10.1126/sciimmunol.ado1710

DNA DAMAGE

NEMO finds DNA damage

DNA damage is recognized and repaired by a series of protein complexes that are often recruited to sites of damage by the kinase ATM. Alonso-Marañón *et al*. found that ATM complexes are directed to sites of damage by the protein NEMO, a regulatory subunit in the NF-κB pathway (see the Focus by Perkins). In cells exposed to ultraviolet light or a chemotherapeutic, NEMO bound to ATM in complex with the kinase IKKα and directed the pair to damaged DNA. Increased expression of NEMO in colorectal cancers correlated with chemoresistance and poor prognosis, suggesting that NEMO may predict tumors, which could then be treated with DNA repair inhibitors. —Leslie. K. Ferrarelli

Sci. Signal. (2025) 10.1126/scisignal.adr0128

IN OTHER JOURNALS

Edited by **Corinne Simonti** *and* **Jesse Smith**

PHOTO: SEAN BAGSHAW/SCIENCE SOURCE

PLANT FLUID DYNAMICS

Seagrasses wave goodbye to erosion

Seagrasses grow in many coastal regions and can help prevent erosion. Vettori *et al.* placed plastic seagrass replicas in a 50-meter-long wave flume to understand how seagrasses affect hydraulic forces of waves and currents. They found that the presence of a current parallel to the waves generated turbulent water flow. This caused damping of the waves in a manner that was not predicted by existing mathematical models of wave attenuation. This work provides a large experimental dataset alongside mathematical modeling to understand how seagrasses might affect the forces that waves exert on the shore. —Madeleine Seale

Proc. Natl. Acad. Sci. U.S.A. (2025) 10.1073/pnas.2414150122

Sea grass protects coastlines from erosion by weakening waves.

PHOTO: NEW YORK DAILY NEWS ARCHIVE/CONTRIBUTOR

QUANTUM CHEMISTRY

Density functionals for large systems

Density functional theory (DFT) is currently the most reliable framework for computational studies of materials based on the fundamental concept of representing the total energy of a system through its electron density, such as Hartree-Fock (HF) density. However, performing quantitative HF-DFT calculations on large systems containing hundreds of atoms remains a challenge. Recent efforts to develop machine learning (ML)–based functionals have shown improvements in accuracy and efficiency; however, because of the black-box nature of ML, these methods compromise transparency, often leading to a lack of confidence. Kim *et al.* demonstrated that the dual-basis scheme with proper dispersion correction, which offers more transparency, can be used to reliably and efficiently extend HF-DFT to large molecular systems such as DNA-ellipticine and HIV-indinavir. —Yury Suleymanov

J. Phys. Chem. Lett. (2025)
10.1021/acs.jpclett.4c02852

CRYOSPHERE

Kept in a cold embrace

Permafrost is a time capsule of ancient flora, fauna, and even the remains of ice sheets. Such relict ice mostly dates back about 20,000 years, to the last glacial advance. Coulombe *et al.* report on much older ice, two massive blocks in the Canadian Arctic that have rested undisturbed for hundreds of thousands of years. Combining radiocarbon and paleomagnetic ages with ice crystallography, the authors observed that the ice survived prior cycles of warming due to an especially thick permafrost. The ice is now exposed in upland slumps because of thaw. —Angela Hessler

Geology (2025)
10.1130/G52446.1

NEUROSCIENCE

Selective astrocyte manipulation

Recent data have shown that astrocytes and other non-neuronal cells can modulate neuronal activity and affect behavioral outcomes in preclinical models. Serra *et al.* developed a tool called AstroLight to manipulate specified astrocyte subpopulations activated during a behavioral task. They then used this tool to identify a subset of astrocytes in the nucleus accumbens that are associated with cue-motivated behaviors. These results demonstrate that selective manipulation of a specific astrocyte subpopulation in the nucleus accumbens can affect cue rewarded-associated behavior in mice, and the potential usefulness of this tool in other functional studies. —Mattia Maroso

Nat. Neurosci. (2025)
10.1038/s41593-025-01870-0

BIOMATERIALS

Sopping up sodium to aid wound repair

Skin wound healing is promoted by a number of factors, including biochemistry, inflammatory responses, and electrical stimulation. For example, endogenous electrical fields promote the proliferation of epithelial cells and their migration from the wound edge, and these are driven by gradients in ions such as Na^+ and Cl^-. Luo *et al.* developed a wound dressing based on polyurethane and carbon, the latter formed from the pyrolysis of ginko leaves. The carbon soaks up Na^+ flowing to the center of the wound site, thus enhancing the electric field. Animal studies showed how this could shorten the inflammatory phase of the wound and accelerate wound closure. —Marc S. Lavine

Adv. Mater. (2025)
10.1002/adma.202419149

PREGNANCY

Polluted pregnancy

Infants are increasingly "born late," after 41 weeks of pregnancy. These prolonged pregnancies increase birth complications and pose other health risks for the child and mother. Nyadanu *et al.* examined the role of two environmental stressors, fine particulate air pollution ($PM_{2.5}$) and extreme weather and temperatures, on pregnancy term for nearly 400,000 Western Australia births from 2000 to 2015. Higher climate-related exposures were associated with elevated odds of prolonged pregnancies, particularly among mothers who live in urban areas, are age 35+, are pregnant for the first time, or are carrying boys. The associations suggest that more research on $PM_{2.5}$ and prolonged pregnancies in other geographic areas is needed to better understand the possible impact of climate change on late births. —Ekeoma Uzogara

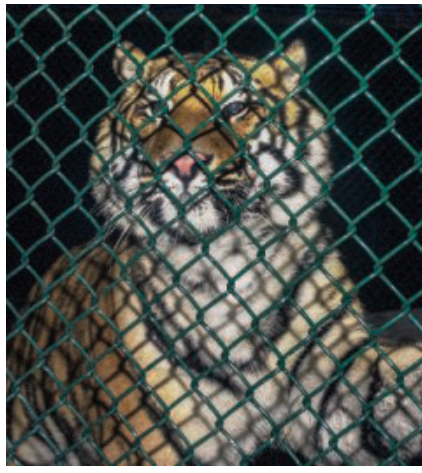
Urban Clim. (2025)
10.1016/j.uclim.2025.102316

CONSERVATION BIOLOGY

The ancestry of Ming

In 2003, a 400-pound tiger named Ming was discovered in a Harlem apartment, sparking media interest and forensic investigation. Using whole-genome sequencing of trace materials from Ming's whiskers, Wu *et al.* performed a forensic genomic analysis of Ming and his putative sibling Cheeky. Both tigers had admixed genetic backgrounds from five subspecies, with Indochinese tiger ancestry being predominant, and were confirmed as full siblings. The study highlights the potential of genomic tools for tiger conservation, particularly in identifying the origins of captive tigers and combating illegal wildlife trafficking. The findings also underscore the importance of genetic surveillance in preserving endangered species. —Di Jiang

J. Hered. (2025)
10.1093/jhered/esaf003



A study uncovered the genetic history of Ming, the tiger found in a New York apartment

RESEARCH ARTICLE SUMMARY

MICROBIOLOGY

Trypanosome doublet microtubule structures reveal flagellum assembly and motility mechanisms

Xian Xia, Michelle M. Shimogawa, Hui Wang, Samuel Liu, Angeline Wijono, Gerasimos Langousis, Ahmad M. Kassem, James A. Wohlschlegel, Kent L. Hill*, Z. Hong Zhou*

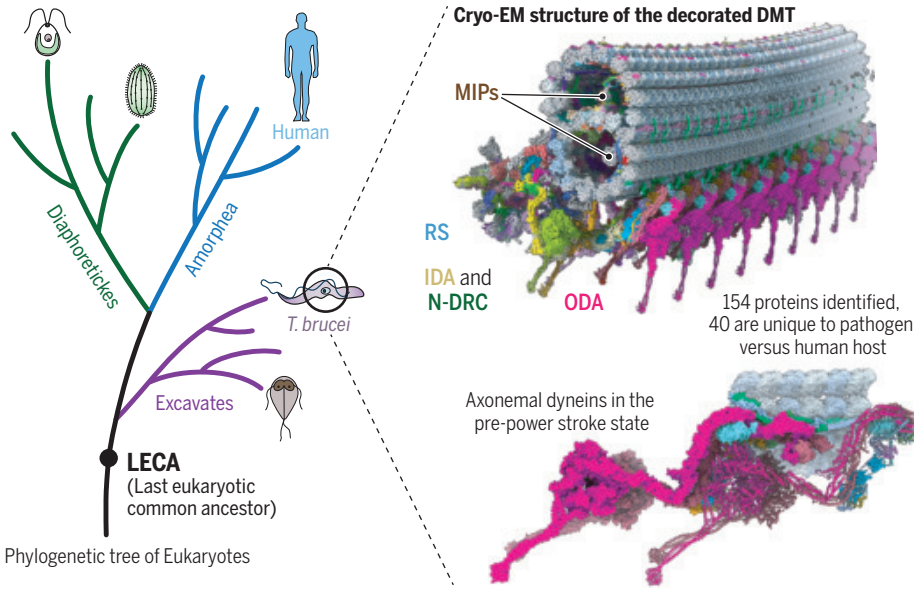
INTRODUCTION: Parasitic trypanosomes, including *Trypanosoma brucei* and related species, cause severe disease in humans and other mammals. As excavates that diverged early within the eukaryotic lineage, trypanosomes also offer great opportunities for novel biological discovery. The single flagellum of *T. brucei* drives its motility and plays key roles in cell division, morphogenesis, and host interactions. Unlike most organisms, the trypanosome flagellum beats with a helical waveform that propagates tip to base. Central to flagellar motility is a conserved 9+2 axoneme, in which nine doublet microtubules (DMTs) surround a pair of singlet microtubules, with radial spokes (RSs) projecting inward from the DMTs to contact the central pair complex. Flagellar beating is powered by thousands of dynein motors arrayed between DMTs, with outer (ODA) and inner (IDA) dynein arms performing complementary roles and regulated by other axonemal complexes such as the nexin-dynein regulatory complex (N-DRC). An extra-axonemal filament unique to trypanosomes and other *Euglenozoa*, the paraflagellar rod (PFR), is attached alongside the axoneme, providing elastic resistance against axoneme bending.

RATIONALE: Recent cryo–electron microscopy (cryo-EM) studies have resolved high-resolution structures of DMTs from different organisms, revealing core features of flagellum assembly and motility and also identifying lineage-specific adaptations. Notably lacking in these analyses are members of the excavata clade, which includes trypanosomes and other devastating pathogens. This presents a critical knowledge gap in our understanding of pathogen biology and evolution of mechanisms for assembly and operation of one of the most iconic structures of eukaryote biology.

RESULTS: We determined the 96-nm repeat structure of split *T. brucei* DMTs with attached ODAs, IDAs, RSs, and N-DRC using high-resolution cryo-EM. A total of 154 different axonemal proteins were identified, including 40 proteins unique to the trypanosome lineage. Trypanosome DMTs are further distinguished from those of other organisms by incorporation of extra paralogs for several microtubule inner proteins (MIPs), inner junction filament proteins, and proteins outside the DMT. We identified all subunits of the trypanosome-specific ponticulus and proteins on the DMT outer surface that may form attachments to the PFR. Using in-depth structural analysis combined with direct interrogation through knockdown of target proteins, we have defined MIP assembly mechanisms and unique features of the trypanosome axoneme.

DMTs were unexpectedly captured in a curved state, bending approximately 3° per 96-nm repeat with the B-tubule on the concave side, with dyneins in a pre–power stroke configuration. This contrasts with prior cryo–electron tomography reconstructions of the *T. brucei* axoneme and all high-resolution structures of split DMTs from other organisms, in which DMTs are straight with dyneins in the post–power stroke configuration. Compared with the post–power stroke state, motor domains of ODAs and IDAs shift toward the microtubule minus end in the pre–power stroke structure. Structural changes in the ODA linker are transmitted through the tail, which pivots upward and forward around the docking site on the DMT. This movement of the tail and associated intermediate chain–light chain tower accommodates concomitant repositioning of the motor domains from the adjacent dynein in the array. These findings led us to propose a "dragon boat" model for dynein-dependent movement of adjacent DMTs in the flagellar axoneme.

CONCLUSION: Our work expands our fundamental understanding of what it takes to build and operate a motile axoneme and also identifies parasite-specific adaptations that present potential targets for therapeutic or transmission-blocking agents. Our studies indicate that mechanisms for bending axonemal DMTs were established at or near the time of the last eukaryotic common ancestor, with eukaryotic diversification accompanied by lineage-specific axonemal proteins that alter DMT properties. These evolutionary adaptations enable each organism and cell type to produce a flagellar waveform that meets the needs of its specific functions and environments. ■



Cryo-EM structure of the decorated DMT from *T. brucei*. The phylogenetic tree of eukaryotes shows that the human parasite *T. brucei* diverged early within the eukaryotic lineage, before the emergence of separate lineages for humans and other organisms used as models for cilium biology. Magnified view of the *T. brucei* flagellum shows the cryo-EM structure of the *T. brucei* DMT with attached ODAs, IDAs, RSs, and N-DRC. Dyneins in the pre–power stroke state are docked to the curved DMT.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: kenthill@microbio.ucla.edu (K.L.H.); hong.zhou@ucla.edu (Z.H.Z.)
Cite this article as X. Xia *et al.*, *Science* **387**, eadr3314 (2025).
DOI: 10.1126/science.adr3314

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adr3314

RESEARCH ARTICLE SUMMARY

MICROBIOLOGY

Evolutionary adaptations of doublet microtubules in trypanosomatid parasites

Matthew H. Doran†, Qingwei Niu†, Jianwei Zeng, Tom Beneke, James Smith, Peter Ren, Sophia Fochler, Adrian Coscia, Johanna L. Höög, Shimi Meleppattu, Polina V. Lishko, Richard J. Wheeler, Eva Gluenz, Rui Zhang*, Alan Brown*

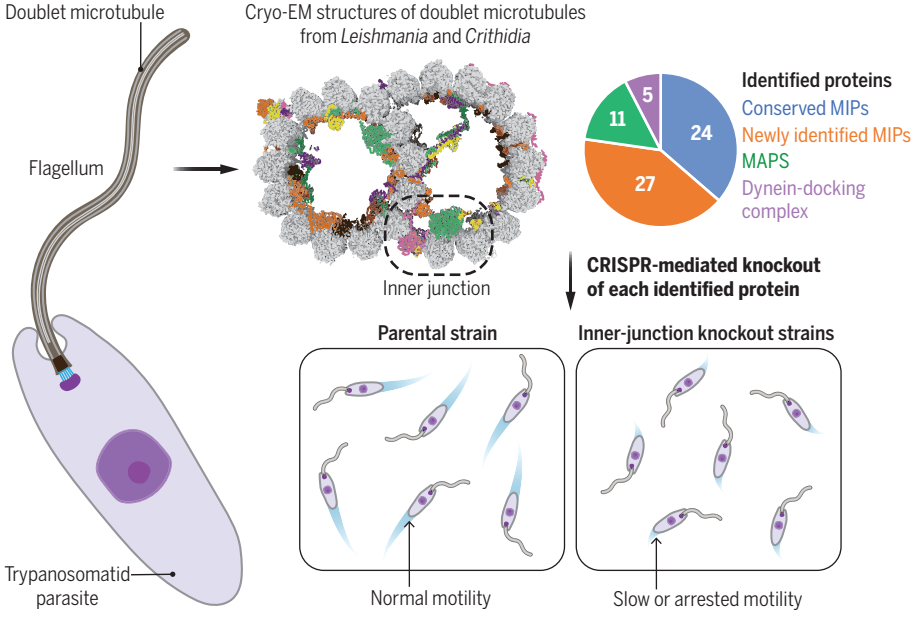
INTRODUCTION: Trypanosomatids are flagellated protists that are responsible for several devastating human diseases, including Chagas disease, trypanosomiasis, and leishmaniasis. These diseases put millions at risk of infection globally, especially those in the poorest communities of developing nations. Trypanosomatid parasites critically depend on a single motile flagellum for movement in either their vertebrate host or insect vector and for successful progression through their life cycles. The oscillatory pattern of the flagellar beat is powered by a protein complex within flagella called the axoneme, an extraordinarily complicated microtubule-based structure containing a collection of microtubule-binding proteins, dynein motors, and mechanoregulatory complexes. Despite its notable structural and molecular conservation across eukaryotes, axonemes display multiple species-specific differences.

RATIONALE: Structural descriptions of axonemal doublet microtubules have been reported for several species, revealing that they are sites of evolutionary adaptation capable of binding a diverse variety of microtubule inner proteins (MIPs) and microtubule-associated proteins (MAPs). However, the role of these molecular adaptations in motility is largely unknown. In this work, we sought to combine structural and functional analyses of trypanosomatid species to better understand the fundamental relationship between doublet microtubule structure and flagellar motility.

RESULTS: We used cryo–electron microscopy (cryo-EM) to resolve structures of doublet microtubules from two members of the trypanosomatid family, *Leishmania tarentolae* and *Crithidia fasciculata*, at resolutions up to 2.7 Å. The structures uncovered a total of 51 different MIPs within a single 48-nm repeat unit, including 27 that are specific to trypanosomatids or the wider kinetoplastid family, as well as a divergent five-membered dynein-docking complex and a class of MAPs called ArcMAPs. Some of the identified proteins come from paralogous genes: The presence of these paralogs within a single repeat unit generates a more complex periodic pattern and MIP network than observed in previously determined doublet microtubule structures. We also found that paralogous proteins allow for the formation of proximal-distal asymmetries. For example, members of the trypanosomatid dynein-docking complex contain paralogs that segregate specifically to the proximal or distal regions of the flagella. Features of paralog incorporation and the formation of proximal-distal asymmetries are likely applicable to most eukaryotic flagella and cilia.

We next leveraged the genetic tractability and the easily recordable motility of trypanosomatids to systematically test the effect of each identified protein on parasite swimming. Using CRISPR-based gene editing, we engineered knockouts (deletions) of each identified protein in *Leishmania mexicana*. For each knockout mutant generated, we used dark-field videomicroscopy to measure swimming speed and directionality. Our data reveal a distinct sensitivity of the inner junction (the join between A and B tubules) to perturbation, because genetic deletion of most inner-junction MIPs caused significantly reduced swimming speeds. Using a structure-based bioinformatics approach, we show that this region is home to a core set of MIPs that are present in all organisms with motile flagella, highlighting the importance of this region for motility.

CONCLUSIONS: Our structural analysis has resolved the species-specific specializations of trypanosomatid doublet microtubules in molecular detail, thereby enhancing our understanding how axonemal microtubules become patterned and diversified by MIPs. By systematically evaluating the contribution of each structure-verified protein to flagellar motility, we demonstrate the singular importance of the inner junction to proper motility. This integrative approach, combining structural analysis with comprehensive genetic knockout, promises to unveil the conserved and trypanosomatid-specific mechanisms that control flagellar beating. ■



Structural and functional analysis of trypanosomatid doublet microtubules. Cryo-EM analysis of doublet microtubules from the flagella of *L. tarentolae* and *C. fasciculata* identified 67 distinct proteins, highlighting the evolutionary adaptations of doublet microtubules that occur in flagellated parasites. A knockout screen determined the contribution of each protein to swimming behavior, revealing a distinct sensitivity of the inner junction to genetic perturbations.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: zhangrui@wustl.edu (R.Z.); alan_brown@hms.harvard.edu (A.B.)
†These authors contributed equally to this work.
Cite this article as M. H. Doran *et al.*, *Science* **387**, eadr5507 (2025). DOI: 10.1126/science.adr5507

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adr5507

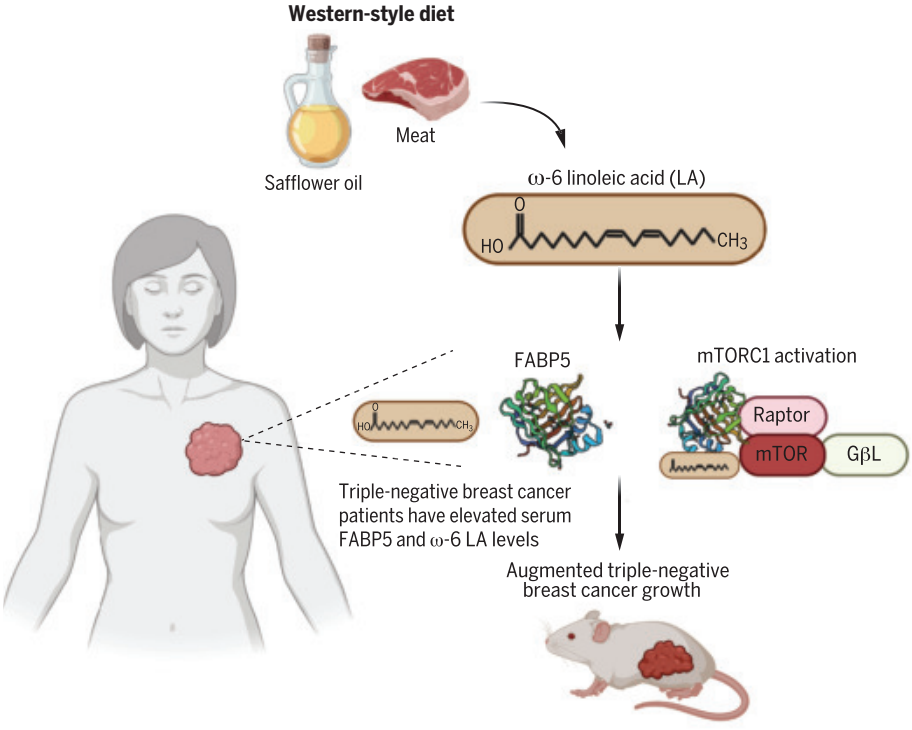
RESEARCH ARTICLE SUMMARY

SIGNAL TRANSDUCTION

Direct sensing of dietary ω-6 linoleic acid through FABP5-mTORC1 signaling

Nikos Koundouros*, Michal J. Nagiec, Nayah Bullen, Evan K. Noch, Guillermo Burgos-Barragan, Zhongchi Li, Long He, Sungyun Cho, Bobak Parang, Dominique Leone, Eleni Andreopoulou, John Blenis*

INTRODUCTION: Cancer cells must coordinate their growth and proliferation with nutrient availability and metabolic demand. This is achieved by the mechanistic target of rapamycin (mTOR) complex 1 (mTORC1) signaling pathway that integrates these extracellular cues with anabolic processes. Given the role of mTOR as a central regulator of cell metabolism, understanding how various nutrient inputs influence the signaling dynamics of this pathway could reveal strategies for integrating nutritional interventions in a cancer patient's journey. This is particularly relevant in breast cancer, where mTOR signaling is dysregulated in up to 50% of cases, and disease incidence is associated with modifiable factors, such as obesity. At the molecular level, the evidence linking amino acid and glucose availability to mTORC1 activity is pervasive, but the importance of other nutrients is understudied. These include the essential polyunsaturated fatty acids (PUFAs) ω-6 linoleic acid (LA) and ω-3 linolenic acid (ALA), which cannot be synthesized endogenously in humans and are only obtained from dietary sources. Physiologically, ω-6 PUFAs are predominantly associated with the synthesis of prostaglandins that mediate the inflammatory response, whereas ω-3 PUFAs have anti-inflammatory properties. Nevertheless, the mechanistic basis for how cancer cells respond to and use these dietary stimuli—and how they influence tumor growth and proliferation—remains poorly understood.



ω-6 LA fuels TNBC growth. A mechanism is revealed through which ω-6 LA—the most abundant polyunsaturated fat in Western diets—activates mTORC1, the main regulator of cell growth. FABP5 links these two components to form a nutrient-signaling axis that enhances the proliferation of TNBC cells in response to dietary ω-6 LA intake. GβL, G protein beta subunit–like; Raptor, regulatory-associated protein of mTOR. [Figure created with BioRender.com]

RATIONALE: ω-6 LA is the most abundant unsaturated fat in Western-style diets and is derived from animal products and processed foods containing vegetable oils, such as safflower oil. Many case-controlled retrospective and prospective studies have been conducted that explore associations between ω-6 LA intake and breast cancer incidence, but the conclusions are often contradictory. Adding to this complexity is breast cancer heterogeneity: Patients are stratified into four main clinical subtypes on the basis of expression of hormone receptors or lack thereof, each with distinct molecular characteristics and therapeutic sensitivities. Because ω-6 LA is an essential nutrient, we hypothesized that the mTOR pathway senses and is activated by its availability, leading to increased breast cancer cell proliferation in a subtype-specific manner.

RESULTS: By leveraging an extensive panel of breast cancer cell lines and patient-derived xenograft (PDX) tumors, we observed that ω-6 LA could activate mTORC1 but only in models of triple-negative breast cancer (TNBC), which is the most aggressive subtype that lacks any targeted therapy. We found that levels of the lipid chaperone fatty acid–binding protein 5 (FABP5) were significantly higher in TNBC compared with hormone receptor–positive tumors and that FABP5 directly interacted with mTORC1 to regulate complex formation, substrate binding, and subcellular localization. Notably, we demonstrated the relevance of this FABP5-mTORC1 signaling pathway in vivo by feeding animals a diet enriched for safflower oil that promoted TNBC tumor growth. FABP5 and ω-6 PUFAs appear to trigger a "perfect storm" of nutrient-driven signaling events, and both factors are also elevated in the serum of newly diagnosed TNBC patients.

CONCLUSION: Accumulating evidence suggests that dietary patterns may influence cancer outcomes, and there is substantial clinical interest in understanding the molecular mechanisms behind these associations to better inform nutritional recommendations. Our findings not only provide a mechanistic explanation for the heterogeneous responses of distinct breast cancer subtypes to dietary fats but also reveal an important perspective on how interactions between ω-6 LA intake and breast cancer need to be studied. Future nutritional studies might consider stratifying patients on the basis of FABP5 expression and triple-negative status. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: job2064@med.cornell.edu (J.B.); nik4009@med.cornell.edu (N.K.)
Cite this article as N. Koundouros *et al.*, *Science* **387**, eadm9805 (2025). DOI: 10.1126/science.adm9805

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adm9805

RESEARCH ARTICLE SUMMARY

IMMUNOLOGY

Identification of antigen-presenting cell–T cell interactions driving immune responses to food

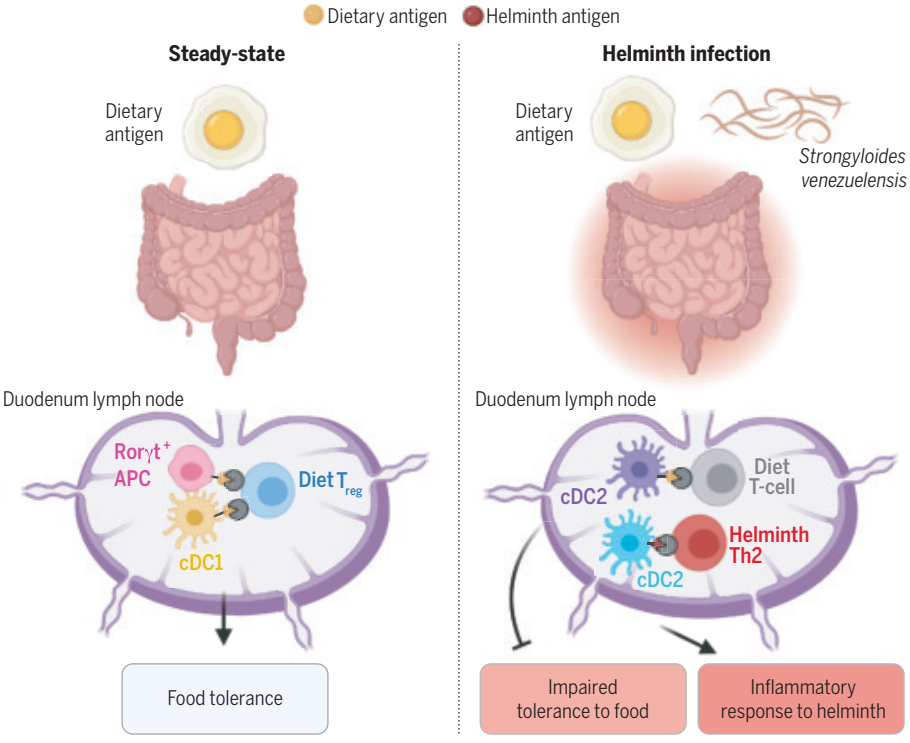
Maria Cecilia Campos Canesso*, Tiago Bruno Rezende de Castro†, Sandra Nakandakari-Higa†, Ainsley Lockhart, Julia Luehr, Juliana Bortolatto, Roham Parsa, Daria Esterházy, Mengze Lyu, Tian-Tian Liu, Kenneth M. Murphy, Gregory F. Sonnenberg, Bernardo S. Reis, Gabriel D. Victora*, Daniel Mucida*

INTRODUCTION: The intestinal immune system protects the organism against enteric pathogens while maintaining tolerance to dietary antigens and commensal bacteria. Failure of such an equilibrium may lead to food allergies, inflammatory bowel diseases, and increased susceptibility to infection. Antigen-presenting cells (APCs) are essential in this process because they present luminal antigens to naïve $CD4^+$ T cells, driving their differentiation into regulatory [peripheral regulatory T cell (pTreg)] or inflammatory [T helper (Th) cell] subsets—the drivers of immune tolerance and protection, respectively. However, it remains unclear how these antagonistic T cell outcomes can be induced independently in a manner that supports the simultaneous induction of tolerogenic and inflammatory reactions to the appropriate antigenic stimuli.

RATIONALE: Tolerance to food requires the generation of pTregs specific for dietary antigens; various APC subsets, including the conventional dendritic cell (cDC) subsets cDC1 and cDC2, as well as the emerging family of Rorγt$^+$ APCs, have been implicated in this process. Given the high density and diversity of luminal antigens constantly available in the gut, elucidating the mechanisms by which APCs induce the differentiation of pTreg or Th cells in vivo remains a long-standing challenge, whose solution requires the ability to identify, characterize, and locate the exact APCs engaged in each of these activities. To address this, we used LIPSTIC (Labeling Immune Partnerships by SorTagging Intercellular Contacts), a proximity labeling method that enables the identification of APCs actively presenting an antigen of interest in vivo. Using LIPSTIC, we sought to identify which APCs present dietary antigens under tolerizing conditions and to determine how enteric infections affect such presentation to disrupt oral tolerance.

RESULTS: By adapting LIPSTIC for use in the intestine, we were able to identify and characterize individual APCs involved in priming food-specific $CD4^+$ T cells in gut-draining lymph nodes (gLNs). Functional experiments and genetic approaches showed that at steady state, migratory cDC1s and Rorγt$^+$ APCs were effective at inducing pTreg differentiation. However, helminth infections, particularly *Strongyloides venezuelensis* (*S.v.*), led to a substantial decrease in the ratio of tolerogenic APCs (cDC1s and Rorγt$^+$ APCs) to inflammatory APCs, primarily cDC2s. This shift in APC populations led to reduced food-specific pTreg generation and disruption in oral tolerance induction. Interaction-based transcriptomics showed that although an inflammatory subset of cDC2s expanded during helminth infection, these DCs were not engaged in the presentation of dietary antigens and thus failed to induce differentiation of diet-specific Th2 cells.

CONCLUSION: Our data indicate a division of labor between different APC subsets in the presentation of luminal antigens. pTreg generation and the establishment of oral tolerance were both associated with dietary antigen presentation by cDC1s and Rorγt$^+$ APCs. Helminth infections impaired oral tolerance induction by excluding these tolerogenic APCs from antigen presentation to food-specific $CD4^+$ T cells, while inducing a large population of inflammatory cDC2s that did not engage in dietary antigen presentation. These findings reveal a compartmentalized system in which presentation of dietary and pathogen-derived antigens is accomplished by distinct APC subsets, likely as a mechanism to prevent food-specific Th2 responses even amid strong type 2 immunity driven by helminth infection. ■



Compartmentalized presentation of dietary versus pathogen-derived antigens by distinct APCs. APCs instruct T cells to differentiate into regulatory (pTreg) and effector (Th) subsets. Using LIPSTIC, we found that at steady state, cDC1s and Rorγt$^+$ APCs both presented dietary antigens, resulting in the induction of pTregs and oral tolerance. *Strongyloides venezuelensis* infection abrogated dietary antigen presentation by these APCs, impairing oral tolerance. Inflammatory helminth Th2-inducing cDC2s did not present dietary antigens, avoiding food-specific Th2 responses. [Figure created with BioRender.com]

The list of author affiliations is available in the full article online.
*Corresponding author. Email: mcamposcan@rockefeller.edu (M.C.C.C.); victora@rockefeller.edu (G.D.V.); mucida@rockefeller.edu (D.M.)
†These authors contributed equally to this work.
Cite this article as M. C. Campos Canesso *et al.*, *Science* **387**, eado5088 (2025). DOI: 10.1126/science.ado5088

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.ado5088

RESEARCH ARTICLE SUMMARY

HEART DEVELOPMENT

Tunneling nanotube–like structures regulate distant cellular interactions during heart formation

Lianjie Miao, Yangyang Lu, Anika Nusrat, Guizhen Fan, Shaohua Zhang, Luqi Zhao, Chia-Ling Wu, Hongyan Guo, Trang Le Nu Huyen, Yi Zheng, Zhen-Chuan Fan, Weinian Shou, Robert J. Schwartz, Yu Liu, Ashok Kumar, Haixin Sui, Irina I. Serysheva, Alan R. Burns, Leo Q. Wan, Bin Zhou, Sylvia M. Evans, Mingfu Wu*

INTRODUCTION: Heart development is a highly orchestrated process dependent on dynamic interactions between the myocardium and the endocardium. The two layers are separated by a noncellular matrix called cardiac jelly and communicate through signaling pathways involving membrane-bound receptors and ligands. However, the mechanisms enabling such signaling interaction over physical distances remain poorly understood. In this work, we characterized tunneling nanotube–like structures (TNTLs), which we found physically connecting cardiomyocytes (CMs) in the myocardium to endocardial cells (ECs) in the endocardium. These structures likely help to facilitate long-distance intercellular communication essential for heart formation.

RATIONALE: Heart formation relies on precise signaling interactions between the myocardium and endocardium, particularly during trabecular development. Signaling pathways, such as Notch1, Vegf, and Nrg1, have ligands and receptors segregated across these two cardiac layers. The mechanisms enabling these long-distance interactions across the intervening cardiac jelly are unclear. We hypothesized that TNTL structures exist between the cardiac layers and could mediate intercellular long-distance communication in the developing heart, allowing for the transport of signaling molecules and cytoplasmic proteins between them.

RESULTS: We used genetic labeling, contact-tracing techniques, and advanced imaging to demonstrate the existence of TNTLs in mouse embryonic hearts. These TNTLs extended from CMs to ECs across the cardiac jelly, establishing direct connections that enable signal transduction and cytoplasmic protein transfer.

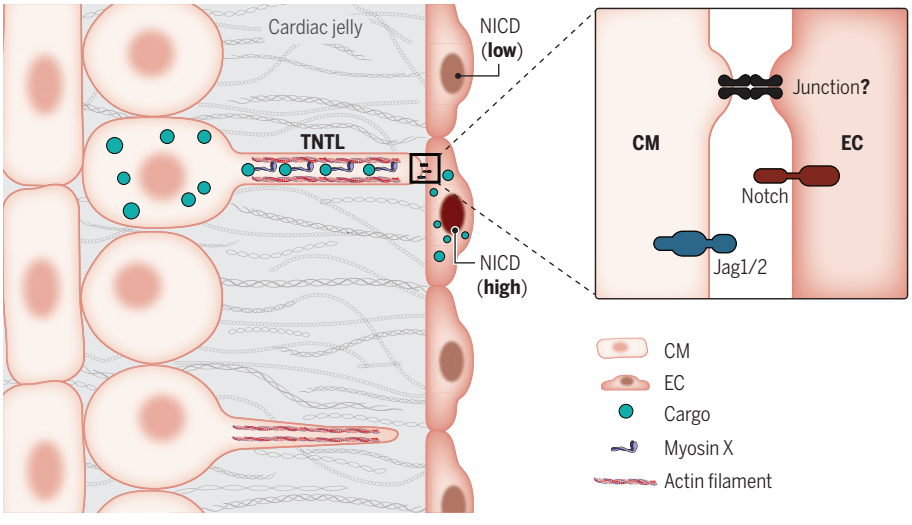
The TNTLs were identified in the heart through the genetic labeling of cellular protrusions. During mouse development TNTLs were shown to form between CMs and ECs as early as embryonic day 8.0. The filamentous structures inside TNTL were characterized by three-dimensional imaging and the reconstruction of the electron microscopy (EM) and cryo-EM images.

The TNTLs contained actin filaments, and TNTL formation depended on actin polymerization. The presence of actin filaments in TNTLs was confirmed in a transgenic mouse line that could label actin filaments with a fluorescent marker. Inhibiting actin polymerization chemically or by ablating the small guanosine triphosphatase, Cdc42, eliminated TNTLs.

The TNTLs were involved in regulating Notch1 signaling and other signaling pathways. TNTLs were sufficient to activate Notch1 signaling in ECs, with ligands from CMs transported through these microstructures to ECs. Loss of TNTLs resulted in reduced Notch signaling and other signaling pathways. TNTLs were able to transport signaling molecules, cytoplasmic proteins, and trafficking vesicles, underscoring their role as conduits for intercellular communication.

The TNTLs were essential in cardiac morphogenesis. Disruption of TNTLs in embryonic hearts resulted in impaired ventricular wall morphogenesis, evidenced by loss of trabeculae and defective myocardial growth.

CONCLUSION: In this work, we identified TNTLs as a critical mechanism for long-distance intercellular communication during heart development. These actin-rich structures physically bridge the myocardium and endocardium, allowing for the efficient exchange of signaling molecules necessary for trabecular formation and ventricular wall morphogenesis. Disruption of TNTLs compromises these interactions, highlighting their essential role in heart patterning. This work provides insights into mechanisms of cellular communication and suggests that TNTL formation might help cells to regulate long-distance cell-cell communication and modulate tissue patterning in mammalian systems. Future research should explore the molecular machinery governing TNTL formation, the structural basis of the interface between TNTLs and endocardial cells (see summary figure), the molecular mechanism of protein trafficking inside the TNTL, the potential functions of cytoplasmic protein transfer between myocardium and endocardium during heart formation, and the potential therapeutic implications of modulating these structures in congenital heart defects and heart failure. ■



Microstructure regulates signaling interaction and cytoplasmic proteins transfer between CMs and ECs. TNTLs linking CMs to ECs function as a bridge across the cardiac jelly for signaling molecule trafficking, cytoplasmic protein transferring, and Myosin X shuttling. Trafficking vesicles and cytoplasmic proteins but not Myosin X can be transferred from CMs to ECs. Disruption of TNTL affects Notch1 signaling activation and several other signaling interaction between CMs and ECs, leading to trabeculation defects. NICD, Notch1 intracellular domain.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: mwu25@central.uh.edu
Cite this article as L. Miao *et al.*, *Science* 387, eadd3417 (2025). DOI: 10.1126/science.add3417

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.add3417

RESEARCH ARTICLES

ORGANIC CHEMISTRY

Excited-state configuration of nitroarenes enables oxidative cleavage of aromatics over alkenes

Wesley J. Olivier[1]†, Piotr Błyszczyk[1]†, Enrique M. Arpa[1], Kenshiro Hitoshio[1], Miguel Gomez-Mendoza[2], Víctor de la Peña O'Shea[2], Isabelle Marchand[3], Thomas Poisson[3], Alessandro Ruffoni[1,4]*, Daniele Leonori[1]*

The ozonolytic deconstruction of aromatics remains a challenge in organic chemistry. Ozone preferentially reacts with alkenes over arenes, meaning that once the initial aromatic cleavage occurs, the dearomatized products are inherently more reactive than the starting materials. Consequently, the process cannot be halted, resulting in full oxidation. Addressing this challenge requires subverting intrinsic rules of chemoselectivity to transform a less reactive substrate in the presence of a more reactive one. We demonstrate that this concept can be achieved by using photoexcited nitroarenes. Crucial to the success of this chemoselective process is the use of a nitroarene that is preferentially excitable to a triplet π,π^* state over the n,π^* state. This switch in excited-state configuration provides an otherwise inaccessible manifold, in which oxidative cleavage is diverted toward aromatics in the presence of alkenes.

The expeditious construction of organic molecules is reliant on our ability to control chemical reactions. Chemoselectivity, in contrast to other forms of selectivity (e.g., stereoselectivity, regioselectivity), is attained when a reactant preferentially targets one functional group over others. High levels of chemoselectivity are generally achieved by exploiting electronic and/or steric effects along a reaction path, as well as other molecular parameters like pK_a and redox potentials (*1*–*3*). However, a challenge arises when intrinsic reactivity patterns need to be subverted to selectively target a less reactive functionality in the presence of more reactive ones (*4*). In these cases, multistep approaches based on protection or deprotection (and/or oxidation and reduction) sequences are often the only viable option (*1*, *5*). We demonstrate that switching of the electronic configuration of an excited-state reactant can be used to control the chemoselectivity of a chemical reaction. Specifically, this blueprint has been used to divert the reactivity of photoexcited nitroarenes away from alkenes, to enable chemoselective "ozonolysis-like" deconstruction of less reactive aromatics.

Aromatic hydrocarbons are produced on a ton scale from petroleum feedstock and then converted into high-value materials (*6*). Most chemical transformations applied to aromatics are based on the introduction or the manipulation of functionalities on their periphery (*7*). Though diverse in terms of mechanistic and synthetic features, these reactions leave the aromatic unit intact. A desirable but overlooked strategy would be to disrupt the aromatic system while introducing functionalities in a selective and predictable manner. Some dearomative methodologies are established (e.g., hydrogenation), but they usually keep the cyclic skeleton intact (*8*–*13*). The possibility of disrupting the aromaticity while also breaking the cyclic framework is much less developed but could open paths to high-value materials with diverse and orthogonal reactivity for further elaboration (*14*–*16*).

Despite attractive synthetic potential, the development of any dearomative strategy is limited by two main factors: (i) The high resonance stabilization that characterizes aromatic molecules makes them comparatively unreactive (*17*) and (ii) consequently, any chemical reagent that overcomes this inertness is inherently more prone to react with the products as they do not possess the initial stabilization energy (*18*). This often leads to unwanted and uncontrolled side or overreactions that limit synthetic potential and applicability. A classic example is arene ozonolysis, in which the intermediate unsaturated products are several orders of magnitude more reactive than the starting aromatic molecule. This generally leads to complete overoxidation whereby an aromatic group becomes a synthon for a carboxylic acid functionality and many of the initial carbon atoms are transformed into synthetically unusable by-products (Fig. 1) (*19*, *20*).

We have recently developed a strategy whereby nitroarenes, upon simple visible-light irradiation, can be used as ozone surrogates for the oxidative cleavage of alkenes (*21*). Photoexcitation of these materials populates the lowest-lying triplet state, which has n,π^* character and engages in a step-wise [3+2]-cycloaddition with πC=C bonds (*21*–*23*). In contrast to standard ozonolysis intermediates which pose explosion risks, "*N*-doped" ozonides (1,3,2-dioxazolidines) are stable chemicals that can be accumulated,

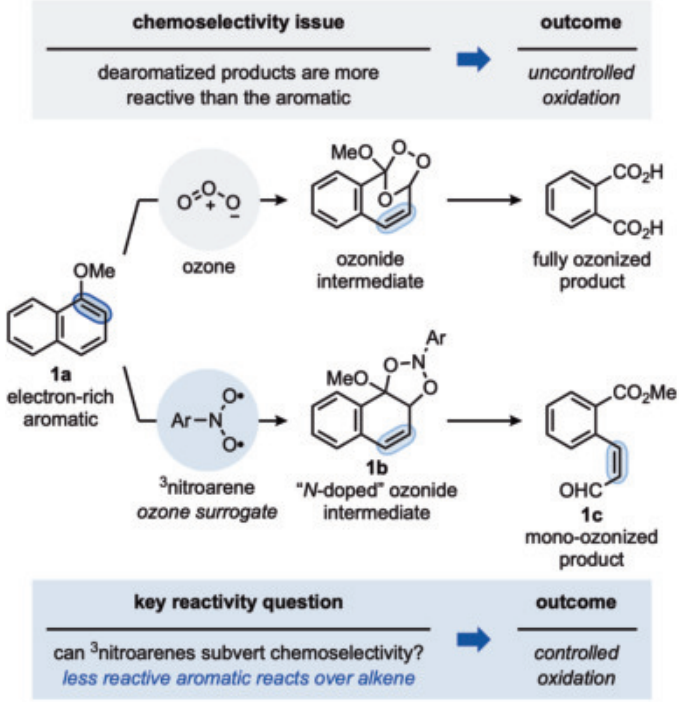


Fig. 1. A chemoselectivity challenge in ozonolysis. The ozonolysis of aromatics is difficult to control because the initial dearomative cleavage generates a more reactive substrate. This work explores the use of triplet nitroarenes to address this chemoselectivity challenge through a "*N*-doped"-ozonide.

[1]Institute of Organic Chemistry, RWTH Aachen University, Landoltweg 1, Aachen, Germany. [2]Photoactivated Processes Unit, IMDEA Energy Institute, Technological Park of Mostoles, Avda. Ramón de la Sagra 3, Madrid, Spain. [3]INSA Rouen Normandie, Univ. Rouen Normandie, CNRS, Normandie Univ., Institut CARMeN UMR 6064, Rouen, France. [4]Otto-Diels-Institut für Organische Chemie, Christian-Albrechts-Universität zu Kiel, Otto-Hahn-Platz 4, Kiel, Germany.
*Corresponding author. Email: aruffoni@oc.uni-kiel.de (A.R.); daniele.leonori@rwth-aachen.de (D.L.)
†These authors contributed equally to this work.

stored, and converted into carbonyl products by heating under hydrolytic conditions (*21*). A salient conclusion of our study was that rational modifications of the nitroarene substitution pattern enable modulation and, ultimately, control of their reactivity. We therefore became interested in understanding whether appropriate tuning could provide a solution to the outstanding challenge of aromatic ozonolysis. Considering 1-OMe-naphthalene **1a** as the model electron-rich aromatic, we envisaged that a photochemical [3+2]-cycloaddition might provide access to "*N*-doped" ozonide **1b** for C–C cleavage to product **1c** (Fig. 1). We expected that the ability of photoexcited nitroarenes to discriminate between electron-rich versus electron-poor olefins would effectively insulate **1c** from additional oxidation. However, it was difficult to anticipate which key features of the nitroarene would have prevented further reactivity on electron-rich olefin intermediate **1b**. Nonetheless, we hoped that nitroarene screening would ultimately provide a solution to this challenge. Pioneering work from Saito hinted that photoexcited nitroarenes could be employed to achieve oxidative arene cleavage (*24*–*26*). Indeed, they reported that Hg-lamp irradiation of benzene solutions of electron donor-acceptor (EDA) complexes, formed between an electron-poor nitroarene and a small number of highly electron-rich arenes, led to oxidative cleavage. However, the yields for this method were generally low and afforded mixtures containing the fully oxidized aromatic, akin to ozonolysis, thus not addressing the issue of selective aromatic ozonolysis over alkenes (see supplementary materials section 5.2).

We started our investigation by evaluating the reactivity of several nitroarenes (**N1** to **N10**) with **1a** (Fig. 2, A and B). The solvent (CH_2Cl_2) and temperature (−65°C) for this screening were selected following full reaction optimization (see supplementary materials section 6 for details). According to our previous work, placing electron-withdrawing substituents on the nitroarene amplifies its excited-state electrophilic character and ensures efficient oxidative cleavage of alkenes (*21*). However, this paradigm did not translate to the reaction with **1a** and only starting material was recovered using highly electron-poor **N1**. Notably, decreasing the electrophilic character of the nitroarene started to produce viable results for oxidative cleavage, leading to the selective formation of *ortho*-substituted benzoate **1c**. Depending on the nitroarene employed (**N2** to **N10**), this reaction occurred in moderate-to-good yields, had high mass recovery, and resulted in no other detectable by-products.

This unexpected and counterintuitive finding motivated us to test the intrinsic reactivity of **N1** to **N10** on cyclooctene **1d**. As expected, decreasing the nitroarene electrophilic character diminished the cleavage efficiency, which clearly contrasts with the outcome obtained on **1a**. Furthermore, we failed in our attempts at correlating these two reactivity profiles (aromatic versus alkene cleavage) with the electronic properties of the nitroarenes [e.g., Hammett σ coefficients (*27*), see also supplementary materials section 4.4.3] (Fig. 2C). This was particularly evident in the unexpected and completely orthogonal reactivity of **N1** and **N8** that seemed able to discriminate between aromatic versus alkene and vice versa.

As it was difficult to rationalize these two chemoselective outcomes purely on the basis of ground-state parameters, we considered the nature of the excited states of the nitroarenes, specifically by calculating the triplet energies and the excited-state reduction potentials for **N1** to **N10** (see tables S4 and S3, respectively). Notably, these parameters also did not appear to correlate with the efficiency of aromatic versus alkene cleavage. Furthermore, **N8** was found to possess the lowest triplet energy (60.5 kcal mol^{-1}) and excited-state reduction potential (E^*_{red} +1.31 V versus SCE in CH_2Cl_2) of all the nitroarenes evaluated, meaning that an adequate explanation for its high reactivity with **1a** remained elusive.

These findings lead us to speculate that the configuration of the lowest-lying triplet excited state might in fact be the key aspect controlling alkene versus aromatic cleavage. As discussed above, **N1** populates, after intersystem crossing (ISC), a short-lived (τ ~ ps) triplet state with n,π* character (*28*) in which the excited state is localized on the nitro group and can be represented as a di-O-radical species. Our calculations indicate that its π,π* triplet state is less stable, lying 16.6 kcal mol^{-1} above the n,π* state (Fig. 2D, left) (*29*–*31*). By contrast, **N8** populates a π,π* triplet state, which is longer lived (τ ~ 1.9 μs, see below), and features delocalization of the excited state over the entire π-system (Fig. 2D, right) (*32*). Crucially, its n,π* triplet state is 5.6 kcal mol^{-1} higher in energy.

The difference of up to six orders of magnitude in the **N8** ^{3}π,π* and **N1** ^{3}n,π* excited-state lifetimes can be understood in terms of the accessibility of the minimum-energy crossing points (CX) that connect these states with the ground state (S_0), allowing nonradiative deactivation (*33*) (fig. S13). For **N8**, we found that the ^{3}π,π*/S_0 CX is characterized by the out-of-plane motion of the NO_2 group (fig. S13). This major change in geometry raises the energy of this CX which is separated from the ^{3}π,π* minimum by > 20 kcal mol^{-1}. Therefore, the most plausible relaxation pathway for the **N8** ^{3}π,π* state involves back-internal conversion to the ^{3}n,π* state in a *contra*-thermodynamic sense (^{3}π,π* → ^{3}π,π*/^{3}n,π* → ^{3}n,π* → ^{3}n,π*/S_0; 11.5 kcal mol^{-1}, fig. S13). Conversely, the ^{3}n,π*/S_0 CX for **N1** involves no relevant geometry deformations, and lies much closer to the ^{3}n,π* minimum (~8 kcal mol^{-1}).

We then calculated the difference in energy between both triplet configurations [E(π,π*) − E(n,π*)] for all nitroarenes used and compared these values with the **1a** versus **1d** cleavage efficiency. Notably, a clear reactivity trend was now observed by ordering the nitroarenes on the basis of their excited-state triplet energy gaps. This outcome provides an experimental guideline whereby excited-state species featuring strong n,π* contributions lead to high reactivity with alkenes, whereas progressive lowering of the π,π* energy level diverts cleavage toward aromatics.

Perhaps the most notable implication of these results is that modulation of the population of nitroarene excited states might now enable chemoselective ozonolysis-like cleavage of aromatics in the presence of alkenes and vice versa. We evaluated this proposal by preparing a series of O-alkylated 1-naphthols containing terminal (**2a**) as well as increasingly more reactive *E*-disubstituted (**3a**), trisubstituted (**4a**), and styrenyl (**5a**) olefins. Using **N8** we obtained **2c** to **5c** as the sole products, with essentially identical yields and mass balances (50 to 56%) as the model substrate **1a**, consistent with full chemoselectivity. Likewise, **N1** enabled chemoselective alkene cleavages, affording **2e** to **5e** as the only products with excellent mass balances (>90%, see supplementary materials section 10 for full details). It is important to note that although targeting of the more reactive alkene in the presence of electron rich aromatic might be possible using O_3 under careful reaction monitoring, there are currently no synthetic methods able to selectively target the less reactive aromatic. This constitutes an example of "excited-state–controlled chemoselectivity" whereby the degree of delocalization of the excited-state over the nitroarene dictates selectivity, overriding standard reactivity patterns in oxidation chemistry. In stark comparison, the use of other nitroarenes with intermediate triplet energy gap values (e.g., **N3**), led to mixtures of products due to nonselective aromatic and alkene oxidation (see supplementary materials section 5.1).

Mechanistic studies

Having identified conditions to oxidatively cleave aromatics over alkenes, we became interested in elucidating specifics of the reaction mechanism. First, UV/Vis absorption spectroscopy studies confirmed that there is no EDA complex built up between **N1** or **N8** and either **1a** or **1d**. Notably, the use of a nitroarene reagent favoring formation of an EDA complex with **1a** (see supplementary materials section 4.3.1) led to complete aromatic decomposition with no traces of **1c**. This suggests that electron-poor nitroarenes capable of ground-state complexation and photoinduced electron transfer have a deleterious impact on the development of this reactivity. Upon purple light irradiation, **N8** populates

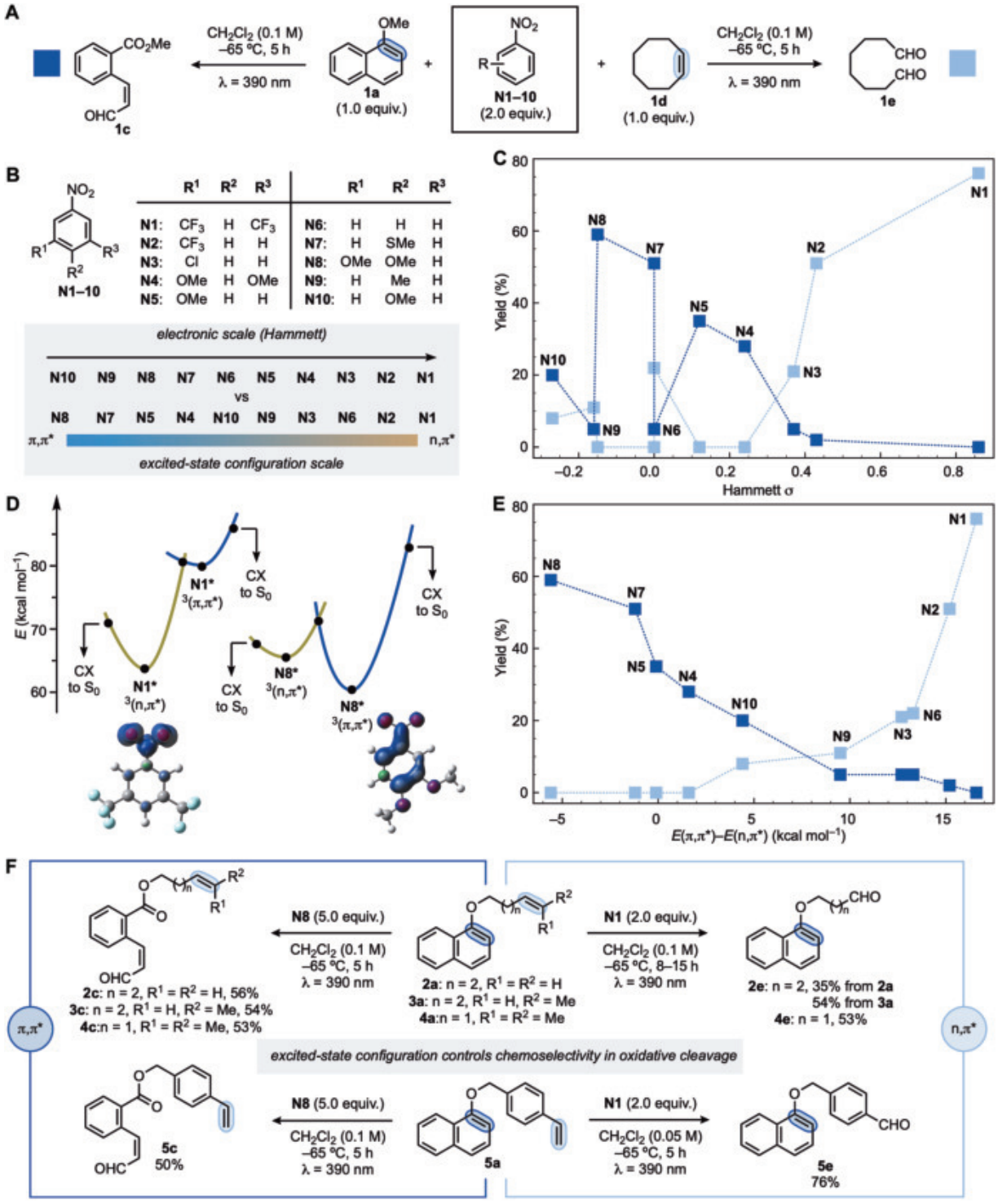


Fig. 2. Reaction optimization. (**A**) Oxidative cleavages of **1a** and **1d** using nitroarenes. (**B**) Nitroarenes tested and their ground- and excited-state properties. (**C**) Nitroarene performance in the cleavage of **1a** and **1d** on the basis of Hammett σ values. (**D**) Excited-state configuration of **N1** and **N8** in CH_2Cl_2. (**E**) Nitroarene performance in the cleavage of **1a** and **1d** on the basis of calculated $E(\pi,\pi^*)$–$E(n,\pi^*)$ values (kcal mol^{-1}). (**F**) Chemoselective oxidative cleavage of alkenes over aromatics and vice versa. Isolated yields are reported.

the π,π* triplet excited state (^{3}**N8**) that can be detected by transient absorption spectroscopy (laser-flash photolysis) (*32*, *34*, *35*). This species has a lifetime of τ = 1.9 μs under air-free conditions (CH_2Cl_2), and Stern-Volmer analysis with **1a** revealed a quenching constant k_q = $3.45 \pm 0.05 \times 10^9$ M^{-1} s^{-1} which means a quenching rate approaching diffusion (fig. S19). Importantly, the ability of ^{3}**N8** to discriminate between aromatics and alkenes was further corroborated by the fact that no excited-state quenching was observed in the presence of **1d** (fig. S20).

We computationally located a transient complex formed between ^{3}**N8** and **1a**, held together

by weak dispersion forces (see supplementary materials section 4.3). From this excited-state encounter complex, the formation of "*N*-doped"-ozonide **1b** can occur through two pathways depending on the initial site of C–O bond formation. Hence, reaction at C1 would lead to triplet biradical $\mathbf{1f_1}$ whereas addition to C2 provides isomeric species $\mathbf{1f_2}$. Either can then undergo ISC followed by cyclization to **1b**. Our computational studies demonstrated that both pathways ([^{3}**N8**•**1a**] → $\mathbf{1f_1}$ and [^{3}**N8**•**1a**] → $\mathbf{1f_2}$) are equally feasible ($\Delta E^{\ddagger}$ ~ 3 kcal mol^{-1}) and that the final transformations to **1b** are respectively barrierless.

Since ^{3}**N8** has π,π* configuration, we were interested in understanding how the initial C–O bond formation takes place as the NO_2 group does not have a di-O-radical character. According to our computational studies, photoexcitation leads to a polarization of the nitroarene and intramolecular charge transfer occurs from the dimethoxybenzene unit to the nitro group (*36*). Upon encountering **1a**, a transfer of electron density from the naphthalene ring to the aromatic ring moiety of ^{3}**N8** occurs at an early stage of the reaction coordinate, prior to the formation of the first C–O bond (at C–O = 2.65 Å). This process imparts **1a** with a partial positive charge and the nitro group of ^{3}**N8** with a partial negative charge thereby facilitating the formation of the first C–O bond.

We propose that the ability of ^{3}**N8** to chemoselectively react with aromatics (**1a**) and not alkenes (**1d**) is the result of several factors. First, since ^{3}**N8** does not have n,π* configuration, a radical-like [3+2]-cycloaddition cannot occur. Therefore, alkenes must react with ^{3}**N8** in a manner analogous to **1a**. However, they will provide substantially less stabilization by dispersion forces and the charge transfer will be energetically more costly, consistent with the generally higher oxidation potentials of alkenes [E^{ox}(**1d**) = +2.27 V versus SCE in CH_3CN] compared with electron-rich aromatics [E^{ox}(**1a**) = +1.27 V versus SCE in CH_3CN].

The overall implication of this reactivity profile is that ^{3}**N8** is intrinsically biased to address the key challenge in aromatic ozonolysis. Indeed, since it is primed to react with aromatics and not alkenes, once the "*N*-doped" ozonide **1b** is generated, there is no excited- or ground-state oxidant available for any further oxidative cleavage. This effectively suppresses any potential side reactions as generally observed using O_3.

A question remains regarding the chemoselective behavior of ^{3}**N8** versus ^{3}**N1**: specifically, why the π,π* configuration leads to aromatic cleavage but the n,π* does not. Both C1 and C2 of **1a** undergo a change in hybridization (sp^2 → sp^3), which leads to partial dearomatization of the naphthalene core. This process is energetically demanding and must occur regardless of the configuration of the triplet nitroarene. Crucially, since ^{3}**N8** possesses π,π* configuration, its excited state is delocalized over the entire (4n+2)π system, which means it is antiaromatic according to Baird's rules (*37*). Hence, reaction with **1a** should allow ^{3}**N8** to regain aromatic character and this stabilizing effect could compensate for the energetic demand of the dearomatization of **1a**. We tested this hypothesis by tracking the variation of the nucleus-independent chemical shift (NICS) values along the reaction coordinate. We found that in the same region where the charge transfer takes place, there is full relief of excited-state antiaromaticity for ^{3}**N8** whereas **1a** becomes nonaromatic (*38*–*40*). By contrast, the excited state of ^{3}**N1** is localized over the NO_2 group, which leaves the electron density of the phenyl ring unperturbed and preserves its Hückel aromaticity. Thus, the reaction between ^{3}**N1** and **1a** lacks the energy compensation provided by the rearomatization of the nitroarene. We also note that the substantially shorter lifetime of the n,π* triplet state may be an additional factor that contributes toward reaction inefficiency. However, no reactivity takes place after prolonged reaction times. Overall, we propose this reactivity mode and its chemoselectivity outcome to be dependent on the electronic nature of the nitroarene lowest-energy triplet state and ultimately driven by relief of excited-state antiaromaticity (*29*, *30*).

A further element of mechanistic investigation deals with the final conversion of **1b** into the *ortho*-functionalized benzoate **1c**. In our previous work on the oxidative cleavage of alkenes, we identified "*N*-doped" ozonides to be stable chemicals requiring thermal hydrolysis to generate the desired carbonyl products (*21*). It was therefore surprising that the reaction between **N8** and **1a** did not lead to the formation of **1b**, but directly provided **1c** without the need for H_2O and at a low temperature. We were unable to detect **1b** using low-temperature (−65°C) ^{1}H NMR spectroscopy, but supporting evidence for the formation of 1,3,2-dioxazolidines was obtained by reacting **N4** with **1a**. This species led to **1b′** which was stable in solution at −65°C and at room temperature for 1 to 2 hours, thus enabling characterization by ^{1}H and ^{13}C NMR spectroscopy, high-resolution mass spectrometry, and UV/VIS absorption spectroscopy (see supplementary materials section 4.4.6). Our calculations indicated that a thermal retro-[3+2] cycloaddition of **1b** involves a prohibitively high barrier of ~25 kcal mol^{-1} at −65°C (see supplementary materials section 4.5.3). However, thermal conversion of **1b′** to **1c** did occur slowly at room temperature in the absence of H_2O. Furthermore, the formation of **1c** was also observed under anhydrous conditions at −65°C with light irradiation (Fig. 3J). Hence, we propose that a different mechanism to the one observed for alkene cleavage might be operative for the final cycloreversion. Specifically, we considered the possibility that **1b** might evolve into **1c** through a photochemically initiated nitrene extrusion. Perhaps the most relevant readout for such a process in the absence of H_2O, is that both O-atoms in **1c** come from the nitroarene. Hence, we prepared bis-^{18}O-labeled **N8** which led to full ^{18}O incorporation at both the ester and the enal sides, strongly supporting nitrene extrusion as the mechanism. We tentatively propose that photoexcitation of **1b** might lead to a stepwise [3+2]-cycloreversion, giving **1c** and a nitrene byproduct (see supplementary materials section 4.5.3). Overall, although the oxidative cleavages of alkenes and aromatics might look related, the switch in triplet nitroarene excited-state configuration (n,π* versus π,π*) translates to fundamental differences in essentially every mechanistic aspect (see supplementary materials section 4.8).

Substrate scope exploration

With conditions in hand for the photochemical deconstruction of aromatics we investigated the scope of the process (Fig. 4). We started by modulating the nature of the O-substituent on the 1-naphthol core. The process was compatible with O*i*-Pr (**7a**) and OBn (**6a**) derivatives, thus demonstrating tolerance of steric hindrance and HAT-activated (H-atom transfer) positions. In terms of limitations, OAc (**8a**) and unprotected phenol (**9a**) derivatives could not be used. The presence of a C2-alkyl group was compatible giving *ortho*-enone **10c**.

C4-substituted 1-MeO-naphthalenes were screened next and we demonstrated compatibility with a broad range of substituents spanning OMe (**12a**), HAT-activated Me group (**13a**) (*41*), Ph (**14a**), all halogens (**15a** to **18a**), and, notably, electron-withdrawing CN (**19a**) and ester (**20a**). These β-functionalized enals derivatives underpin a class of largely unknown derivatives.

We then moved onto screening C8-functionalized derivatives that enabled the preparation of *ortho*,*ortho*-disubstituted benzoates (**21c** and **22c**), a class of synthetically challenging derivatives for both electrophilic aromatic substitution (S_EAr) and C–H activation. The chemistry was also applied on an extended phenanthrene derivative (**26a**) that gave the *ortho*,*ortho'*-biphenyl **26c**.

We then became interested in evaluating the reactivity of 2-OMe-naphthalene **25a** as oxidative cleavage could take place across C1 to C2 or C2 to C3. Deconstruction occurred selectively across C1 to C2 delivering, in good yield, **25c** which features the opposite disposition of ester and aldehyde functionalities with respect to **1c**. The ability of **25a** to interact with ^{3}**N8** was also studied by transient spectroscopy and this revealed the bimolecular quenching process to be ~1 order of magnitude slower than the one observed for **1a** (k_q(**25a**): 5.3 ± 0.1 × 10^8 versus k_q(**1a**): 3.45 ± 0.05 × 10^9 M^{-1} s^{-1}). This kinetic outcome suggests that site selectivity

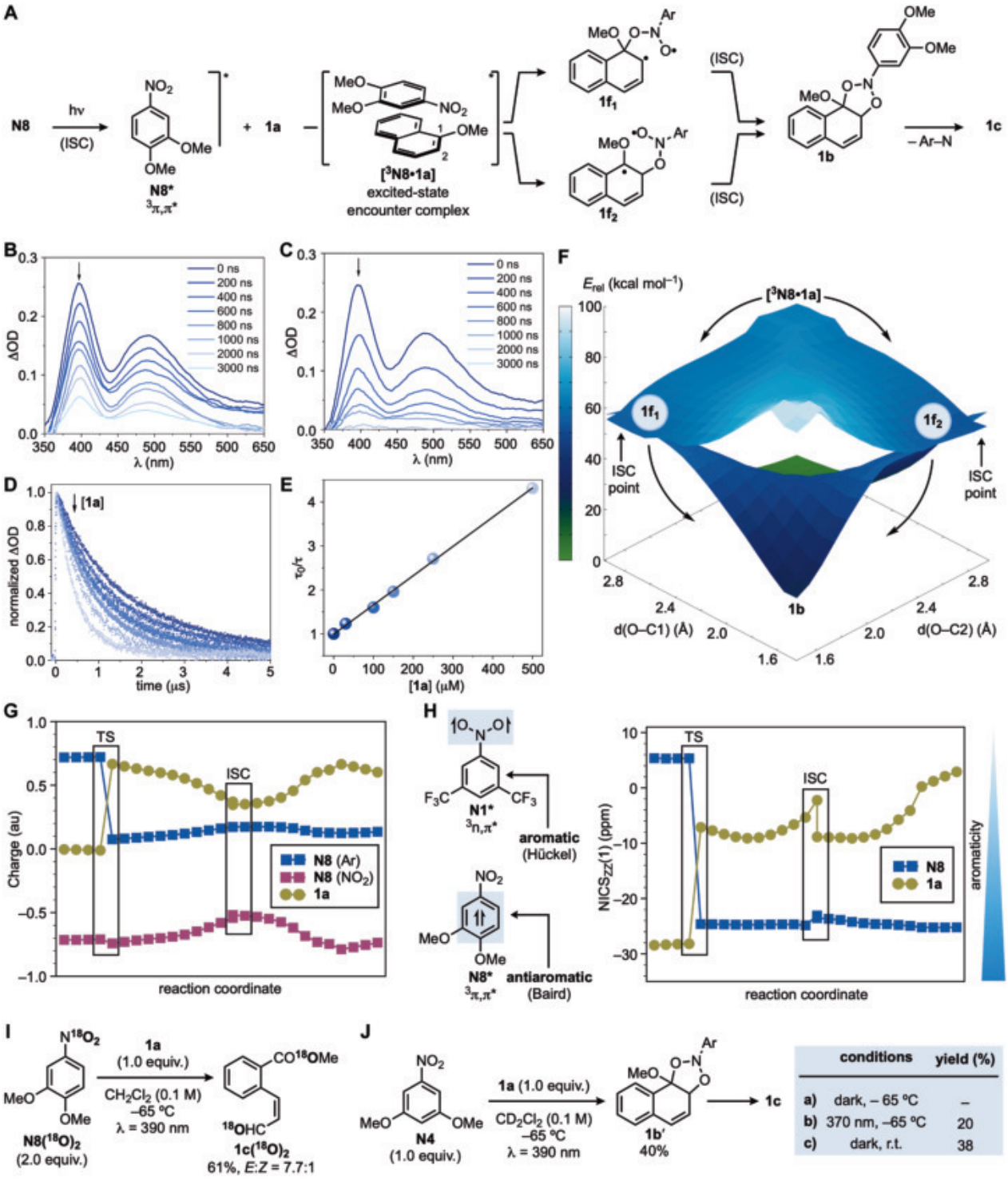


Fig. 3. Mechanistic studies. (**A**) Proposed reaction mechanism for the oxidative cleavage of **1a**; Ar = 3,4(OMe)$_2$-C$_6$H$_3$. (**B**) Transient absorption spectra for **N8** in deaerated CH_2Cl_2 (λ_{exc} = 355 or 390 nm). (**C**) Transient absorption spectra for **N8** in deaerated CH_2Cl_2 (λ_{exc} = 355 or 390 nm) upon addition of **1a** (500 μM). (**D**) Normalized transient decay traces for **N8** in deaerated CH_2Cl_2 (λ_{exc} = 355 or 390 nm) monitored at 395 or 490 nm upon addition of increasing amounts of **1a**. In all cases, the decays were fitted to a monoexponential function. (**E**) Corresponding Stern-Volmer plot. All goodness-of-fit (R^2) values exceed 0.983 (see fig. S19 for further information). The obtained K_q value was 3.45 ± 0.05 × 10^9 M^{-1} s^{-1}. (**F**) Representation of the M06-2X/cc-pVTZ/SMD(DCM)//M06-2X/cc-pVDZ ground (S_0) and lowest-lying triplet (T_1) states for the cycloaddition reaction between **1a** and **N8.** (**G**) Variation of the partial NBO charges of different molecular fragments along the [^{3}**N8•1a**] → **1b** reaction coordinate. (**H**) Variation of the NICS$_{zz}$(1) values for **N8** and **1a** along the reaction coordinate. (**I**) ^{18}O-Labeling experiments. (**J**) Formation and decomposition of 1,3,2-dioxazolidine **1b**′; Ar = 3,5(OMe)$_2$-C$_6$H$_3$. See supplementary materials section 4 for full experimental details.

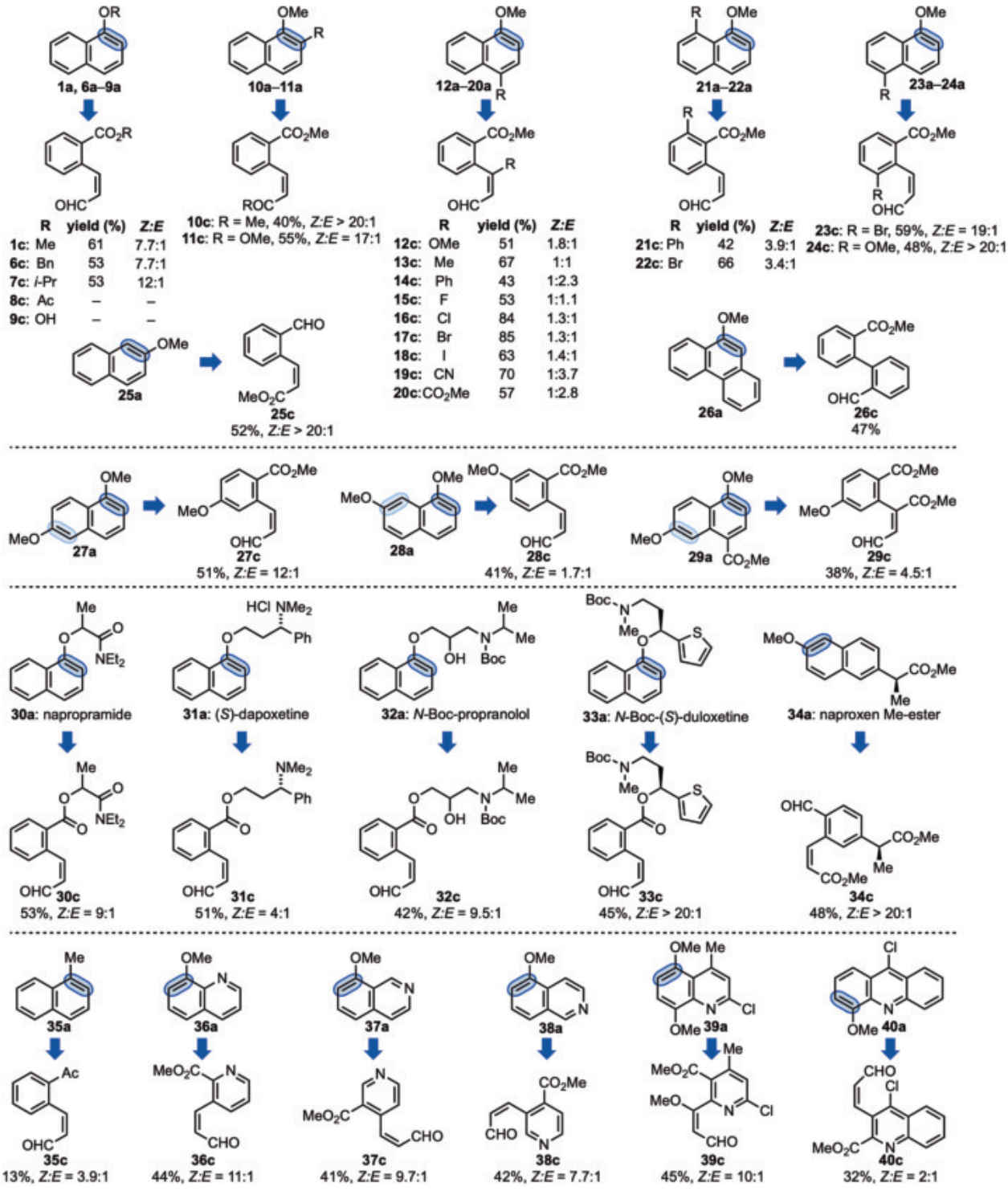


Fig. 4. Oxidative cleavage of naphthalenes and bicyclic azines. General reaction conditions: arene (1.0 equiv.), **N8** (5.0 equiv.), CH_2Cl_2 (0.1 M), −65°C, Kessil LEDs (λ = 390 nm). Isolated yields are reported. See supplementary materials sections 1 and 10 for full experimental details. The mass balance for the scope varies from good to excellent, see supplementary materials section 10 for more information.

might be achieved in the cleavage of naphthalenes containing two different OMe groups. This was evaluated using **27a** and **28a**. The selective formation of **27c** and **28c** demonstrates that cleavage across C1 to C2 occurs preferentially with respect to either C5 to C6 or C7 to C8. Furthermore, the strong preference for 1-MeO versus 2-MeO-naphthalene cleavage was also confirmed by the formation of **29a** in which C1 to C2 cleavage takes place in the presence of a potentially deactivating C4-ester functionality.

Since alkylated 1- and 2-naphthols are often encountered in the core structure of many commercial medicines, we decided to benchmark the chemistry in the oxidative deconstruction of complex and high-value materials. The blockbuster drugs napropramide (**30a**), (*S*)-dapoxetine (**31a**),

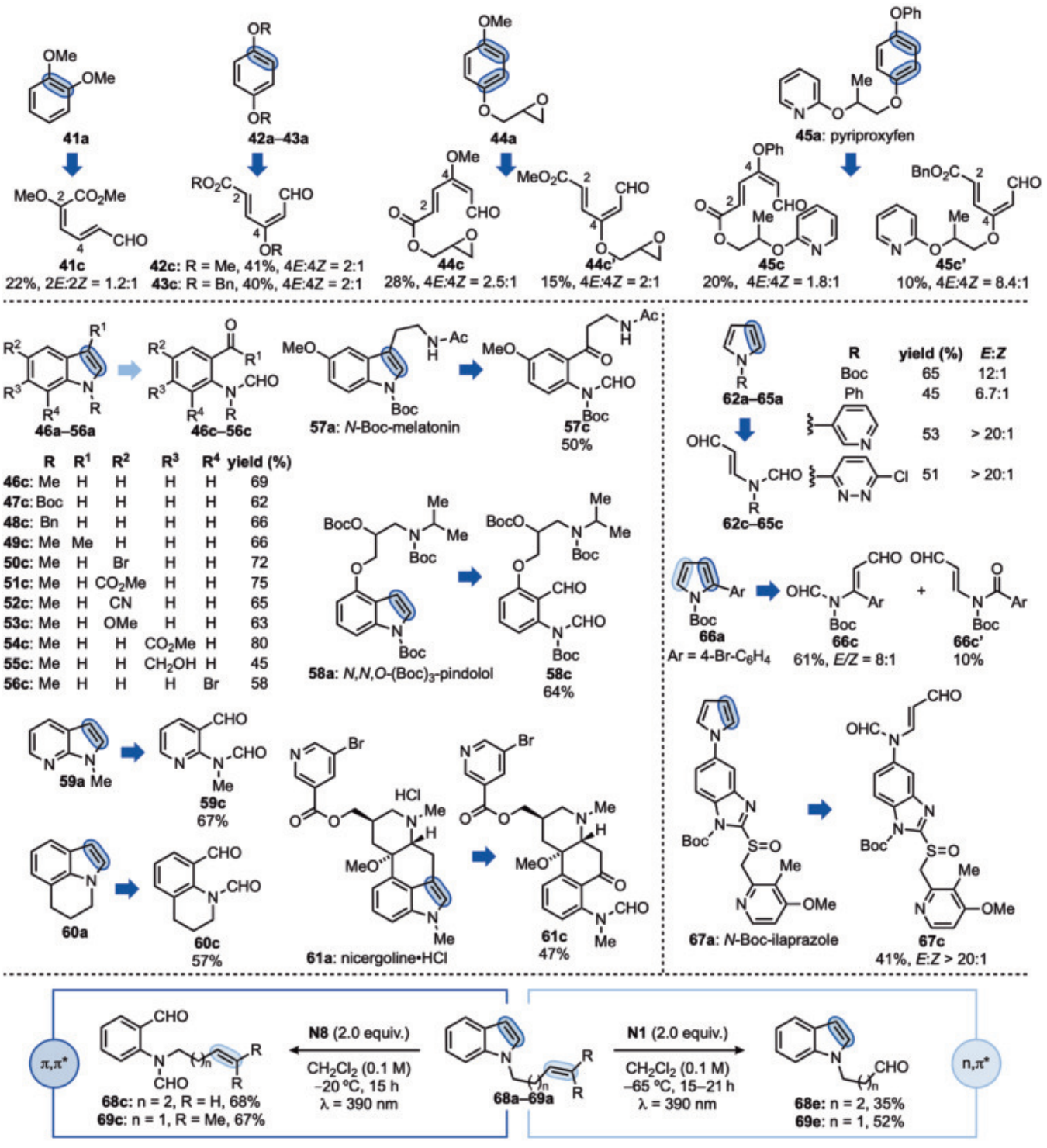


Fig. 5. Oxidative cleavage of anisoles, indoles, and pyrroles. General reaction conditions for anisoles and pyrroles: arene (1.0 equiv.), **N8** (4.0 to 5.0 equiv.), CH_2Cl_2 (0.01 M), −65°C, Kessil LED (λ = 390 nm). General reaction conditions for indoles: indole (1.0 equiv.), **N8** (5.0 equiv.), CH_2Cl_2 (0.1 M), −20°C, Kessil LED (λ = 390 nm). See supplementary materials sections 1 and 10 for full experimental details. The mass balance for the scope varies from good to excellent; see supplementary materials section 10 for more information.

and propanolol (**32a**, Boc-protected) feature amide, tertiary, and secondary amines as well as free alcohol functionalities, and gave **30c** to **32c** in good yields. Although the chemistry tolerated the presence of a free OH group, highly electron-rich tertiary amines required protonation, presumably to protect against oxidation by the nitroarene. Secondary amines were tolerated under the reaction conditions but also Boc-protected to aid with purification of the amino-aldehyde. Duloxetine (**33a**) contains a highly HAT-activated position (benzylic and α-O) as well as a thiophene ring that would not be compatible under O_3. This substrate, upon *N*-Boc protection, gave **33c** in moderate yield.

Finally, the 2-MeO-naphthalene–based naproxen methyl ester (**34a**) was deconstructed into **34c**, which can be further functionalized to explore additional chemical space around the aromatic core. Though naphthalene could not be engaged in the process, we successfully cleaved its 1-Me derivative **35a** albeit in low yield.

Quinoline and isoquinoline are too deactivated to react with ^{3}**N8** but the introduction MeO-groups in their benzenoid rings enabled efficient cleavage to ester- and enal-containing pyridines (**36c** to **38c**). The synthetic value of this approach was further demonstrated with the commercial quinoline **39a** and acridine **40a** that provided tetrasubstituted pyridine **39c** and trisubstituted quinoline **40c** in good to moderate yields. Accessing polyfunctionalized azines is often challenging, and this method introduces a straightforward retrosynthetic tactic for the expeditious preparation of these materials.

We then tested this approach in the cleavage of anisoles (fig. 5). We anticipated this to be much more challenging considering the higher stabilization energy of these derivatives (*17*). Nonetheless, although anisole could not be used, we successfully engaged 1,2- and 1,4-dimethoxybenzenes **41a** and **42a** in low and moderate yield, respectively. In this case, the cleavage products **41c** and **42c** represent a previously unknown class of push-pull dienes that might find application in Diels-Alder cycloaddition chemistry. The reactivity of **42a** with ^{3}**N8** was also investigated by transient spectroscopy that confirmed a considerably lower quenching performance ($k_q = 1.8 \times 10^8\ M^{-1}\ s^{-1}$), which is in line with the increasingly challenging nature of the process and the experimental outcome. We also evaluated other derivatives based on the 1,4-hydroquinone core. **43a** further highlighted the compatibility of the system toward HAT-labile positions even in the presence of a derivative, which is inherently more difficult to dearomatize. **44a** and the drug pyriproxyfen (**45a**) are two unsymmetrical substrates featuring epoxide functionality and 2-hydroxypyridine functionalities, respectively. In these two cases, oxidative deconstruction took place but non-selectively, delivering two isomeric dienes (**44c** and **44c′**; **45c** and **45c′**).

As an additional element of substrate scope, we attempted the deconstruction of electron-rich azoles. Indoles could be efficiently cleaved across C2 to C3 giving the corresponding *ortho*-formamide-containing benzaldehydes. Although this reactivity is synthetically related to that of singlet oxygen (*42*), our method enabled use of *N*-Me, -Bn, and -Boc derivatives (**46c** to **48c**) as well as substrates featuring substitutions at either C3, C5, C6, and C7 (**49a** to **56a**). The chemistry was also applied to an azaindole derivative (**59a**) which gave pyridine **59c** and the structurally complex melatonin (**57a**), pindolol (**58a**) (both Boc-protected), and nicergoline (**61a**). Notably, despite the indole core being electron rich, use of **N8** and **N1** also enabled chemoselective aromatic versus olefin cleavage on **68a** and **69a**.

Finally, we aimed to apply the method to the deconstruction of pyrroles, for which there are no general methods for cleavage across C2 to C3, and which often undergo uncontrolled decomposition under oxidative settings (*43*). The reactivity was extended to both *N*-Boc and *N*-aryl derivatives (**62a** to **65a**), including the blockbuster drug ilaprazole (**67a**). This further demonstrated tolerance toward benzimidazole and electron-rich pyridine rings as well as oxidable sulfoxide functionality. The unsymmetrical *N*-Boc-2-aryl derivative **66a** underwent cleavage with good selectivity across the less-substituted C4 to C5 side giving **66c** as the major product.

Conclusions

The ozonolysis of aromatics is a classical example of when formation of an intrinsically more reactive product leads to over-oxidation and therefore limited synthetic versatility. The results presented demonstrate that photoexcited nitroarenes can be used to overcome this reactivity challenge and cleave, in a controlled and selective manner, many electron-rich aromatics. We hope that this strategy will find application in the deconstruction of other high value molecules and will further stimulate the exploration of other reactivity patterns that can be controlled by specific excited-state configurations.

REFERENCES AND NOTES

1. N. A. Afagh, A. K. Yudin, *Angew. Chem. Int. Ed.* **49**, 262–310 (2010).
2. T. Newhouse, P. S. Baran, *Angew. Chem. Int. Ed.* **50**, 3362–3374 (2011).
3. I. P. Beletskaya, C. Nájera, M. Yus, *Chem. Soc. Rev.* **49**, 7101–7166 (2020).
4. T. Iwasaki, K. Nozaki, *Nat. Rev. Chem.* **8**, 518–534 (2024).
5. I. S. Young, P. S. Baran, *Nat. Chem.* **1**, 193–205 (2009).
6. H.-G. Franck, J. W. Stadelhofer, *Industrial Aromatic Chemistry: Raw Materials· Processes· Products* (Springer, 2012).
7. J. Mortier, Ed., *Arene chemistry: Reaction Mechanisms and Methods for Aromatic Compounds* (Wiley, 2015).
8. M. P. Wiesenfeldt, Z. Nairoukh, W. Li, F. Glorius, *Science* **357**, 908–912 (2017).
9. D. H. Liu, J. Ma, *Angew. Chem. Int. Ed.* **63**, e202402819 (2024).
10. Y.-Z. Cheng, Z. Feng, X. Zhang, S.-L. You, *Chem. Soc. Rev.* **51**, 2145–2170 (2022).
11. C. J. Huck, D. Sarlah, *Chem* **6**, 1589–1603 (2020).
12. R. Mykura *et al*., *Nat. Chem.* **16**, 771–779 (2024).
13. E. H. Southgate, J. Pospech, J. Fu, D. R. Holycross, D. Sarlah, *Nat. Chem.* **8**, 922–928 (2016).
14. X. Qiu *et al*., *Nature* **597**, 64–69 (2021).
15. T. D. H. Bugg, C. J. Winfield, *Nat. Prod. Rep.* **15**, 513–530 (1998).
16. K. N. Parida, J. N. Moorthy, *J. Org. Chem.* **80**, 8354–8360 (2015).
17. R. Gleiter, G. Haberhauer, *Aromaticity and Other Conjugation Effects* (Wiley, 2012).
18. W. B. Motherwell, A. S. Williams, *Angew. Chem. Int. Ed.* **34**, 2031–2033 (1995).
19. M. D. Gaul, G. A. Junk, H. J. Svec, *Environ. Sci. Technol.* **21**, 777–784 (1987).
20. B. Rindone, F. Saliu, R. Suarez-Bertoa, *Ozone Sci. Eng.* **32**, 161–165 (2010).
21. A. Ruffoni, C. Hampton, M. Simonetti, D. Leonori, *Nature* **610**, 81–86 (2022).
22. P. De Mayo, J. Charlton, C. Liao, *J. Am. Chem. Soc.* **93**, 2463–2471 (1971).
23. D. E. Wise *et al*., *J. Am. Chem. Soc.* **144**, 15437–15442 (2022).
24. I. Saito, M. Takami, T. Matsuura, *Bull. Chem. Soc. Jpn.* **48**, 2865–2871 (1975).
25. I. Saito, M. Takami, T. Matsuura, *Chem. Lett.* **1**, 1195–1196 (1972).
26. I. Saito, M. Takami, T. Matsuura, *Tetrahedron Lett.* **15**, 659–660 (1974).
27. C. Hansch, A. Leo, R. W. Taft, *Chem. Rev.* **91**, 165–195 (1991).
28. R. Hurley, A. Testa, *J. Am. Chem. Soc.* **90**, 1949–1952 (1968).
29. M. Takezaki, N. Hirota, M. Terazima, *J. Phys. Chem. A* **101**, 3443–3448 (1997).
30. M. Takezaki *et al*., *J. Phys. Chem. A* **101**, 5190–5195 (1997).
31. R. Hurley, A. Testa, *J. Am. Chem. Soc.* **88**, 4330–4332 (1966).
32. M. Mir, L. M. G. Jansen, F. Wilkinson, J. L. Bourdelande, J. Marquet, *J. Photochem. Photobiol. Chem.* **113**, 113–117 (1998).
33. A. Cu, A. Testa, *J. Phys. Chem.* **79**, 644–646 (1975).
34. R. Gonzalez-Blanco, J. L. Bourdelande, J. Marquet, *J. Org. Chem.* **62**, 6903–6910 (1997).
35. G. G. Wubbels, D. P. Susens, E. B. Coughlin, *J. Am. Chem. Soc.* **110**, 2538–2542 (1988).
36. J. Cerveló *et al*., *Tetrahedron Lett.* **25**, 4147–4150 (1984).
37. N. C. Baird, *J. Am. Chem. Soc.* **94**, 4941–4948 (1972).
38. M. Rosenberg, C. Dahlstrand, K. Kilså, H. Ottosson, *Chem. Rev.* **114**, 5379–5425 (2014).
39. J. Yan, T. Slanina, J. Bergman, H. Ottosson, *Chemistry (Basel)* **29**, e202203748 (2023).
40. M. Baranac-Stojanović, *J. Org. Chem.* **85**, 4289–4297 (2020).
41. S. Negele, K. Wieser, T. Severin, *J. Org. Chem.* **63**, 1138–1143 (1998).
42. C. Zhang *et al*., *ACS Catal.* **6**, 6853–6860 (2016).
43. J. K. Howard, K. J. Rihak, A. C. Bissember, J. A. Smith, *Chem. Asian J.* **11**, 155–167 (2016).

ACKNOWLEDGMENTS

Funding: D.L. acknowledges the European Research Council for a grant (101086901); W.J.O. and E.M.A. thank the Marie Skłodowska-Curie Actions for fellowships (101102820 and 101150311 respectively). The authors gratefully acknowledge the computing time provided to them at the NHR Center NHR4CES at RWTH Aachen University (project number p0021519). This is funded by the Federal Ministry of Education and Research, and the state governments participating on the basis of the resolutions of the GWK for national high-performance computing at universities (www.nhr-verein.de/unsere-partner). M.G.-M. and V.A.d.l.P.O'S. acknowledge financial support from the Spanish national project SolarChem 5.0 (TED2021-130173B-C41) and SOL-Future (PLEC2021-0079069). I.M. and T.P. acknowledge the API program supported by the European Union through the operational program (ERDF/FSE 2014–2020, no. 20E04976), the RIN ZIRCO_PHOTO_COP program (ERDF/FSE 2021–2027, no. 21E05645) and FNADT-DRACCARE [program 2020–2023 (N)15179] for funding, as well as Normandie Univ. (NU), Région Normandie, CNRS, Univ. Rouen Normandie, INSA Rouen Normandie, Labex SynOrg (ANR-11-LABX0029), the Graduate School for Research XL–Chem (ANR-18- EURE-0020 XL CHEM) and Innovation Chimie Carnot (I2C) for support. K.H. acknowledges the JSPS Overseas Challenge Program for Young Researchers. **Author contributions:** A.R. and D.L. conceptualized and designed the project. W.J.O., P.T.B., and K.H. designed and performed the synthetic experiments. E.M.A. designed and performed the computational studies. M.G.M. and V.d.l.P.O. designed and performed the laser flash photolysis experiments. I.M. and T.P. designed and performed optimization in flow. All authors analyzed the results, contributed to the mechanistic understanding of the process, and wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ads3955
Materials and Methods
Figs. S1 to S48
Tables S1 to S11
NMR Spectra
References (*44*–*91*)

Submitted 11 August 2024; resubmitted 13 November 2024
Accepted 11 February 2025
10.1126/science.ads3955

CHIRAL ELECTRONICS

Circularly polarized electroluminescence from chiral supramolecular semiconductor thin films

Rituparno Chowdhury[1]†, Marco D. Preuss[2]†, Hwan-Hee Cho[1]‡, Joshua J. P. Thompson[3], Samarpita Sen[4], Tomi K. Baikie[1], Pratyush Ghosh[1], Yorrick Boeije[1,5], Xian Wei Chua[1,5], Kai-Wei Chang[3], Erjuan Guo[6], Joost van der Tol[2], Bart W. L. van den Bersselaar[2], Andrea Taddeucci[7,8], Nicolas Daub[2], Daphne M. Dekker[9], Scott T. Keene[1], Ghislaine Vantomme[2], Bruno Ehrler[9], Stefan C. J. Meskers[2], Akshay Rao[1], Bartomeu Monserrat[3], E. W. Meijer[2]*, Richard H. Friend[1]*

Current organic light-emitting diode (OLED) technology uses light-emitting molecules in a molecular host. We report green circularly polarized luminescence (CPL) in a chirally ordered supramolecular assembly, with 24% dissymmetry in a triazatruxene (TAT) system. We found that TAT assembled into helices with a pitch of six molecules, associating angular momentum to the valence and conduction bands and obtaining the observed CPL. Cosublimation of TAT as the "guest" in a structurally mismatched "host" enabled fabrication of thin films in which chiral crystallization was achieved in situ by thermally triggered nanophase segregation of dopant and host while preserving film integrity. The OLEDs showed external quantum efficiencies of up to 16% and electroluminescence dissymmetries ≥10%. Vacuum deposition of chiral superstructures opens new opportunities to explore chiral-driven optical and transport phenomena.

Obtaining chiral nanostructures in semiconductor devices is inherently complex owing to the challenge of manipulating molecular arrangements with precise chirality at the nanoscale (*1*). Maintaining uniform chirality across large areas is a further challenge using established device fabrication methods such as vacuum sublimation, which have not been reported so far. The dynamic and noncovalent nature of supramolecular interactions enables the fabrication of chiral structures using a bottom-up approach (*1*–*4*) by transferring asymmetry from the molecular to the nanoscopic or microscopic scale (*5*). Such chiral materials have been used for advanced spintronic and optoelectronic applications (*6*). A commonly followed approach is to attach chiral side groups to functional semiconductor units, so that the chiral side groups cause chiral organization of the functional blocks. Both molecular systems, as described below, and also two-dimensional (2D) metal halide perovskites (*7*–*10*) have used this approach.

Since the first reports of a circularly polarized organic light-emitting diode (CP-OLED) (*11*–*13*), substantial efforts (*13*–*15*) have been made to improve the efficiency of the device and the degree of dissymmetry (g_{EL}) in circularly polarized electroluminescence (CP-EL). Strong dissymmetry in the EL, close to the theoretical maximum of $g_{EL} = \pm 2$, has been realized for materials forming chiral, cholesteric supramolecular aggregates in which the dissymmetry arises from the photonic structure (*13*). Current efficient OLED systems use light-emitting molecules spatially isolated in a host and arranged to harvest both triplet and singlet excitons. The emitter molecule is generally isolated in a host matrix and, although it may have chiral symmetry, it has little impact on the electronic structure and the levels of CPL are low and below 10^{-3} (*16*). However, systems with higher levels of dissymmetry—for example, systems using lanthanide emitters—show low electroluminescence quantum yield (ELQE) (*17*). This observed trade-off between device performance and EL dissymmetry limits the application of CP-OLEDs to date.

Here, we combined high dissymmetry in EL with high OLED-device performance. To this end, we exploited thermally induced phase segregation in vacuum-sublimed host-dopant films to fabricate chiral supramolecular nanostructures. We showed that delocalization of the exciton on the helically ordered supramolecular structure creates an orbital angular momentum that yields highly dissymmetric emission ($g_{EL} \sim 10^{-1}$) together with efficient (maximum external quantum efficiency $EQE_{max} \sim$ 16%) and bright (maximum luminscence $L_{max} > 5 \times 10^4$ cd m^{-2}) OLED devices.

Supramolecular assembly and structure

The molecules synthesized (Fig. 1A) share the triazatruxene (TAT) core with different N-substituents: *S*-3,7-dimethyloctyl (*S*-TAT), *n*-octyl (*n*-TAT), and phenyl (*phen*-TAT). Out of these molecules, only *S*-TAT possesses intrinsic chemical chirality. The remaining two molecules, *n*-TAT and *phen*-TAT, served as nonchiral and aromatic structural controls to study the emergent chiral supramolecular behavior of *S*-TAT, which is described later. Free-standing single crystals were obtained by slow evaporation from toluene-hexane solution. The crystal structure of *S*-TAT (Fig. 1B) exhibited a noncentrosymmetric packing with a primitive unit cell conforming to the chiral, $P6_122$ space group. When viewed along the *b* axis, the *S*-TAT appears dimerized and helically organized through cofacial π-stacking and exhibits a spacing of 0.38 nm in the dimer and 0.39 nm in between dimers along the *a* axis. Viewing along the *c* axis revealed the sixfold rotation axis of the stacking and face-to-face stacking without any slippage. From this structure, we determined the helical pitch of the supramolecular assembly to be 2.3 nm and an intermolecular rotation angle of 40° within the dimer and 120° in between dimers. The crystal structures of other TAT derivatives and *S*-TAT polymorphs are discussed in the supplementary materials (SM). We also fabricated vacuum-cosublimed thin films of TAT with molecular semiconductor hosts, including 4,4′-bis(*N*-carbazolyl)-1,1′-biphenyl (CBP). As presented later, in situ crystallization of the TAT within the host matrix was readily achieved, which was triggered by the shape dissimilarity between the *N*-substituent of TAT and the semiconducting molecular host, as illustrated schematically in Fig. 1C. These thin films were used as the emitting layers in LEDs, as discussed below.

Optical absorption measured using photothermal deflection spectroscopy (PDS) and photoluminescence (PL) spectroscopy results for both single crystals and films of *S*-TAT are shown in Fig. 1D. Pristine vacuum-sublimed, amorphous thin films of *S*-TAT showed that the isolated noninteracting molecule had a sharp absorption edge at 380 nm, displaying PL with vibronic features at 390 and 407 nm with a photoluminescence quantum efficiency (PLQE) of 14%. The *S*-TAT chiral crystal had its absorption edge red-shifted by ~0.65 eV to 510 nm and gave a PLQE of 58%. PLQE is maintained for the green emission in annealed vacuum-sublimed thin-films (table S1).

Pressure-dependent measurements of PL spectra were obtained on single crystals and thin films, up to 350 MPa (figs. S27 and S28). There was a pronounced reduction in PL intensity with increasing pressure (up to 16% at 350 MPa) and

[1]Cavendish Laboratory, University of Cambridge, Cambridge, UK. [2]Institute for Complex Molecular Systems and Laboratory of Macromolecular and Organic Chemistry, Eindhoven University of Technology, Eindhoven, Netherlands. [3]Department of Materials Science and Metallurgy, University of Cambridge, Cambridge, UK. [4]The Gurdon Institute and the Department of Genetics, University of Cambridge, Cambridge, UK. [5]Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge, UK. [6]State Key Laboratory of Materials Processing and Die and Mould Technology, School of Materials Science and Engineering, Huazhong University of Science and Technology, Wuhan, China. [7]B23 Beamline, Diamond Light Source Ltd, Didcot, UK. [8]Dipartimento di Chimica e Chimica Industriale, University of Pisa, Pisa, Italy. [9]MPV-SEM Department, AMOLF, Amsterdam, Netherlands.
*Corresponding author. Email: e.w.meijer@tue.nl (E.W.J.); rhf10@cam.ac.uk (R.H.F.)
‡These authors contributed equally to this work.
‡Present address: Department of Materials Science and Engineering, Yonsei University, Seoul, Republic of Korea.

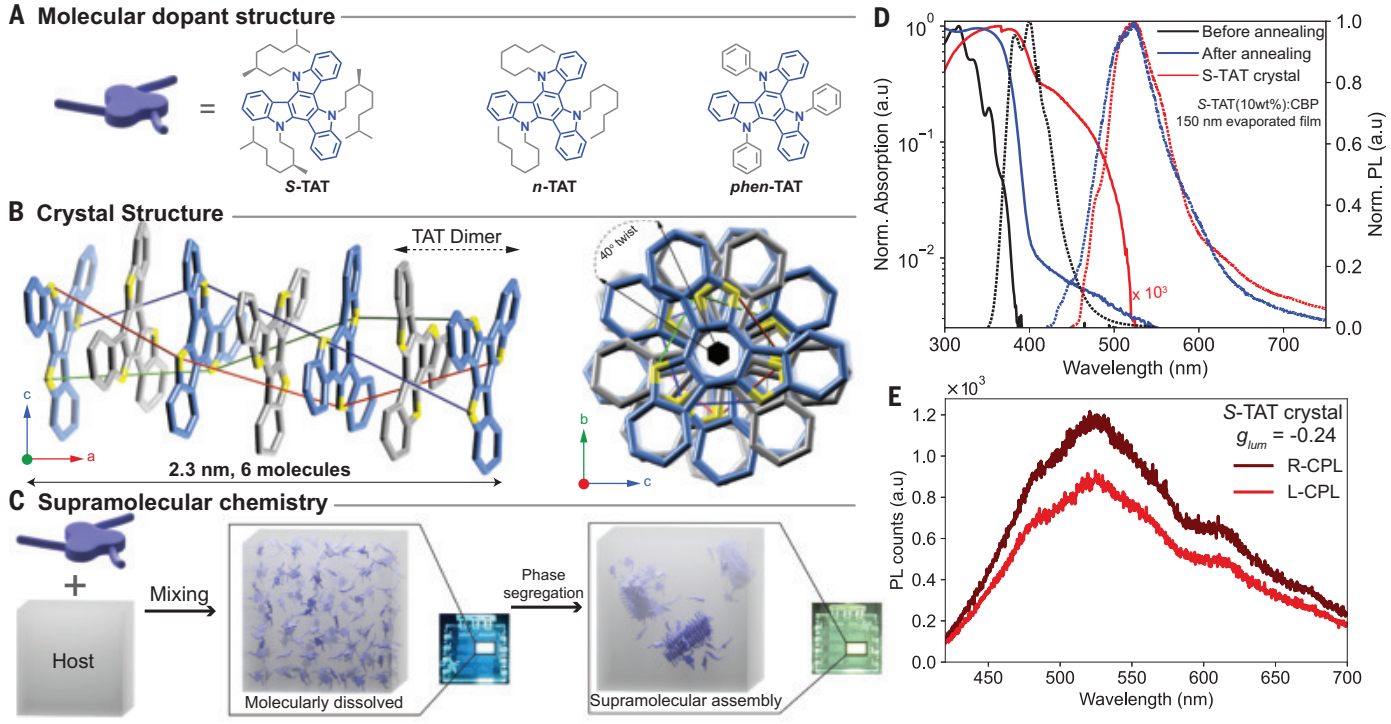


Fig. 1. Supramolecular chiral assembly of chromophores. (**A**) The triazatruxene molecules synthesized and studied. *S*-3,7-dimethyloctyltriazatruxene (*S*-TAT), *n*-octyltriazatruxene (*n*-TAT), phenyltriazatruxene (*phen*-TAT). (**B**) The crystal structure of *S*-TAT. (Left) Viewed along the *b* axis of the unit cell, the components of the *S*-TAT dimer are colored blue and gray; this is the repeating unit of the helix. There are six molecules in a full turn of the helix, which extends along a pitch of 2.3 nm. (Right) View along the *a* axis of the unit cell. (**C**) Supramolecular assembly of TAT within a molecular semiconductor host; a cartoon of the proposed arrangement and a picture of the device. (**D**) Optical absorption measured by photothermal deflection spectroscopy (solid lines) and emission (dashed lines) of the *S*-TAT pure crystal (red), and the 150-nm-thick vacuum-sublimed thin film of 10 wt % TAT in CBP before annealing (black) and after annealing (blue). Further analysis of the spectra is provided in the supplementary materials. (**E**) Circularly polarized photoluminescence (CP-PL) spectrum of the pure *S*-TAT crystal. We observe a CP-PL dissymmetry, g_{lum}, of −0.24. We have used a 400-nm pulsed laser excitation at a fluence 5 μJ cm^{-2} throughout.

some broadening of the spectra, but relatively little shift in the peak PL energy. These were small changes, and we note that redshifts have been seen at higher pressures in π-stacked organic systems (*18*).

We performed measurements of circularly polarized photoluminescence (CP-PL) by distinguishing between the left-handed (L-CPL) and right-handed (R-CPL) components of emitted photons, using a super-achromatic quarter waveplate followed by a linear polarizer that served as an analyzer (see materials and methods) (*9*, *15*, *19*, *20*). We also checked the CP-PL using a photoelastic modulator and obtained identical results, confirming appropriate use of the quarter waveplate method (*21*, *22*). The CP-PL spectrum of the *S*-TAT crystals is shown in Fig. 1E. The spectrum was centered at 540 nm, which clearly showed a difference between the L-CPL and R-CPL components with $g_{lum} = -0.24$. Plots of the intensity against the angle of the rotating quarter waveplate have been shown for all relevant systems (figs. S10 to S13).

Electronic structure and chiral emission

We performed first-principles electronic calculations on a single chiral stack of *S*-TAT; for computational ease, we reduced the alkyl side chains to methyl groups (for details, see materials and methods). We found that, moving from the isolated TAT molecule to the stack, the energy of the π-π* gap decreased from 3.2 to 2.67 eV, a redshift of ~0.53 eV, owing to the considerable π-π intermolecular stacking with relatively strong orbital overlap and strong dipole-dipole interactions. We attributed the experimentally observed redshift from blue to green emission on crystallization to these intermolecular interactions. The intermolecular coupling gave rise to a dispersive electronic band structure (Fig. 2A).

Isolated TAT molecules possess doubly degenerate highest occupied molecular orbital (HOMO) (H) and lowest occupied molecular orbital (LUMO) (L) energy levels. In the stacked phase (Fig. 2A), these energy levels are degenerate at the Z point; however, we found that intermolecular coupling lifted their degeneracy, with the two bands dispersing differently. We found that the two degenerate levels carry opposite magnetic quantum number, $m_l = \pm 1$, and couple differently to their twisted neighboring molecule owing to differences in the orbital overlap between neighboring molecules. The two π and two π* bands were folded back within the reduced Brillouin zone to give six bands, which were gapped between the third and fourth bands because of the dimerization, opening gaps at −0.12 eV and at 2.9 eV, in the HOMO and LUMO, respectively.

The two π HOMO and two π* LUMO bands would be degenerate in the absence of a chiral structure, as they were for the isolated molecule (fig. S45). However, in the presence of the chiral intermolecular π and π* contacts, this degeneracy was lifted, to form two bands with $L = 1$, with energy-separated $m_L = +1$ and $m_L = -1$ bands. These bands dispersed differently at general k values in the Brillouin zone but converged at the band minima and maxima (Fig. 2A). We also observed an additional band at 2.88 eV, stemming from the LUMO+2 orbital. This band couples with the neighboring LUMO and LUMO+1 orbitals, leading to an avoided crossing.

We extracted the optical selection rules of each pair of bands via the momentum matrix element, $\langle \psi_f | \hat{p} | \psi_i \rangle$. We evaluated these matrix elements at all k points; these elements contain the transition dipole selection rules and directly determined the handedness of the

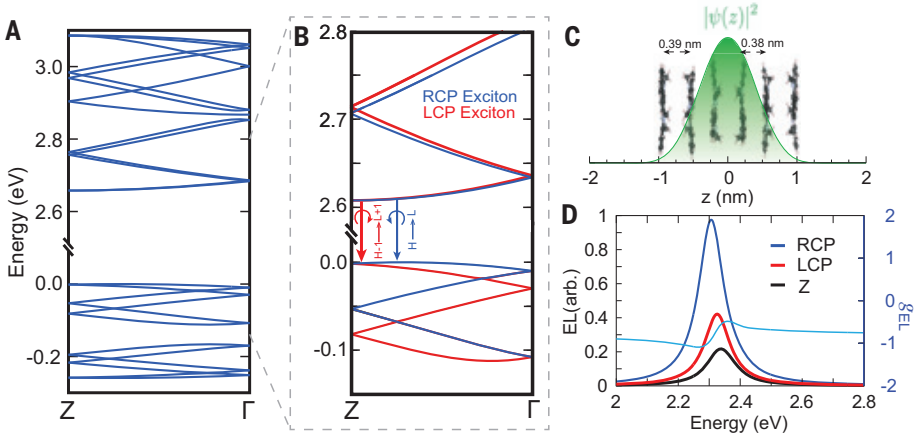


Fig. 2. Ab initio calculations on the chiral electronic structure of *S*-TAT. (**A**) Electronic band structure of stacked-TAT (blue) with lattice spacing a = 0.39 nm, thus Z = $\pi/6a$. (**B**) Zoomed-in electronic band structure near the band extrema, showing the corresponding right-handed (blue) and left-handed (red) bands. The HOMO (LUMO) and HOMO-1 (LUMO+1) are highlighted in blue and red, respectively. Relevant optical transitions are marked with vertical arrows with their circular polarization indicated by the curved arrow. Note that the valence band maximum for the RH (blue) band is away from Z at the position shown by the blue vertical arrow. (**C**) Real-space excitonic wave function, $\psi(R)^2$, overlaid onto the unit cell, showing that the exciton is spread over multiple molecules. (**D**) Modeled luminescence spectrum at 300 K of right-handed (blue), left-handed (red), and Z (black) polarized light. The calculated dissymmetry factor is shown in cyan.

optical transition. We found that the lowest-energy transition H→L carries right-handed circular polarization (RCP) and the H-1→L+1 transition carries left-handed circular polarization (LCP) (Fig. 2, A and B). These selection rules can be understood from the change in magnetic quantum number Δm, between the initial and final state. Note that the magnetic quantum number is defined as modulo 3 (due to C_3 symmetry), and hence $\Delta m_l = \pm 2$ is equivalent to $\Delta m_l = \pm 1$. We found that for the H→L (or H-1→L+1), $\Delta m_l = +1(-1)$, corresponding to RCP (LCP). By contrast, we found that H→L+1 and H-1→L result in $\Delta m_l = 0$, corresponding to Z polarization.

The separation of the $m_L = +1$ and $m_L = -1$ bands at general k points can be viewed intuitively to arise from the different intermolecular orbital overlaps: the $m_L = +1$ circularly polarized band has the same handedness as the helical stack in the unit cell, whereas the $m_L = -1$ band has opposite handedness. A similar analysis (*23*) of cyclic π-systems attributed the origin of magnetic circular dichroism (CD) in these systems to the wave function at general k points away from the high-symmetry points; this is also called the perimeter model (*24*). Our first-principles results indicate that the top of the first ($m_L = -1$) valence band was away from the Z point (Fig. 2B), which we attributed to the interplay between the chiral stacking and low symmetry of the electronic orbitals, leading to enhanced intermolecular interaction at this momentum. It was interesting to see the size of the band splitting as caused by the chiral structure. This reached values of up to 30 meV for the valence band at $k = \pi/3a$.

The Wannier equation approach was used to calculate the excitonic envelope functions (*25*, *26*) (Fig. 2C). We found that the singlet, $S_{1,stack}$, exciton was spread over four molecules with a Bohr radius of ~0.5 nm, which is larger than the intermolecular separation; this result suggests delocalization of the exciton along the TAT stack. Near the band gap, the weaker intermolecular coupling for the HOMO is clearly lower than for HOMO-1. As such, the exciton is composed of a hole in the HOMO that is much more spatially localized than the hole in the HOMO-1 level, leading to distinct excitonic resonances for the HOMO→LUMO and HOMO-1→LUMO+1 excitons.

We emphasize that in reciprocal space, the construction of the exciton from band states required a spread of k that extended from Z to ≈ $\pi/4a$. Together with the substantial energy splitting of the two bands, it was sufficient to drive population selection for the two excitons at room temperature. Quantitatively, we used the Elliot formula (*25*, *26*) (see materials and methods) to calculate PL spectra of the TAT stack (Fig. 2D). Assuming the exciton population is thermalized, we found that the R-CPL was larger than the L-CPL and Z-polarized PL. The calculated g_{lum} (Fig. 2D) clearly showed the expected CP-PL. We attributed the CP-PL to the larger chiral stacking modulated intermolecular coupling of the HOMO-1. In a single molecule, there is no intermolecular coupling, hence the L-CPL and R-CPL transitions happened with equal probability, leading to negligible CPL. We explored other structural arrangements of the TAT molecules (SM section 4.2) that may be present because of stacking faults, including a nondimerized structure (figs. S44 to S46). We found that these alternate structures gave neither CPL nor bright optical transitions.

Our model captured the origin of the large redshift and large circular polarization at the supramolecular level: The intermolecular π-π stacking extends the exciton across multiple molecules, thus lowering its energy, and the extension across multiple molecules can generate a structurally imposed lifting of the exciton degeneracy, resulting in defined angular momentum selection rules that allow for handed photons to be emitted from the chiral supramolecular assembly. We note that circular polarization arose here through the electric dipole matrix element but only at low-symmetry k points, as directly calculated from the band structure. We did not introduced spin-orbit coupling because this has a much smaller energy scale for simple hydrocarbons. For this structure, we achieved the substantial splitting of the $m_L = +1$ and $m_L = -1$ bands needed for circular polarization of luminescence. This offers a starting point for the design of new chirally organized supramolecular structures.

Vacuum sublimation and fabrication of chiral thin films

For practical purposes, single crystals are not feasible. We performed vacuum cosublimation to fabricate thin films with 10 weight % (wt %) of the *S*-TAT molecule alongside a large selection of *p*-type and *n*-type organic semiconducting hosts (table S2). We focused on the results obtained from thin films with 10 wt % *S*-TAT doped in CBP and 3′,5′-di(carbazol-9-yl)-[1,1'-biphenyl]-3,5-dicarbonitrile (DCzDCN). CBP is a hole-transporting semiconductor (*27*), whereas DCzDCN is an ambipolar semiconductor (*28*–*30*). We found that crystallization in all systems could be triggered through annealing above the melting temperature (T_m) of *S*-TAT of ~353 K (fig. S35). This led to the formation of spherulitic domains in the thin films, whose development could be tracked through polarized optical microscopy (fig. S32). Small angle x-Ray scattering (SAXS) confirmed that thermally triggered phase-segregated structures were formed in the bulk (fig. S36). Atomic force microscopy (AFM) studies showed that these spherulites were built up from three-dimensional interpenetrated long crystalline structures that run parallel to the substrate (figs. S30 to S33).

Figure 3A shows the evolution of the CD spectrum upon annealing of a pristine film. A large CD signal was observed after annealing, indicating the breaking of symmetry in the

crystallization process. When spatially resolved, using Mueller matrix imaging (MMPi) (fig. S34), we observed that this large CD signal was present across the spherulitic domains of the annealed thin film and did not contain linear artefacts. The absorption edge and PL spectrum were red-shifted by 0.65 eV after annealing, which electronically resembles the *S*-TAT crystal, indicating that a similar electronic structure and chiral π-stacked packing were achieved by annealing the thin films (Fig. 1D).

The CP-PL spectrum of the annealed films was measured using a home-built setup with a rotating quarter-waveplate and fixed linear polarizer (*9*, *19*, *20*, *31*, *32*) (see materials and methods). We measured a large dissymmetry factor, g_{lum} ~ 10^{-1}, for both thin films in the 500- to 650-nm domain. The pristine films with blue emission showed a PLQE of ~12% and a lifetime of 2.7 ns, whereas the annealed films with green PL showed a delayed lifetime of 23 ns, with ~20% of PL delayed beyond 100 ns. The transient response of the PL shows that there is little change in the spectrum out to 100 μs. Further information is shown in figs. S6 to S9.

Circularly polarized organic light-emitting diodes

We used successive vacuum sublimation of charge injection, transport and recombination layers to make LED structures using 20-nm-thick, 10 wt % *S*-TAT doped thin films as the recombination and emissive layer (EML). This architecture (*29*) is shown in Fig. 4A.A broad range of different electron-transporting and hole-transporting host materials and neat *S*-TAT as EMLs (figs. S19 to S24 and table S2) was screened. In the case of a pure *S*-TAT EML, similar performances were obtained compared with *S*-TAT (10 wt %):CBP blends. Irrespective of the host material, all devices exhibited green EL with EQE values >10%, high current efficiencies, and moderate to excellent brightness (table S2). For both *S*-TAT(10%):CBP and *S*-TAT(10%):DCzDCN emitting layers (EMLs), dissymmetric EL was recorded (Fig. 4, C and D), with EL dissymmetry g_{EL} ~ 10^{-1} (angular resolution of EL dissymmetry is shown in fig. S10). The best-performing OLED was obtained from the ambipolar host DCzDCN (*30*). An EQE_{max} of 15.7% was obtained with a CE_{max} of 45.0 cd A^{-1} at a low turn-on voltage of 2.2 V (Fig. 4, D to F). This device exhibited bright electroluminescence with a L_{max} as high as 57,000 cd m^{-2} (Fig. 4E). The efficiency roll-off (Fig. 4D) was substantially improved in DCzDCN-containing devices compared with the CBP-containing devices.

Our vacuum deposition system required that the vacuum be broken to change organic evaporation sources (a total of seven sources were used). LEDs were not encapsulated beyond the deposition of the top aluminum layer and were measured in air at room temperature. These conditions are not optimized for lifetime studies, but nevertheless we measured lifetimes of >100 hours at 100 cd/m^2 to reach 50% brightness for DCzDCN devices (fig. S25). These lifetimes were substantially longer than those that we reported for spin radical LEDs (*28*) and for hyperfluorescent LEDs (*33*). Our devices did not suffer from the spectral instability reported for 2D perovskite devices (*9*). We consider this is a promising indication that these are inherently stable structures and devices.

The LEDs that we fabricated showed an EQE of >15%. Considering the PLQE (up to 60% for the EML), the 3:1 spin statistics of triplet: singlet excitons (*34*), and ideal light-coupling efficiency (*35*), we would expect a maximum EQE of only ~5% from our thin films. We notably found this enhanced EQE for both *S*-TAT and *n*-TAT, and emphasize that both show local chiral domains, with equal amounts of left-handed and right-handed domains for *n*-TAT (Fig. S34). By contrast, nonassembled blue-emitting *phen*-TAT showed an EQE that was a factor of about 4 lower than the PLQE (table S2). We propose two possible explanations for this unexpected behavior, both of which depend on the presence of the supramolecular π-π packing:

1) If both singlet and triplet excitons are generated by electron hole recombination, efficient luminescence would require a mechanism similar to that of TADF (thermally activated delayed fluorescence) OLEDs (*36*). Such a mechanism requires that the exchange energy between singlet and triplet be thermally accessible and is achieved in TADF systems by forming charge-transfer excitons. However, for isolated TAT molecules, this cannot be the case. We therefore consider that the large redshift of the singlet from isolated molecule to supramolecular stack (0.65 eV) brings the singlet close in energy to the triplet. The T_1 energy of TAT was recently reported (*37*) to be ~2.3 eV, close to the observed luminescence. There is evidence for some delayed emission after photoexcitation, up to 20% of the total (fig. S7). This slow emission may arise from thermal activation from the nonemissive triplet, as observed in TADF systems, although the component of slow emission seen here is smaller than for most TADF systems. We might expect that under pressure, the singlet exciton would fall to a lower energy with respect to the triplet exciton, but note the weak pressure dependence of emission energy and reduced PL yield under pressure (figs. S27 and S28).

2) The origin of CPL for this material is the result of the energy separation of the two valence

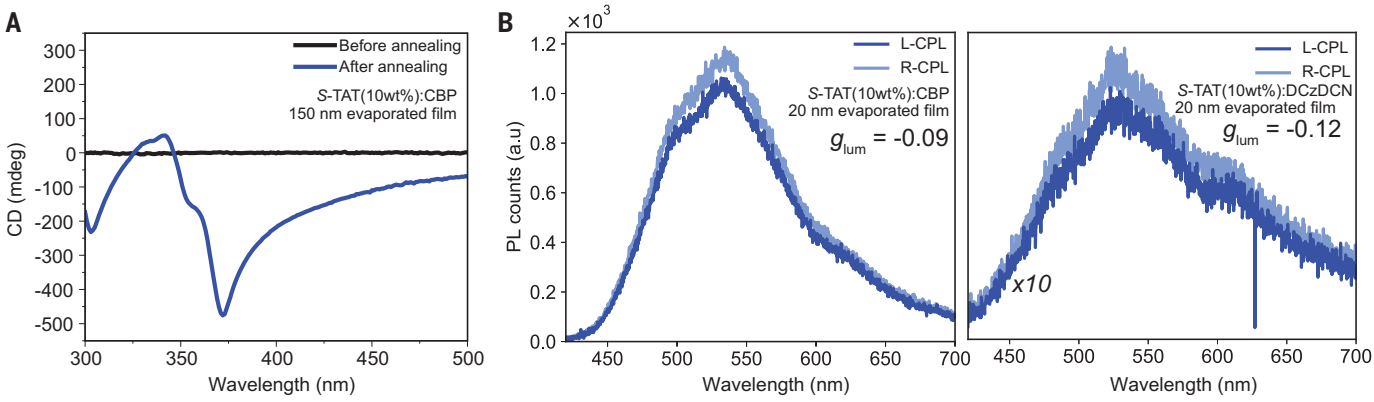


Fig. 3. Photophysics of the vacuum-sublimed *S*-TAT–doped thin films. (**A**) CD spectra obtained by averaging the front and back CD component measured on vacuum-sublimed *S*-TAT(10%):CBP thin films before annealing (black) and after annealing (blue). For the full Mueller matrix imaging, see fig. S34. (**B**) CPL spectrum on annealed, vacuum-sublimed 20-nm-thick films of (left) *S*-TAT(10%):CBP and (right) *S*-TAT(10%):DCzDCN with a g_{lum} = −0.09 and g_{lum} = −0.12, respectively. A fluence of 12 μJ cm^{-2} for the CBP films and 2.0 μJ cm^{-2} for the DCzDCN films was used. In both cases, a 400-nm pulsed laser excitation was used that lies 0.15 eV below the first excited state of amorphous and molecular *S*-TAT.

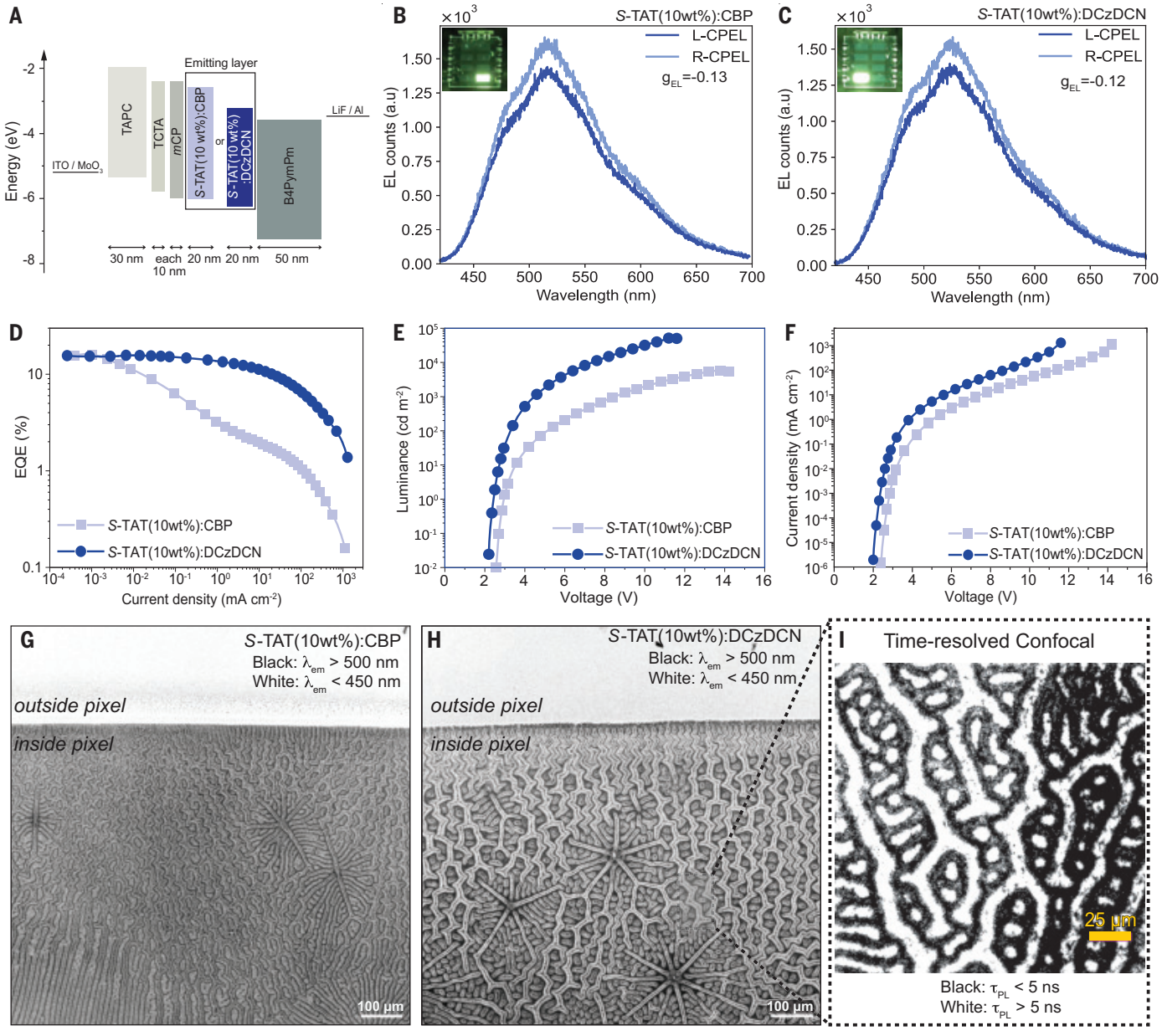


Fig. 4. Circularly polarized organic light-emitting diodes (CP-OLEDs). (**A**) Device architecture for the fabrication of CP-OLEDs using vacuum sublimation. (**B** and **C**) Circularly polarized electroluminescence spectra for *S*-TAT(10%):CBP and *S*-TAT(10%):DCzDCN emitting layers. (**D**) External quantum efficiency (EQE) versus current density curves for devices using *S*-TAT (10%):CBP (light-blue) and *S*-TAT(10%):DCzDCN (dark blue) emitting layers. (**E**) Luminance versus voltage curves for devices using *S*-TAT(10%):CBP (light-blue) and *S*-TAT(10%):DCzDCN (dark blue) emitting layers. (**F**) Current density versus voltage curves for devices using *S*-TAT(10%):CBP (light-blue) and *S*-TAT(10%):DCzDCN (dark blue) emitting layers. Steady-state confocal images of the emitting layer in devices using (**G**) *S*-TAT(10%):CBP and (**H**) *S*-TAT (10%):DCzDCN as the emitting layer. A 450-nm excitation was used to capture emission image with λ_{em} > 500 nm; a 380-nm excitation was used to capture emission image with λ_{em} < 450 nm. (**I**) Time-resolved confocal image on a smaller region in the emitting layer where the regions with τ_{PL} > 5ns and τ_{PL} < 5ns are mapped. Images [(G) and (H)] show that crystalline green domains are phase segregated after being formed; these same domains show slow PL kinetics seen in (I).

and two conduction bands that were split by their orbital angular momentum. The distinction between the LH and RH bands only develops for nonzero values of k points. This has been explored theoretically by Joseph Michl (*23*, *24*). Spin-orbit coupling is, as expected, low for these hydrocarbon materials (as evident from the absence of spin-orbit splitting in the calculated electronic structure) (Fig. 2). However, it is interesting to consider possible coupling of electron spin to orbital angular momentum for injected carriers. One such process is called chiral-induced spin selectivity, which has been identified as a mechanism for spin-polarized injection and transport in chiral semiconductor structures (*38*–*41*).

We also investigated the microstructural origin of the emission regions within the TAT or host emissive layers. Confocal microscopy images on the EML of the driven pixels in the CP-OLEDs with *S*-TAT (10%):CBP and *S*-TAT (10%):DCzDCN are shown in Fig. 4, G and H. Upon 450-nm excitation, the regions with a photoluminescence wavelength (λ_{PL}) > 500 nm displayed a large,

patterned network that resembled the aftermath of viscoelastic phase-separated mixtures (*42*). Time-resolved confocal microscopy maps (Fig. 4I) showed that the same pattern emerged with regions that had a lifetime $\tau_{PL} > 5$ ns. Conversely, when using a 350-nm excitation region with blue PL, which was also the region with $\tau_{PL} < 5$ ns, the patterns were found in between the green-emitting regions. Comparing the microscopy results in Fig. 4, G and H, to the PL spectrum in Fig. 1D, we attributed the green-emitting regions in the EML to the chiral helical, π-stacked *S*-TAT molecules, which emit green light. Additionally (Fig. 3, C and D), these green-emitting domains had a lifetime of τ_{PL} ~ 23 ns, which was verified spatially using the time-resolved-confocal image (Fig. 4I and figs. S26 and S27). Upon close inspection, the emergent patterns formed by the $\lambda_{PL} > 500$ nm regions resembled those obtained after viscoelastic phase separation of kinetically asymmetric mixtures (*43*, *44*). In the thin film, local heating could quickly exceed 100°C during device operation, above the melting temperature of *S*-TAT but below the melting temperature of the host. Under such conditions, the quickly melted *S*-TAT phase could possibly reorganize and flow inside the host matrix to trigger a strong bias for driving phase separation between the slow and fast components (*45*). In the case of *S*-TAT, the long alkyl chains prevent host-guest interactions, providing a strong driving force for phase segregation and crystallization of *S*-TAT into the preferred helical π-π stacked supramolecular assemblies (*46*, *47*). When the interaction between the host and dopant was larger, such as when 10 wt % *phen*-TAT was used (fig. S37), this phase separation was hampered, resulting in only a slight redshift in the PL and no emergent CP-PL from annealed films. The vacuum-sublimed devices with *phen*-TAT (10 wt %):CBP EML also had a much lower EQE of ~3%, which is close to the expected theoretical value, although obviously no CP-EL from the vacuum-sublimed OLEDs was observed. Furthermore, these *phen*-TAT–doped OLEDs did not show any emergent phase-separated structures when subjected to confocal microscopy. Finally, when investigating the *n*-TAT devices with *n*-TAT (10 wt %):CBP, most optical properties and device characteristics were found to be similar to the S-TAT-based devices, except that a dissymmetry of g_{EL} ~ 0 was obtained, consistent with the racemic nature of the *n*-TAT stacks. These observations indicate that the nature of the N-substituent is directly involved in the formation of supramolecular structures in the films, which is further linked to the photophysical and electrical properties of the thin films.

A clear picture for the photophysical properties that we observed (Fig. 1C) emerged: In situ crystallization of *S*-TAT rich domains into the chiral supramolecular architecture was triggered by the local heating produced by electric driving. Crystallization was propelled by the large entropy of the molten *S*-TAT phase (it has the lower T_m in all the cases we studied), which did not strongly interact with the slow-melting host matrix. The crystallization led to a viscoelastic microphase segregation. As discussed earlier, the strong chirality of the *S*-TAT supramolecular stacking imposes angular momentum selection rules on emitted photons (Fig. 2).

In summary, we have set out a scheme for the fabrication of efficient OLEDs, which contains chiral structures to give circularly polarized emission without lowering the efficiency and brightness. Vacuum cosublimation of guest-host systems selected to allow in situ crystallization to the chiral structure allowed controllable thin-film processing that is fully compatible with OLED fabrication. Other approaches to achieve CP emission from LEDs use spin-selective electrodes, such as magnetically polarized metal spin injectors (*48*) and chiral spin-injecting 2D metal halide perovskites (*8*–*10*). We have shown here a molecular structure in which CPL arises from the electronic bandstructure and does not require spin-orbit coupling (Rashba splitting), as recently invoked for 2D perovskites (*49*). This result may have broad applications: Circularly polarized emission is of potential value for light control in LED displays, and this offers opportunities to explore spin-selective transport processes (*38*, *41*).

REFERENCES AND NOTES

1. K. Ariga, T. Mori, T. Kitao, T. Uemura, *Adv. Mater.* **32**, e1905657 (2020).
2. M. B. Baker *et al*., *Nat. Commun.* **6**, 6234 (2015).
3. P. Besenius *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **107**, 17888–17893 (2010).
4. S. M. C. Schoenmakers *et al*., *ACS Macro Lett.* **11**, 711–715 (2022).
5. J. H. K. K. Hirschberg *et al*., *Nature* **407**, 167–170 (2000).
6. C. Kulkarni *et al*., *Adv. Mater.* **32**, e1904965 (2020).
7. A. Ashoka *et al*., *Nat. Mater.* **22**, 977–984 (2023).
8. M. P. Hautzinger *et al*., *Nature* **631**, 307–312 (2024).
9. Y.-H. Kim *et al*., *Science* **371**, 1129–1133 (2021).
10. H. Lu, Z. V. Vardeny, M. C. Beard, *Nat. Rev. Chem.* **6**, 470–485 (2022).
11. B. M. W. Langeveld-Voss *et al*., *J. Am. Chem. Soc.* **118**, 4908–4909 (1996).
12. E. Peeters *et al*., *J. Am. Chem. Soc.* **119**, 9909–9910 (1997).
13. D. Di Nuzzo *et al*., *ACS Nano* **11**, 12713–12722 (2017).
14. Y. Yang *et al*., *ACS Nano* **11**, 8329–8338 (2017).
15. J. R. Brandt, X. Wang, Y. Yang, A. J. Campbell, M. J. Fuchter, *J. Am. Chem. Soc.* **138**, 9743–9746 (2016).
16. Y. Xu, Q. Wang, X. Cai, C. Li, Y. Wang, *Adv. Mater.* **33**, e2100652 (2021).
17. F. Zinna, U. Giovanella, L. Di Bari, *Adv. Mater.* **27**, 1791–1795 (2015).
18. J. P. Schmidtke, J.-S. Kim, J. Gierschner, C. Silva, R. H. Friend, *Phys. Rev. Lett.* **99**, 167401 (2007).
19. F. Zinna *et al*., *Chemistry* **22**, 16089–16098 (2016).
20. D. F. De Rosa, P. Stachelek, D. J. Black, R. Pal, *Nat. Commun.* **14**, 1537 (2023).
21. S. C. J. Meskers, *ChemPhotoChem* **6**, e202100154 (2022).
22. W. R. Kitzmann, J. Freudenthal, A. M. Reponen, Z. A. VanOrman, S. Feldmann, *Adv. Mater.* **35**, e2302279 (2023).
23. J. Michl, *J. Am. Chem. Soc.* **100**, 6801–6811 (1978).
24. J. Michl, *Tetrahedron* **40**, 3845–3934 (1984).
25. M. Kira, S. W. Koch, *Prog. Quantum Electron.* **30**, 155–296 (2006).
26. J. J. P. Thompson *et al*., *Nat. Sci.* **3**, e20220040 (2023).
27. Q. Zhang *et al*., *Nat. Photonics* **8**, 326–332 (2014).
28. H.-H. Cho *et al*., *Nat. Photonics* **18**, 905–912 (2024).
29. H. H. Cho *et al*., *Adv. Mater.* **35**, e2303666 (2023).
30. Y. J. Cho, K. S. Yook, J. Y. Lee, *Adv. Mater.* **26**, 4050–4055 (2014).
31. B. Baguenard *et al*., *Nat. Commun.* **14**, 1065 (2023).
32. M. Morgenroth *et al*., *Nat. Commun.* **13**, 210 (2022).
33. H.-H. Cho *et al*., *Nat. Mater.* **23**, 519–526 (2024).
34. C. J. Bardeen, *Annu. Rev. Phys. Chem.* **65**, 127–148 (2014).
35. R. Meerheim, M. Furno, S. Hofmann, B. Lüssem, K. Leo, *Appl. Phys. Lett.* **97**, 253305 (2010).
36. H. Uoyama, K. Goushi, K. Shizu, H. Nomura, C. Adachi, *Nature* **492**, 234–238 (2012).
37. C.-Y. Lin *et al*., *Nat. Chem.* **16**, 98–106 (2024).
38. B. Göhler *et al*., *Science* **331**, 894–897 (2011).
39. R. Naaman, Y. Paltiel, D. H. Waldeck, *Nat. Rev. Chem.* **3**, 250–260 (2019).
40. A. Chiesa *et al*., *Adv. Mater.* **35**, e2300472 (2023).
41. B. P. Bloom, Y. Paltiel, R. Naaman, D. H. Waldeck, *Chem. Rev.* **124**, 1950–1991 (2024).
42. H. Tanaka, *J. Phys. Condens. Matter* **12**, R207–R264 (2000).
43. H. Tanaka, T. Araki, *Chem. Eng. Sci.* **61**, 2108–2141 (2006).
44. H. Tanaka, T. Araki, T. Koyama, Y. Nishikawa, *J. Phys. Condens. Matter* **17**, S3195–S3204 (2005).
45. A. Brunk *et al*., *J. Phys. Condens. Matter* **33**, 234002 (2021).
46. W. M. Jacobs, D. W. Oxtoby, D. Frenkel, *J. Chem. Phys.* **140**, 204109 (2014).
47. D. Frenkel, A. A. Louis, *Phys. Rev. Lett.* **68**, 3363–3365 (1992).
48. N. Nishizawa, K. Nishibayashi, H. Munekata, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 1783–1788 (2017).
49. P. C. Sercel, M. P. Hautzinger, R. Song, V. Blum, M. C. Beard, *Adv. Mater.* e2415901 (2025).
50. R. Chowdhury, Circularly polarized electroluminescence from chiral supramolecular semiconductor thin films, Zenodo (2025). https://doi.org/10.5281/zenodo.14782007.

ACKNOWLEDGMENTS

We acknowledge the Diamond Light Source B23 beamline for the MMP mapping. We thank L. Minion and M. J. Fuchter for measurements of CPL using quarter-waveplate and photoelastic modulator methods and M. Beslac for synthetic support. **Funding:** European Research Council (ERC) grant agreement number SCORS – 101020167 (RHF, HHC, RC); European Research Council (ERC), Horizon 2020 research and innovation programme grant agreement number SYNMAT – 788618 (EWM); European Research Council (ERC), Horizon 2020 Marie Skłodowska-Curie Actions grant agreement no. HEL4CHIROLED-859752 (R.C., M.D.P., A.T.); Dutch Ministry of Education, Culture and Science Gravity program no. 024.001.035 (E.W.M., S.C.J.M., J.v.d.T., B.W.L.v.d.B., G.V.); Engineering and Physical Sciences Research Council (EPSRC) Programme Grant no. EP/W017091/1 (B.M., J.J.P.T.); United Kingdom Research and Innovation (UKRI) Future Leaders Fellowship no. MR/V023926/1 (B.M.); Gianna Angelopoulos Programme for Science (BM); Bill & Melinda Gates Foundation Gates-Cambridge scholarship no. OPP1144 (S.S.); Lindemann Trust (T.K.B); The Winton Programme for the Physics of Sustainability (B.M., Y.B.); Agency of Science, Technology and Research (A*STAR, Singapore) National Science Scholarship (X.W.C.); PhD scholarship from Cambridge Trust and the George and Lilian Schiff Foundation (P.G.); Dutch Research Council (NWO) research programme on 'Innovations for wind and solar energy (KIC)' project PerFoRM with file no. KICH1.ED02.20.007 (D.M.D., B.E.); European Research Council (ERC), Horizon 2020 Marie Skłodowska-Curie Individual Fellowship agreement number 101022365 (S.T.K.); European Research Council (ERC) Horizon 2020 research and innovation programme grant agreement no. 758826 (A.R.). **Author contributions:** R.H.F. and E.W.M. conceived the project. R.C. conducted steady-state spectroscopy, transient spectroscopy and magneto-optical experiments, and x-ray structure determination. M.D.P synthesized the molecules and carried out the supramolecular assembly studies. H.H.C. designed and fabricated the CP-OLEDs and all vacuum-sublimed films and measured OLED devices. T.K.B. assisted with CPL measurements.

J.J.P.T, K.W.C., and B.M. performed electronic structure calculations. S.S conducted steady-state confocal imaging. P.G. performed Raman spectroscopy. Y.B and A.R. conducted photothermal-deflection spectroscopy. E.G. fabricated and measured CP-OLED devices. X.W.C performed time-resolved confocal microscopy. J.v.T. performed AFM measurements. B.W.L.v.d.B. performed SAXS measurements. A.T. performed Muller-Matrix imaging experiments. N.D carried out cyclic voltammetry measurements. D.M.D and B.E. performed pressure-dependent PL studies. G.V. carried out assembly studies of the molecules. S.C.J.M. critically analyzed, interpreted, and designed the polarized optical studies. R.C., R.H.F., E.W.M., M.D.P., J.J.P.T., and B.M. wrote the manuscript with feedback and approval from all authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials. Further data are available (*50*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adt3011
Materials and Methods
Supplementary Text
Figs. S1 to S46
Tables S1 to S5
References (*51*–*63*)

Submitted 20 September 2024; resubmitted 10 December 2024
Accepted 5 February 2025
10.1126/science.adt3011

ENVIRONMENTAL SCIENCE

Environmental effects of the Kakhovka Dam destruction by warfare in Ukraine

O. Shumilova[1]*†, A. Sukhodolov[2]†, N. Osadcha[3], A. Oreshchenko[3], G. Constantinescu[4], S. Afanasyev[5], M. Koken[6], V. Osadchyi[3], B. Rhoads[7], K. Tockner[8,9], M. T. Monaghan[1,10], B. Schröder[11], J. Nabyvanets[3], C. Wolter[12], O. Lietytska[5], J. van de Koppel[13,14], N. Magas[15], S. C. Jähnig[16,17], V. Lakisova[18], G. Trokhymenko[15], M. Venohr[2], V. Komorin[19], S. Stepanenko[20], V. Khilchevskyi[21], S. Domisch[16], M. Blettler[22], P. Gleick[23], L. De Meester[1,10,24], H.-P. Grossart[25,26]

The use of water as a weapon in highly industrialized areas in the Russo-Ukrainian war has resulted in catastrophic economic and environmental damages. We analyze environmental effects caused by the military destruction of the Kakhovka Dam. We link field, remote sensing, and modeling data to demarcate the disaster's spatial-temporal scales and outline trends in reestablishment of damaged ecosystems. Although media attention has focused on the immediate impacts of flooding on society, politics, and the economy, our results show that toxic contamination within newly exposed sediments of the former reservoir bed poses a largely overlooked long-term threat to freshwater, estuarine, and marine ecosystems. The continued use of water as a weapon may lead to even greater risks for people and the environment.

Modern warfare strategies have reduced the need for ground-based operations, however, rivers not only continue to be combat barriers but are increasingly used as weapons (*1*). Intentional destruction of dams has occurred repeatedly in Ukraine. In 1941, detonation of the Dnipro Dam occurred, killing thousands (*2*). In 2022, several dams were destroyed along the Irpen, Oskil, and Inhulets rivers (*2*, *3*). The most devastating event occurred on 6 June 2023, when the Kakhovka Dam on the Dnipro River collapsed following repeated attacks (*4*–*9*). In March 2024, Russian missiles destroyed power plants of the Dnipro (3.3 km^3), Kaniv (2.6 km^3), and Novodnistrovsk (3 km^3) reservoirs.

Dams are major components of contemporary human infrastructure on rivers throughout the world. More than 50,000 large reservoir facilities (LRFs) exist worldwide (*10*). Although the risk of failure is low (~1%) (*11*), concern is growing given the state of deterioration of many structures (*12*, *13*). The recent collapse of dams in Libya (*14*) highlights the danger of dam failures in the context of global climatic change. However, human conflict has been overlooked as an aspect of risk partly because the intentional destruction of dams is banned by the Geneva Conventions.

Until recently, the Dnipro River included a cascade of six reservoirs stretching across from Ukraine to the Black Sea and containing ~43.6 km^3 of water (Fig. 1). The Kakhovka LRF (hereafter K-LRF, 18 km^3, Fig. 2A) was at the downstream end of this cascade. Its main purpose was to supply water to a 12,000-km irrigation system for 500,000 hectares (ha) of croplands and to a 400-km-long canal for provision of 85% of Crimea's water (*7*, *15*).

The destruction of the Kakhovka Dam caused catastrophic draining of the reservoir, downstream flooding, and contamination of freshwater and marine environments (*8*). However, access to data on this failure has been hindered by ongoing combat which constrains field research (*4*, *8*, *16*), limiting analysis thus far to early rapid assessment (*6*, *7*). We examine the environmental impacts and scales of the K-LRF catastrophe by linking empirical data from ground-based surveys and remote sensing within an analytical framework encompassing insights from dam removal practices (*17*), hydrodynamic modeling, flood hazard assessment (*18*), and analysis of ecosystem reestablishment (*19*) (Fig. 2, materials and methods, figs. S1 to S10, and tables S1 to S8). We show that (i) large amounts of pollutants accumulated in the reservoir sediment before the catastrophe; (ii) excessive drainage and release of water likely caused massive mortality of benthic organisms and fish; and (iii) a massive freshwater plume was present which undoubtedly affected marine life in the northwestern

[1]Department of Evolutionary and Integrative Ecology, Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany. [2]Department of Ecohydrology and Biogeochemistry, Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany. [3]Hydrometeorological Institute, State Emergency Service and the National Academy of Sciences of Ukraine, Kyiv, Ukraine. [4]Department of Civil and Environmental Engineering and IIHR Hydroscience and Engineering, University of Iowa, IA, USA. [5]Institute of Hydrobiology, National Academy of Sciences of Ukraine, Kyiv, Ukraine. [6]Department of Civil Engineering, Middle East Technical University, Ankara, Turkey. [7]Department of Geography and Geographic Information Science, University of Illinois at Urbana-Champaign, Champaign, IL, USA. [8]Senckenberg-Leibniz Institution for Biodiversity and Earth System Research, Frankfurt a.M., Germany. [9]Faculty of Biological Sciences, Goethe-University, Frankfurt a.M., Germany. [10]Institute of Biology, Freie Universität Berlin, Berlin, Germany. [11]Department of Plant Ecology, Technical University of Berlin, Berlin, Germany. [12]Department of Fish Biology, Fisheries and Aquaculture, Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany. [13]Department of Estuarine and Delta Systems, NIOZ Royal Netherlands Institute for Sea Research, AC Yerseke, the Netherlands. [14]Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, the Netherlands. [15]Department of Ecology and Environmental Technologies, Admiral Makarov National University of Shipbuilding, Mykolayiv, Ukraine. [16]Department of Community and Ecosystem Ecology, Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany. [17]Geography Department, Humboldt University of Berlin, Berlin, Germany. [18]International Progressive Education Foundation, Odesa, Ukraine. [19]Ukrainian Scientific Centre of Ecology of the Sea, Odesa, Ukraine. [20]Hydrometeorological Institute, Odesa State Environmental University, Odesa, Ukraine. [21]Department of Hydrology and Hydroecology, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine. [22]The National Institute of Limnology, Ciudad University, Santa Fe, Argentina. [23]Pacific Institute for Studies in Development, Environment and Security, Oakland, CA, USA. [24]Department of Biology, University of Leuven, Leuven, Belgium. [25]Department of Plankton and Microbial Ecology, Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany. [26]Institute of Biochemistry and Biology, Potsdam University, Potsdam, Germany.
***Corresponding author. Email: oleksandra.shumilova@igb-berlin.de**
†These authors contributed equally to this work.

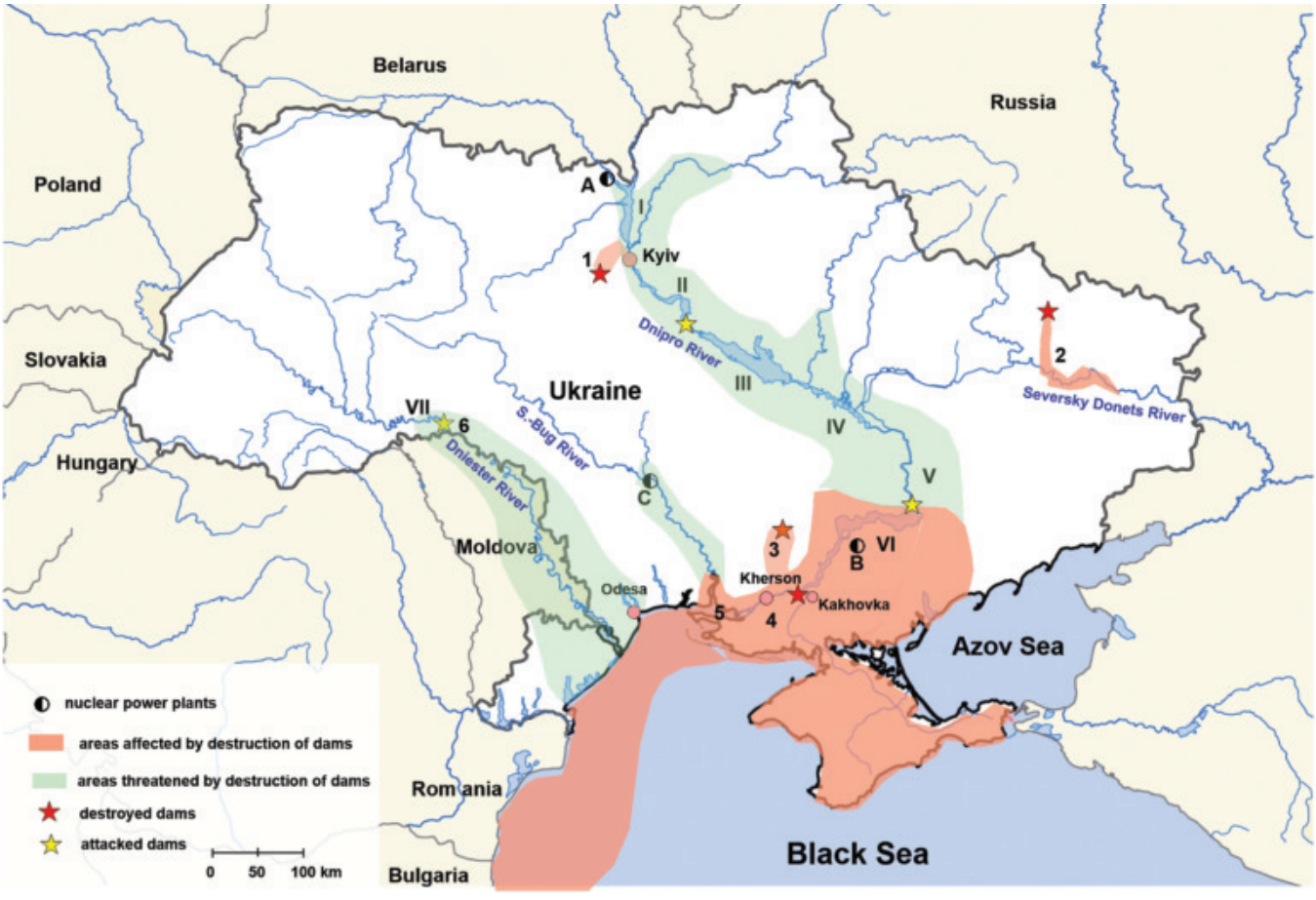


Fig. 1. Areas of Ukraine affected or threatened by dam destruction in military operations. Arabic numbers 1 to 6 indicate rivers: Irpen, Oskil, Inhulets, Dnipro, Dnipro-Bug Estuary, and Dniester, respectively. Roman numbers I to VII indicate large reservoir facilities: Kyiv, Kaniv, Kremenchuk, Kaminske, Dnipro, Kakhovka, and Dniester, respectively. (**A** to **C**) indicate nuclear power plants: Chornobyl, Zaporizhzhia, and South Ukraine, respectively.

Black Sea; (iv) the exposed sediment has become a long-term source of contaminants that can be mobilized by floods; and (v) high growth rates of riparian vegetation suggest that conversion of the former lakebed to vegetated floodplain can be accomplished within five years.

Framework and data

To analyze hazards to humans caused by the Kakhovka flooding, we adopt the US Bureau of Reclamation approach (*18*), which relates loss of human life to values of *DV*, where *D* and *V* are maximum flow depth and velocity, respectively. We extend this analysis by developing an approach to estimate impacts on populations of rodents, which is based on a species-area relation and evaluates population losses as a ratio between the area of danger for swimming animals and the total inundated area. Dangerous areas are identified by critical values of the survival index S_{sw}, which relates swimming abilities of animals (*20*, *21*) to *DV* values (fig. S10).

To calculate spatial patterns of *DV,* we developed a hydrodynamic model of the lower Dnipro with a computational domain (mesh size 16×16 m) based on bathymetric data derived from field and photogrammetric surveys and remotely sensed water surface extent during the peak of the flood (fig. S9). A boundary condition, a water discharge of 29,000 m^3s^{-1}, was estimated based on the decrease in water volume in the reservoir through time. Model validation was performed using ground-based observations (figs. S10 to S12). The model allowed assessment of damage to riparian habitats caused by soil erosion and uprooting of vegetation (figs. S13 to S18).

Drainage of the K-LRF exposed lakebed sediment (Fig. 2, A to C, and figs. S19 to S21), the release of which is a major concern in dam removal projects (*12*, *13*, *17*, *22*). Three concepts developed through dam removal research guided our analysis: (i) Abrupt fall of base level is a key factor controlling mobilization of sediment; (ii) sediment volume, grain size, chemical content, and colonization of sediment by vegetation define the long-term mobilization of toxic substances; and (iii) in highly developed regions, contaminated sediment may require intentional removal. Examples include polychlorinated biphenyls in the Hudson River (*23*) and Agent Orange in the Passaic River (*24*) in the USA. We estimate sediment release using a method of non-erodible velocities (*25*), combined with analysis of remote sensing images and data from water quality monitoring (fig. S18).

Reestablishment of ecosystem biodiversity in the aftermath of dam removal (*12*, *17*, *22*) and levee breaching (*23*, *24*, *26*) is an important aspect of remediation that has been conceptualized in process-based models (*27*, *28*) and recent models of self-organization (*19*, *29*, *30*). We analyze the processes of ecosystem reestablishment using results of field surveys and remote sensing within the context of process-based modeling (Fig. 2D and figs. S1 and S4). We apply our analytical framework to data collected before the dam breach (figs. S5 to S8), during the flood and reservoir draining (figs. S10 to S12, S18, and tables S1 to S6), and over a one-year period following the disaster (table S5 and figs. S23 to S26).

The Kakhovka Dam triggered a toxic "time-bomb"

After construction in the mid-1950s, an increase in nutrient concentration and water temperature

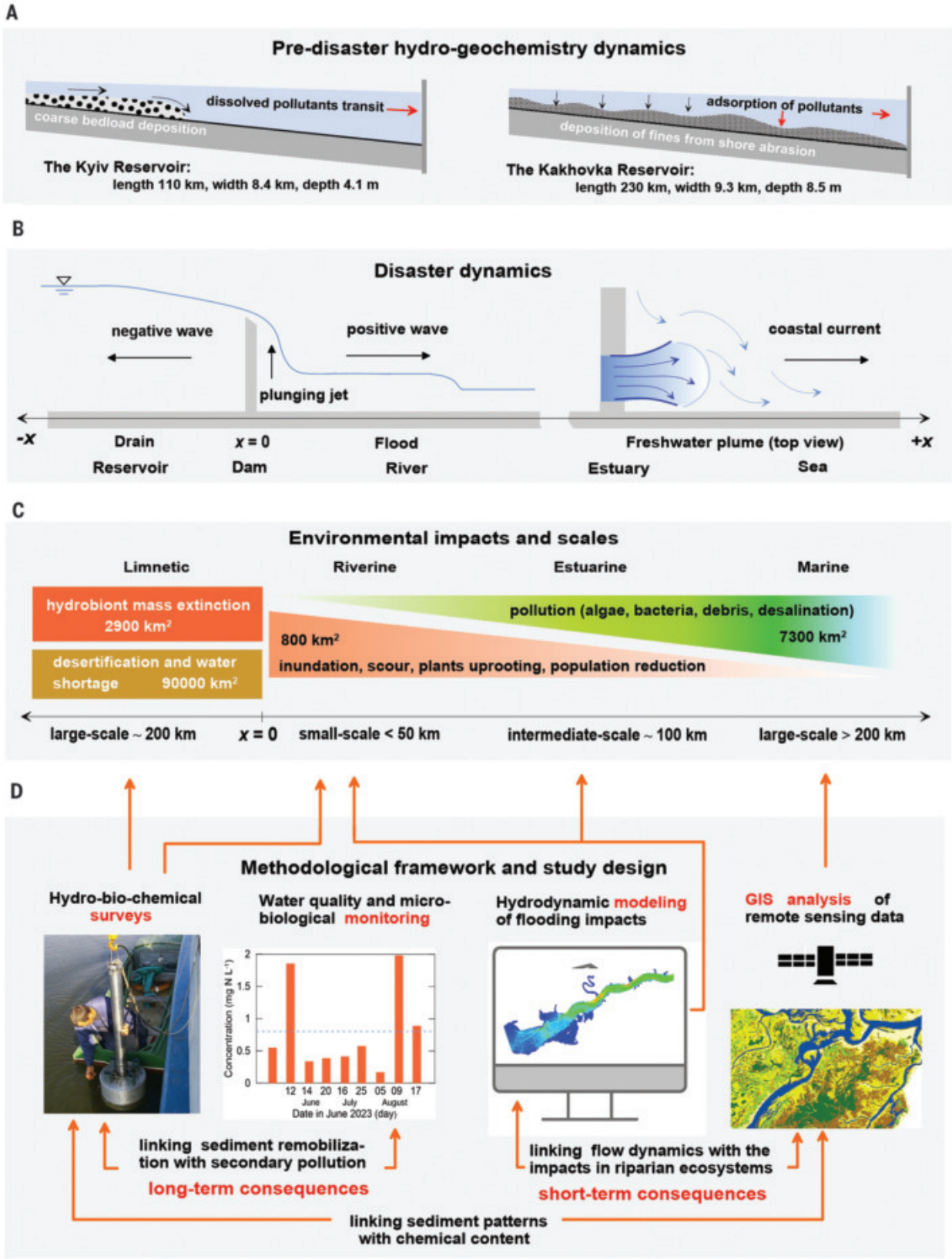


Fig. 2. Conceptual framework of the study. (**A**) Pre-disaster sedimentation dynamics in the Dnipro cascade showing the Kyiv reservoir (upstream) and the Kakhovka reservoir (downstream). (**B**) Schematic of the dam-breaching event (side view on the left) and downstream of the Dnipro-Bug Estuary (top view on the right). (**C**) Environmental impacts and characteristic scale of the Kakhovka Dam disaster. (**D**) Methodology, components, interrelations, and relevance to spatial and temporal scales.

in the K-LRF boosted biological productivity (*31*–*33*). Based on field research prior to 2022, we estimate total standing biomasses of 100,000 to 150,000 tonnes for periphyton, 30,000 to 50,000 tonnes for aquatic plants, 200,000 to 500,000 tonnes for macroinvertebrates, and 6000 to 10,000 tonnes for fish. The diversion of water for agriculture and industry decreased the maximum mean water discharge of the lower Dnipro from 3500 m^3s^{-1} to 2000 m^3s^{-1} (fig. S7). The river system downstream of the dam consists of the main channel, a floodplain, and several meandering side channels. The riverine ecosystem was very diverse (*8*, *15*) with more than 70 fish species, of which 18 were protected species (*33*). The floodplain provided habitat for breeding, migration, and wintering for approximately 350 bird species. Mammals included rodents, canids, wild pigs, and deer (*7*, *33*).

Because the K-LRF is located at the downstream end of the Dnipro reservoir cascade, approximately 98% of the sediment load was retained in upstream reservoirs (*15*) (Fig. 2A). Pre-disaster field surveys reveal that the lakebed contained a layer of sediment 0.3 to 1.5 m thick, composed almost entirely of fine silt eroded from the shoreline. Based on the morphology (fig. S6B) and average thickness of the sediment layer, we estimate the total volume of sediment to be approximately 1.3 to 1.7 km^3. Sorption, hydrolysis, settling, and biological consumption mediated the adsorption of contaminants from the water column to the sediment (*34*). Contaminants include heavy metals, nitrogen, and phosphorus delivered from industrial and agricultural sources (tables S7 and S8). Within the reservoir, some heavy metals such as Pb and Ni are distributed relatively uniformly (figs. S19 and S20) due to mixing by waves and currents, whereas others such as Zn exhibit locally elevated values at industrial outputs near Nikopol (fig. S19). Nondegradable heavy metals can damage the nervous system, disrupt endocrine system functioning, and cause congenital disorders (*35*).

Short-term impacts of flood, drain, and plume spread

Hydraulic theory predicts that two waves formed at the breach (*36*), producing surges in both the down- and upstream directions (Fig. 2B). Floodwater surging downstream formed a plume that spread into the Black Sea. The release of ~16.4 km^3 of water continued for two weeks and imposed large-scale impacts on riverine and marine ecosystems (*3*, *7*, *8*).

Modeling indicates that between Kakhovka and Kherson (Fig. 3B), the reach-averaged *DV* was approximately 10 m^2s^{-1}. Along this section of the river, in which 110,000 people and 60,000 buildings were affected by the flood, 84 lives were lost (*7*). Although this flood hazard is classified as medium severity (*18*), public perception was nonetheless elevated, driven mainly by the scale of economic losses rather than loss of life (*7*, *8*).

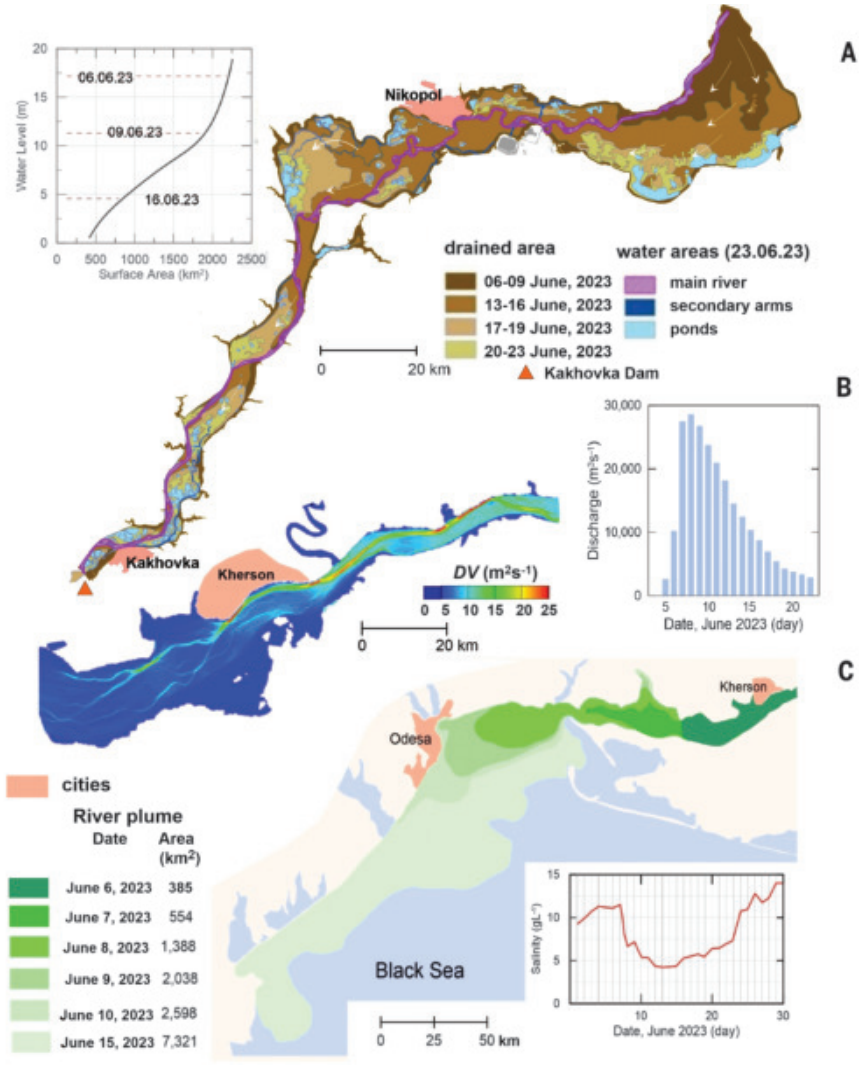


Fig. 3. Dynamics of the dam-breach catastrophe on the Dnipro river, Dnipro-Bug Estuary, and the northern part of the Black Sea. (**A**) Draining of the K-LRF (inset shows how surface area relates to water level of the reservoir; white arrows show the direction of propagating drained land front). (**B**) Modeled depth-average values of *DV* downstream of the K-LRF at maximal water stage (inset shows water discharge at Kherson). (**C**) Dynamics of the river plume in the north shelf of the Black Sea reconstructed from remote sensing data [inset shows measured salinity at Odesa after (*6*)].

Predicted flow velocities (figs. S10C and S12E) range between 1 and 4 ms^{-1} on the floodplain and in the main channel. The highest predicted values are close to 5 ms^{-1}, which commonly occur over rapids (*37*). Breaking waves—a distinctive feature of rapids—were unlikely, due to low Froude numbers (fig. S10D). Strong turbulence developed between the main channel and the floodplains (figs. S13 and S14), enhancing riverbed scour and plant uprooting. However, compared with rapids, these turbulent zones reduced swimming risks by directing currents toward slow-moving margins of the flow (*38*). Estimating hazards to floodplain rodents from patterns of the S_{SW} index suggests a 20 to 30% reduction of the total population (fig. S15C). Rodents occupy lower trophic positions in the food web and losses in this population likely affected populations of canids and birds.

Floods can greatly affect ecosystems by mobilizing large amounts of fine particles (*39*). Our results show that flow scoured sediment within a high-speed core immediately downstream of the breach (figs. S16 and S17) and within turbulent shear layers on the floodplains (fig. S14). Flow entrained about 2.97×10^{-3} km^3 of sediment within the core, 1.9×10^{-3} km^3 from a knick zone (fig. S21), and about 1.8×10^{-3} km^3 within shear layers. Using a reconstructed hydrograph (fig. S22A) and extrapolating measured suspended solids concentrations (fig. S22B), we estimate that reservoir drainage released about 0.78×10^{-3} km^3 of contaminated fine sediment. Destruction

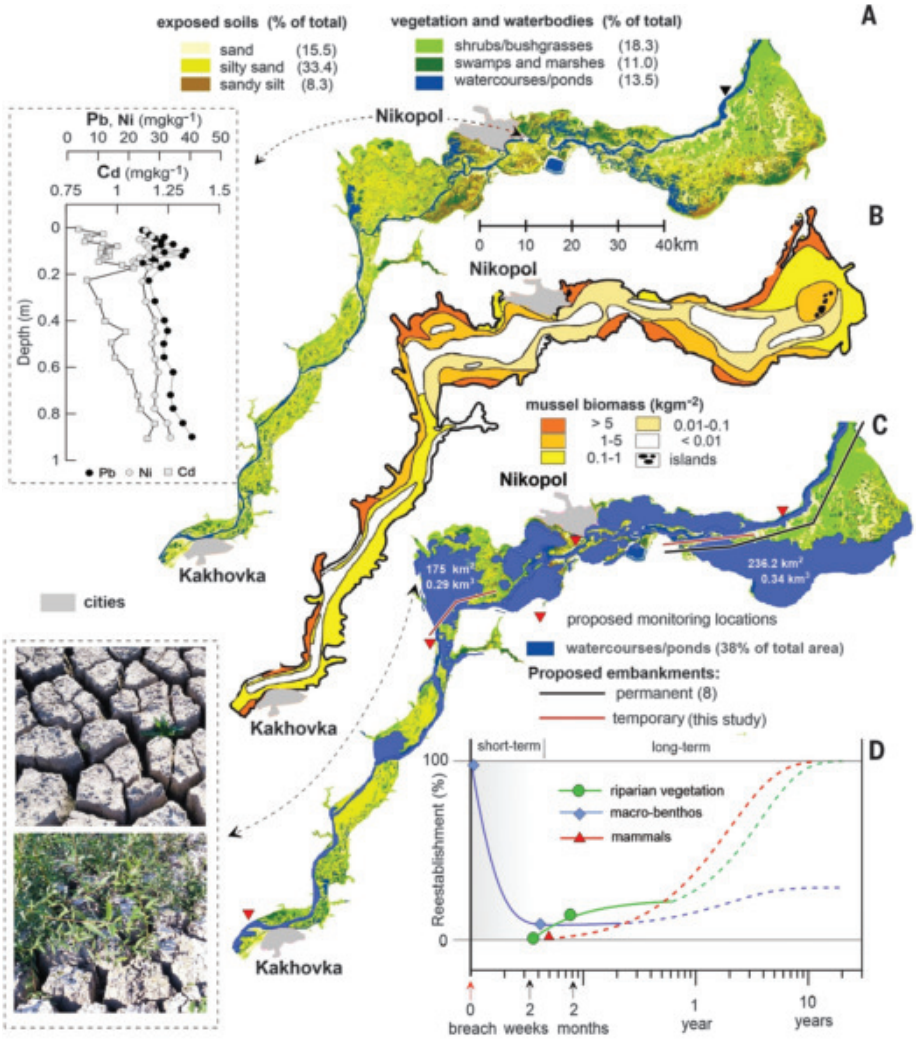


Fig. 4. Development of the area of the former Kakhovka LRF. (**A**) Land cover based on Sentinel-2 satellite images (19 August 2023; resolution 10 m pixel^{-1}; inset shows concentration of heavy metals in the sediment of K-LRF (mg kg^{-1} dry weight) sampled in 2020, dashed arrow line indicates sampling location). (**B**) Spatial pattern of zebra mussels biomass reconstructed from field surveys in the period from 2000 to 2021. (**C**) Extent of inundation of the floodplain by a seasonal flood in March 2024 (inset shows slabs separated by cracks on the former lakebed after drainage and after colonization by plants, above and below, respectively). (**D**) Predicted ecosystem reestablishment on the former lakebed (solid lines indicate observed and dashed lines predicted trends).

of the Kakhovka power plant and flooding of retail gas (petrol) stations released about 450 tonnes of oil products (*7*).

The first week after the dam breach, the reservoir released 9000 to 17,000 tonnes of phytoplankton per day (fig. S18A). Turbidity increased by a factor of nearly 50, concomitant with a rise in toxicity (fig. S18 and tables S1 to S5), resulting in the probable loss of 10,000 tonnes of macroinvertebrates, considering data on standing biomass prior to the breach. Most likely the entire 0+ juvenile fish stock was lost as the flood occurred immediately after spawning. Floodplain grasses were buried by sediment, destroying habitat for amphibians, birds, and mammals.

Drainage of the reservoir exposed 1944 km^2 of the reservoir bottom (Fig. 3A and figs. S21 to S23) and based on this area we estimate that ~80% of the reservoir biomass was lost. The river returned to its historical course, but large ponds appeared on the floodplain (Fig. 3A). As exposed sediment dried out, it formed patterns composed of regular slabs separated by cracks (*40*, *41*) (Fig. 4C, inset).

The river flood affected water quality in the Dnipro-Bug Estuary and in the northern part of the Black Sea (Fig. 3C and table S1 to S5). The river plume, visible from high levels of turbidity in remote sensed imagery, reached the Danube Delta on 17 June 2023, covering ~7300 km^2. Because of restricted vertical mixing, the freshwater formed a layer 10 m thick over underlying saline water. At Odesa, salinity decreased from 11 gL^{-1} to 4.2 gL^{-1} (Fig. 3C), concomitant with five- to tenfold increases in concentrations of nitrogen and phosphorus (table S3). Low salinity and high toxicity caused declines in mussel populations at some locations by as much as 50%. Debris from dislodged plants and entrained rubbish was dispersed along 250 km of coastline. The total biomass of plants transported by the flood amounted to 1600 tonnes with detrital material stranded on beaches at a density of 0.1 kgm^{-2}.

Long-term pollution threat and ecosystem reestablishment trends

The greatest effects of the breach occurred in the former reservoir where the water-surface elevation decreased by 86.4%. Full recovery of the reservoir ecosystem requires rebuilding of the dam (*7*), an unlikely scenario given the ongoing conflict. Therefore, environmental threats and ecosystem responses must be framed within the context of extant conditions. The breach of the dam has produced two main long-term threats: (i) exposure of large areas of contaminated sediment, and (ii) a shortage of water for irrigation. Here we discuss the first threat, while details on the second are provided in (*7*).

Our results show that draining of the reservoir exposed lakebed sediment cumulatively containing around 83.3 thousand tonnes of highly toxic heavy metals (Pb, Cd, Ni) (Fig. 4A, and tables S6 and S8), of which <1% was likely released during the drainage. Remaining heavy metals can affect human populations in the region because river water is widely used by households to compensate for shortages in municipal water supply (*2*, *3*, *42*). Heavy metals in bottom sediment enter the river through erosion of sediment by surface runoff and by seasonal floods. In August 2023, surface runoff rapidly increased levels of turbidity and nutrients in water sampled at Kherson (fig. S18). In March and May 2024, spring floods inundated up to 888.5 km^2 (Fig. 4C), increasing contaminant concentrations (tables S4 and S5).

Heavy metal pollution can be mitigated by bioremediation methods, including biostimulation of uptake by microorganisms (*43*) and phytoremediation (*44*). For the K-LRF, phytoremediation may be the only available strategy over short timescales (2 to 10 years). We analyze the dynamics of riparian vegetation establishment by combining conceptual models of dam removal (*28*, *45*), remote sensing data, and ground observations to gain a comprehensive understanding of self-organizing ecosystem processes driven by flow-soil-plant interactions (*19*, *40*). By mid-August 2023, pioneer riparian vegetation had become established over 18% of the new floodplain area (Fig. 4A). Although rapid colonization of bare ground is usually attributed to dispersal by wind (*41*), our analysis

indicates that wind contributed only 25% to recolonization (*46*) and water played the major role (75%), amplified by self-organization processes in soils (*40*) (figs. S24 and S25). Before breaching, seeds of willows and poplars were dispersed on the water surface by wind over distances of ~200 to 300 m from shorelines (*47*), and were deposited at margins of the ebbing water over distances of 3 to 5 km during drainage of the reservoir (Fig. 3A). Moist cracks that developed in the drying sediment trapped seeds (Fig. 4E), ensuring access to wet soil (fig. S24). The next phase of ecosystem self-organization, the emergence of patterns in initially randomly seeded vegetation, involved flow-plant interactions during spring floods (Fig. 4C). Flow is strongest over soil cracks during the rising and falling stages of floods (*40*), leading to erosion that increases the wetted area of the cracks and enhances releases of contaminants. Over time these erosional processes also uproot plants and form erosional channels, thereby increasing the complexity of riparian habitats (*29*).

Using remotely sensed data on ecosystem conditions in the former reservoir after dam breaching (Fig. 3A and Fig. 4, A and B), we parameterized conceptual models of biomass growth for plants and organisms as indicators of ecosystem reestablishment (Fig. 4D). Initial conditions are 0% for terrestrial species whereas values for aquatic species correspond to 100% at breaching and reduce to 13.5% for remaining aquatic areas (Fig. 4A). Upper limits of the timescale for reestablishment of terrestrial species equaled those for fully grown willow trees with an average life span of 30 years (*41*). Our results suggest that reestablishment equivalent to 80% of an undammed ecosystem is expected within five years. Field observations in spring 2024 documented colonization by canids and wild pigs (fig. S26). Restored connectivity among marine, estuarine, and riverine environments supports the expectation (*48*) that biodiversity of the riverine environment will increase within two years (Fig. 4D).

Scales and perspectives

Our work highlights the far-reaching environmental consequences of the K-LRF destruction and raises concerns not only about the use of water as a weapon, but also about risks posed by aging dams around the world. As the war continues, discussions have begun among decision-makers, scientists, and practitioners about the future of the Kakhovka Dam. Opposing opinions exist about whether to rebuild the dam (*8*). Environmentalists argue that the river ecosystem is quickly reestablishing its pre-dam state but neglect threats posed by releases of heavy metals and their accumulation in food webs. Impounding the reservoir might mitigate this problem and promote economic recovery in the region. As a compromise between these two options, scientists have proposed to "build back better" (*8*) by constructing a 50-km-long barrier that would separate a large boggy area in the upper part of the drained reservoir from the lower reservoir (Fig. 4C). Reconstruction of the dam and construction of the barrier would require years of effort during which the release of contaminants would need to be reduced. We suggest that contaminant release can be effectively controlled by constructing two 15-km-long temporary barriers separating the main channel from the two largest areas of bogs (Fig. 4C). Any plans for the recovery of Ukraine's conflict-damaged water ecosystems require the war to end, but considerable risks persist for new missile attacks on dams in the Dnipro and Dniester cascades. If more dams are targeted, the human toll and environmental damage could be cataclysmic as revealed by the collapse of the Kakhovka Dam. Protection of dams in military zones should be a priority concern for international law given the potential of conflict-related breaches to produce large-scale and long-term environmental impacts.

REFERENCES AND NOTES

1. P. H. Gleick, *Water Resour. Manage.* **33**, 1737–1751 (2019).
2. O. O. Shumilova *et al*., *Nat. Sustain.* **6**, 578–586 (2023).
3. P. H. Gleick, V. Vyshnevskyi, S. Shevchuk, *Earth's Future* **11**, e2023EF003910 (2023).
4. E. Stokstad, *Science* **380**, 1099 (2023).
5. M. Naddaf, *Nature* **618**, 440–441 (2023).
6. V. Vyshnevskyi, S. Shevchuk, V. Komorin, Y. Oleynik, P. Gleick, *Water Int.* **48**, 631–647 (2023).
7. B. M. Spears *et al*., *Nat. Ecol. Evol.* **8**, 834–836 (2024).
8. R. Stone, *Science* **383**, 18–23 (2024).
9. Q. Yang *et al*., *Commun. Earth Environ.* **5**, 230 (2024).
10. B. Lehner *et al*., *Front. Ecol. Environ.* **9**, 494–502 (2011).
11. N. M. Rana *et al*., *Earth Sci. Rev.* **232**, 104144 (2022).
12. G. Grant, *Hydrol. Processes* **15**, 1531–1532 (2001).
13. M. W. Doyle *et al*., *Science* **319**, 286–287 (2008).
14. J. Klemm, I. Winkler, "Is the disaster in Lybia coming soon to an aging dam near you?" (The New York Times, 2023); https://www.nytimes.com/2023/09/17/opinion/libya-floods-dams.html
15. A. N. Sukhodolov *et al*., in *Rivers of Europe*, K. Tockner, C. Zarfl, C. Robinson, Eds. (Elsevier, 2022), pp. 687–718.
16. I. Khurshudyan, P. Sonne, S. Morgunov, K. Hrabchuk, "Inside the Ukrainian counter-offensive that shocked Putin and reshaped the war" (The Washington Post, 2022); https://www.washingtonpost.com/world/2022/12/29/ukraine-offensive-kharkiv-kherson-donetsk/
17. M. M. Foley *et al*., *Water Resour. Res.* **53**, 5229–5246 (2017).
18. "Dam failure and flood event case history compilation" (RCEM, U.S. Bureau of Reclamation, 2015).
19. M. Rietkerk, S. C. Dekker, P. C. de Ruiter, J. van de Koppel, *Science* **305**, 1926–1929 (2004).
20. M. Gazzola, M. Argentina, L. Mahadevan, *Nat. Phys.* **10**, 758–761 (2014).
21. A. S. Veskoukis, A. Kyparos, V. Paschalis, M. G. Nikolaidis, *Chin. J. Physiol.* **61**, 144–151 (2018).
22. J. E. O'Connor, J. J. Duda, G. E. Grant, *Science* **348**, 496–497 (2015).
23. L. W. Morton, K. R. Olson, *J. Soil Water Conserv.* **75**, 13A–19A (2020).
24. K. R. Olson, M. Tharp, *J. Soil Water Conserv.* **75**, 33A–37A (2020).
25. T. E. Mirtskhulava, *Erosion of Channels and Methods for Assessing their Stability* (Kolos, 1967).
26. K. R. Olson, L. W. Morton, *Managing Mississippi and Ohio River Landscapes* (Soil and Water Conservation Society, 2016).
27. J. R. Bellmore *et al*., *Bioscience* **69**, 26–39 (2019).
28. R. L. Flitcroft *et al*., *Front. Environ. Sci.* **10**, 1–27 (2022).
29. C. Schwarz *et al*., *Nat. Geosci.* **11**, 672–677 (2018).
30. J. Schotanus *et al*., *J. Appl. Ecol.* **57**, 1958–1968 (2020).
31. A. D. Primachenko, *Phytoplankton and Primary Production of the Dnieper and its Reservoirs* (Naukova Dumka, 1981).
32. A. A. Protasov, S. A. Afanasyev, *Hydrobiol. J.* **26**, 15–23 (1990).
33. S. A. Afanasyev, *Hydrobiol. J.* **59**, 3–16 (2023).
34. V. Osadchy, B. Nabyvanets, P. Linnik, N. Osadcha, J. Nabyvanets, *Processes Determining Surface Water Chemistry* (Springer, 2016).
35. P. B. Tchounwou, C. G. Yedjou, A. K. Patlolla, D. J. Sutton, in *Molecular, Clinical and Environmental Toxicology*. A. Luch, Ed (Springer, 2012); pp. 133–164.
36. H. Chanson, *Coast. Eng. J.* **48**, 355–370 (2018).
37. C. S. Magirl, J. W. Gartner, G. M. Smart, R. H. Webb, *Water Resour. Res.* **45**, W05427 (2009).
38. A. N. Sukhodolov, O. O. Shumilova, G. S. Constantinescu, Q. W. Lewis, B. L. Rhoads, *Nat. Geosci.* **16**, 89–93 (2023).
39. S. W. Kieffer, *Rev. Geophys.* **27**, 3–38 (1989).
40. K. Zhang *et al*., *Sci. Adv.* **9**, eabq3520 (2023).
41. A. Kuzemko *et al*., Reach the bottom: plant cover of the former Kakhovka Reservoir, Ukraine. Research Square rs.3.rs-4137799 [Preprint] (2024); doi:10.21203/rs.3.rs-4137799/v1.
42. M. G. Macklin *et al*., *Science* **381**, 1345–1350 (2023).
43. G. Haferburg, E. Kothe, *J. Basic Microbiol.* **47**, 453–467 (2007).
44. S. A. Bhat *et al*., *Chemosphere* **303**, 134788 (2022).
45. P. B. Shafroth, L. G. Perry, J. M. Helfield, J. Chenoweth, R. L. Brown, *Front. Ecol. Evol.* **12**, 1272921 (2024).
46. R. Nathan *et al*., *Theoretical Ecology* **4**, 113–132 (2011).
47. E. A. Gage, D. J. Cooper, *Can. J. Bot.* **83**, 678–687 (2005).
48. C. Garcia de Leaniz *et al*., *Front. Ecol. Evol.* **11**, 1110413 (2023).
49. O. Shumilova, Supplementary dataset for a research paper "A Farewell to a Dam: ecological impacts of the Kakhovka Reservoir ravaging by warfare in Ukraine" version v2 Zenodo (2024); https://doi.org/10.5281/zenodo.14135939.

ACKNOWLEDGMENTS

This paper is dedicated to the memory of L. Shevtsova. A few days before the first submission of the paper, we received the sad news that L. Shevtsova was mortally wounded in a missile attack on Kyiv. She took an active part in research, whose results were partly used in our paper. We thank the numerous people who helped with the field work. Constructive comments of G.E. Grant and three anonymous reviewers helped to improve the paper. I. Antipenko, N. Yarmola, O. Medvedenko, and A. Karman are thanked for providing photo and video materials. O.S. acknowledges personal support from the Berlin-Brandenburg Academy of Sciences and Humanities. **Funding:** This work was funded by the following: Leibniz Institute of Freshwater Ecology and Inland Fisheries, Programme Area Dimensions of Complexity of Aquatic Systems (to O.S., A.S., H.P.G., M.V., L.M., and C.W.) Leibniz Competition J45/2018 (to S.D.) German Science Foundation, DFG SU405/11 (to A.S.) Berlin-Brandenburg Academy of Sciences and Humanities (BBAW), project "Young Network TransEurope" (to O.S.). **Author contributions:** Conceptualization: O.S. and A.S. Data curation: N.O., V.O., J.N., S.A., O.L., V.K., M.B., and S.S. Formal analysis: O.S., A.S., and H.P.G. Funding acquisition: A.S. and O.S. Investigation: N.O., V.O., J.N., S.A., O.L., G.T., N.M., and V.K. Methodology: O.S., A.S., G.C., S.A., N.O., and V.O. Project administration: L.M. Software: G.C., M.K., and A.O. Validation: A.S., A.O., M.K., and V.L. Visualization: O.S., A.S., A.O., M.V., and N.M. Writing - original draft: O.S., A.S., H.P.G., and B.L.R. Writing - review & editing: O.S., A.S., H.P.G., B.L.R., B.S., J.d.K., P.G., M.M., K.T., S.J., S.D., L.M., and C.W. **Competing interests:** Authors declare that they have no competing interests. **Data and materials availability:** The data used in the final analyses are available in the main text or the supplementary materials. Digital elevation model, results of modeling, observations, and photo and video materials are available at Zenodo (*49*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adn8655
Materials and Methods
Figs. S1 to S26
Tables S1 to S8
References (*50–102*)
MDAR Reproducibility Checklist

Submitted 4 January 2024; resubmitted 5 July 2024
Accepted 24 January 2025
10.1126/science.adn8655

PIEZOELECTRICS

Piezoelectricity in half-Heusler narrow-bandgap semiconductors

Yi Huang[1]†, Fu Lv[1]†, Shen Han[1]†, Mengzhao Chen[1], Yuechu Wang[1], Qianhui Lou[1], Chenguang Fu[1]*, Yuhui Huang[1]*, Di Wu[2], Fei Li[3], Tiejun Zhu[1]*

Piezoelectricity is primarily observed in noncentrosymmetric insulators or wide bandgap semiconductors. We report the observation of the piezoelectric (PE) effect in half-Heusler (HH) narrow-bandgap semiconductors TiNiSn, ZrNiSn, and TiCoSb. These materials exhibit shear PE strain coefficients that reach ~38 and 33 picocoulombs per newton in ZrNiSn and TiCoSb, respectively, which are high values for noncentrosymmetric nonpolar materials. We demonstrated a TiCoSb-based PE sensor with a large voltage response and capable of charging a capacitor. The PE effect in HHs remains thermally stable up to 1173 kelvin, underscoring their potential for high-temperature applications. Our observations suggest that these HH narrow-bandgap semiconductors may find promising applications for advanced multifunctional technologies.

Energy harvesting can enable electronic devices to achieve self sufficiency in power. Piezoelectric (PE) materials generate electrical voltage when mechanically stressed and can be used to make PE power generators that convert vibrational energy into electricity. The PE strain coefficient, d, is an essential parameter for characterizing the performance of PE materials, which reflects the coupling relationship between the mechanical properties and dielectric properties of these materials. Generally, a higher PE coefficient indicates superior PE performance.

Research on PE materials has primarily been focused on ceramics and single crystals (SCs) with a wide bandgap (E_g > 2.0 eV) (*1*) and low electrical conductivity (~10^{-15} S/m) (*2*–*4*). For example, $Pb(Zr,Ti)O_3$ (PZT)–based ceramics are notable for having d values that exceed 1000 pC/N (*5*–*9*). Lead-free ceramics such as $BaTiO_3$ (*10*), $(Bi, Na)TiO_3$ (*11*), $(K, Na)NbO_3$ (*12*), and $BiFeO_3$-based ceramics (*13*), which have PE coefficients of around hundreds of picocoulombs per newton, have gained interest because of the demand for environmentally friendly PE materials. Moreover, developments in producing SCs, such as $Pb(Mg_{1/3}Nb_{2/3})O_3$-$PbTiO_3$ (PMN-PT) and $Pb(Zn_{1/3}Nb_{2/3})O_3$-$PbTiO_3$ (PZN-PT), have yielded enormous PE strain coefficients of around 1000 pC/N (*14*). Nevertheless, PE materials such as PZT-based ceramics and PMN-PT–based SCs generally require an external electric field to induce polarization to obtain the PE response. By contrast, noncentrosymmetric crystalline materials such as quartz (SiO_2), GaSb, and ZnO naturally manifest the PE effect but typically exhibit a lower PE coefficient of around 10 pC/N (*15*–*19*).

By contrast, narrow-bandgap ($E_g \leq 0.5$ eV) (*20*) semiconductors with noncentrosymmetric crystal structures have received very limited attention regarding their PE properties, owing to their generally higher electrical conductivity, which hinders the effective charge accumulation and stable maintenance of the voltage response. One example is rhombohedral α-GeTe with a $R3m$ structure that possesses a ferroelectric phase, demonstrated with piezoresponse force microscopy measurements (*21*). Although use of density functional theory (DFT) predicted a large PE strain coefficient of 170 pC/N (*22*), experimental confirmation is lacking. The failure to record high strain coefficients is probably due to the native cationic vacancies–induced high electrical conductivity (~5×10^5 S/m at 300 K) (*23*).

Half-Heusler (HH) compounds are ternary intermetallics with a stoichiometric formula of XYZ, where X is the most electropositive element, Y is a less electropositive transition-metal element, and Z is a main-group element (*24*, *25*). HH compounds crystallize in a cubic MgAgAs-type structure with space group $F\bar{4}3m$ (no. 216), comprising three interpenetrating ordered face-centered cubic (fcc) sublattices, occupied by the X, Y, and Z atoms, respectively (Fig. 1A). The corresponding occupied Wyckoff positions are 4*a* (0, 0, 0), 4*c* (1/4, 1/4, 1/4), and 4*b* (1/2, 1/2, 1/2), leaving 4*d* (3/4, 3/4, 3/4) vacant. HH compounds conforming to the 18-valence electron rule are categorized as narrow-bandgap semiconductors (*24*, *26*–*28*), which are distinguished by their excellent mechanical properties (*29*) and hold substantial potential for multifunctional applications in various fields, including semiconductors, thermoelectrics (*25*, *30*), half-metallic ferromagnetism (*31*), superconductivity (*32*), topological insulators (*33*), shape memory effects (*34*), along with others (*24*). These compounds belong to the noncentrosymmetric nonpolar $F\bar{4}3m$ space group (*15*). Generally, crystals within the $F43m$ space group permit one nonzero PE response that is described as a shear PE strain coefficient, d_{14} (the description of the PE coefficient that we used is provided in table S1). The d_{14} can be derived from the

Table 1. Experimental and theoretical PE properties of HH compounds. The experimental (Expt.) and theoretical (Cal.) lattice parameter a; bandgap E_g; electrical conductivity σ; elastic moduli C^*_{eff}, C_{44}, and $C'^{[111]}$; shear PE stress coefficient e_{14}; PE strain coefficients d_{14} and $d'^{[111]}_{33}$; and averaged longitudinal PE strain coefficient $\bar{d}_{eff}$ of TiNiSn, ZrNiSn, and TiCoSb compounds. The experimental values were recorded at room temperature. Dashes indicate no data.

		TiNiSn		ZrNiSn		TiCoSb	
		Expt.	Cal.	Expt.	Cal.	Expt.	Cal.
a (Å)		5.9182(1)	5.9445	6.0986(1)	6.1462	5.8787(1)	5.8866
E_g (eV)		–	0.43	–	0.48	–	1.10
σ (S/m)		1.3×10^4	–	1.1×10^4	–	0.1×10^4	–
C^*_{eff} (GPa)		255.1(1)	–	210.7(1)	–	273.5(1)	–
$C'^{[111]}$ (GPa)		–	238.5	–	221.8	–	261.5
C_{44} (GPa)		–	84.2	–	74.8	–	88.7
e_{14} (C/m^2)		–	0.84	–	0.11	–	1.16
d_{14} (pC/N)		8	10.0	38	1.5	33	13.1
$d'^{[111]}_{33}$ (pC/N)		–	5.8	–	0.9	–	7.6
$\bar{d}_{eff}$ (pC/N)	Upperside	5 ± 2	–	20 ± 2	–	19 ± 1	–
	Underside	–5 ± 2	–	–23 ± 4	–	–19 ± 3	–
	Both sides	5 ± 2	–	22 ± 3	–	19 ± 2	–

[1]State Key Laboratory of Silicon and Advanced Semiconductor Materials, School of Materials Science and Engineering, Zhejiang University, Hangzhou, P. R. China. [2]National Laboratory of Solid State Microstructures, College of Engineering and Applied Sciences and Collaborative Innovation Center of Advanced Microstructures, Nanjing University, Nanjing, P. R. China. [3]Electronic Materials Research Lab, Key Lab of Education Ministry, International Center for Dielectric Research, School of Electronic and Information Engineering, State Key Laboratory for Mechanical Behavior of Materials, Xi'an Jiaotong University, Xi'an, P. R. China.
*Corresponding author. Email: zhutj@zju.edu.cn (T.Z.); huangyuhui@zju.edu.cn (Yu.H.); chenguang_fu@zju.edu.cn (C.F.) †These authors contributed equally to this work.

perpendicular PE strain coefficient measured in the [111]-direction, $d'^{[111]}_{33}$. Over the past decade, however, the PE effect in HH compounds has remained only in the theoretical prediction stage (*35*–*37*). DFT calculations predict that the VFeSb compound has a promising shear PE strain coefficient of about 100 pC/N, which is comparable with that of $BaTiO_3$-based ceramics (*10*). The PE coefficients of TiNiSn and ZrNiSn compounds are also predicted to be on the same order of magnitude as SiO_2 (*19*). However, a large research gap exists in the experimental observation of the PE effect in HH compounds. HHs exhibit much greater electrical conductivity (~10^4 S/m at room temperature) (*26*–*28*) as compared with that of traditional PE materials (~10^{-15} S/m of PZT at room temperature) (*38*), posing substantial challenges for observing the PE response. Experimental reports on the PE measurements of HH compounds appear to be absent in the existing literature, underscoring a critical area in research.

We present experimental observations of the PE effect in HH TiNiSn, ZrNiSn, and TiCoSb compounds. Our findings reveal that SCs ZrNiSn and TiCoSb exhibit a large PE response, with a PE coefficient of ~35 pC/N. Furthermore, we have confirmed the thermal stability of the PE effect in these compounds up to a high temperature of 1173 K. We have demonstrated a PE sensor using a TiCoSb SC without induced polarization, which exhibits the prospect of HHs in practical PE application. Our observations position these HH narrow-bandgap semiconductors with the noncentrosymmetric nonpolar structures as attractive candidates for advanced PE materials. The successful observation of PE effect in HH compounds establishes a foundation for their broader use in PEs and promotes the exploration of different PE materials. Combined with their narrow-bandgap semiconductor characteristics, this synergy could drive the development of multifunctional devices.

Experimental PE coefficient of HHs

We investigated the PE effect of polycrystalline (PC) TiNiSn, ZrNiSn, and TiCoSb. We tested the longitudinal PE strain coefficient, d_{eff}, using the quasi-static method (fig. S1A). Although these had high electrical conductivity σ (1.8×10^4, 1.6×10^4, and 0.7×10^3 S/m at 300 K for TiNiSn, ZrNiSn, and TiCoSb, respectively), all three compounds exhibited observable PE responses with the d_{eff} values ranging around 5 to 13 pC/N (fig. S1B). This PE effect motivated

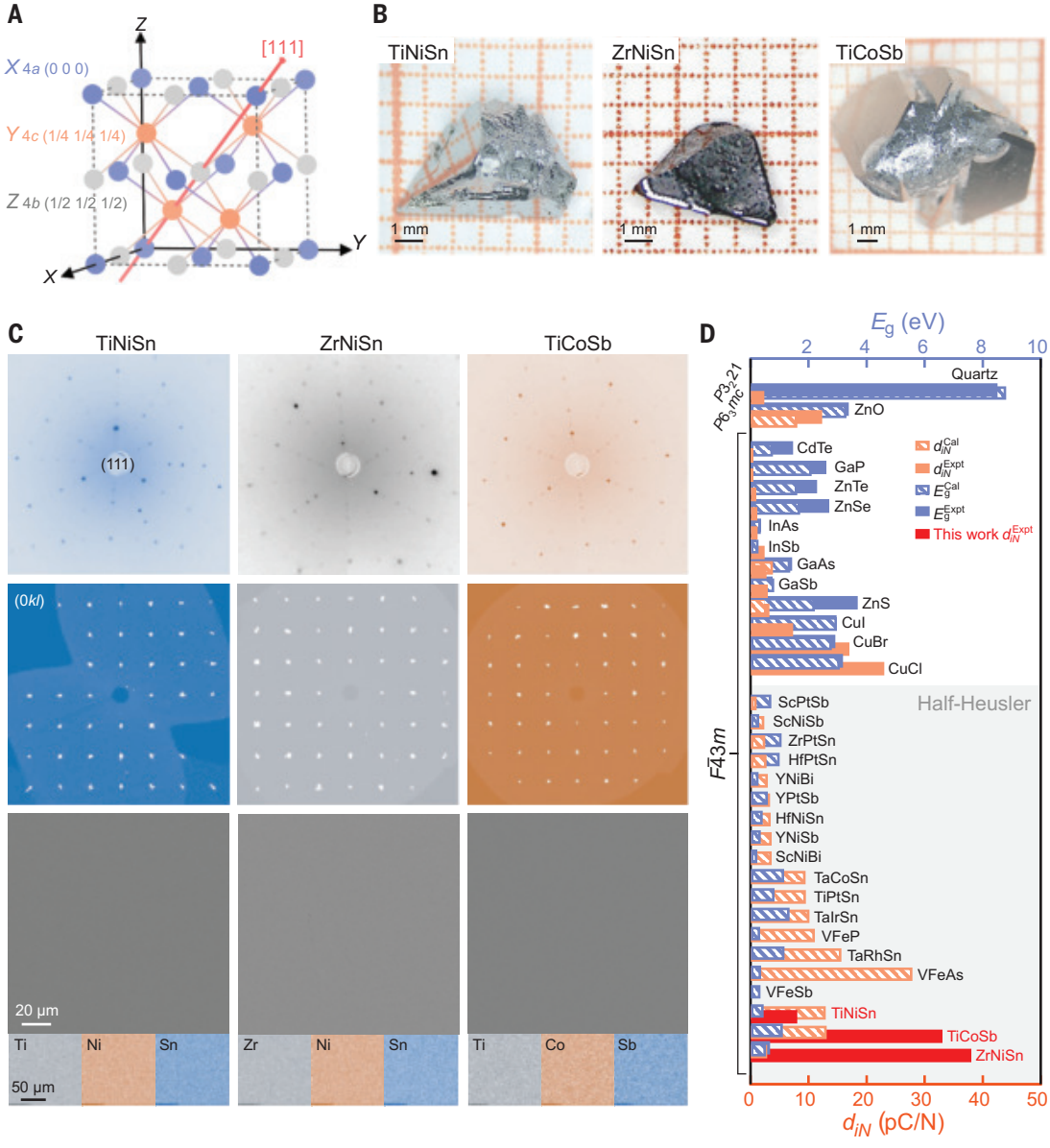


Fig. 1. Characterization and comparison of PE properties for HH SCs. (**A**) Illustration for the noncentrosymmetric nonpolar crystal structure of HH *XYZ* compounds. The [111] direction, characterized by one of the threefold rotational C_3 symmetry operations in the noncentrosymmetric nonpolar $F\bar{4}3m$ space group, serves as the basis for the intrinsic PE effect observed in HH compounds. (**B**) Photographs of the as-grown single-crystalline TiNiSn, ZrNiSn, and TiCoSb. (**C**) (Top) Laue diffraction pattern for a shining trapezoidal face of the grown single-crystal TiNiSn, ZrNiSn, and TiCoSb. (Top middle) Precession diffraction patterns for TiNiSn, ZrNiSn, and TiCoSb samples along the [0*kl*] direction. (Bottom middle) Back-scattered electron images confirm the compositional uniformity of the samples. (Bottom) The elemental distribution obtained from energy-dispersive x-ray spectroscopy analysis. (**D**) Summary of PE coefficient d_{iN} values at room temperature obtained in the currently studied HH crystals and other compounds with $F\bar{4}3m$, $P6_3mc$, and $P3_221$ space groups in references (detailed data are listed in table S2).

us to carry out further investigations on HH SCs, which could help in understanding the intrinsic coupling between mechanical response and dielectric properties because the response might be maximized when used with SCs of specific orientations. However, acquiring a large SC for ternary alloy compounds is challenging.

We grew high-quality single-crystal TiNiSn, ZrNiSn, and TiCoSb using the self-flux method (*26*–*28*). Our as-grown single-crystal TiNiSn, ZrNiSn, and TiCoSb crystals have dimensions of several millimeters. The exposed shiny surfaces of the as-grown crystals can be observed in the photographs (Fig. 1B), displaying the characteristic pyramid shape of HH crystals with preferential growth along the [111] direction. The Laue x-ray diffraction (XRD) patterns (Fig. 1C) reveal distinct spots, confirming the single-crystalline nature of the as-grown crystals. The trapezoid and equilateral triangle shiny surfaces correspond to plates perpendicular to the growth direction, aligned along the [111] direction within the cubic $F\bar{4}3m$ space group. This observation is consistent with the simulated pattern, demonstrating the good quality and crystallinity of the as-grown single-crystal TiNiSn, ZrNiSn, and TiCoSb.

We measured the single crystal XRD (SC-XRD) scattering intensities in the [*0kl*] (Fig. 1C), [*h0l*] (fig. S2A), and [*hk*0] directions (fig. S2B) of reciprocal space confirmed the crystallinity of the TiNiSn, ZrNiSn, and TiCoSb samples. We resolved the crystal structures as an ideal HH with a cubic structure (space group $F\bar{4}3m$), in which *X*, *Y*, and *Z* atoms are positioned at the 4*a* (0 0 0), 4*c* (1/4 1/4 1/4), and 4*b* (1/2 1/2 1/2) Wyckoff sites, respectively. The refinement results from our analyses are summarized in table S3. Additionally, powder synchrotron radiation XRD (SR-XRD) patterns (fig. S3) displayed the absence of obvious secondary phases, in addition to only minimal traces of Sn or Sb flux identified in the samples. To mitigate the influence of the Sn and Sb flux residues, we carefully mechanically polished the samples from the central regions. Our Rietveld refinements of the powder SR-XRD patterns confirmed lattice parameters and atomic coordinates similar to those refined from SC-XRD (table S4). Energy-dispersive x-ray (EDX) compositional mapping of the polished sample surfaces demonstrated a nearly uniform elemental distribution (Fig. 1C). Furthermore, electron probe microanalysis (EPMA) verified that the actual composition closely matched the nominal composition for all three crystals (table S5). We did not detect secondary phases in the polished samples according to the EPMA results. Overall, we demonstrated the successful growth of high-quality SC-TiNiSn, ZrNiSn, and TiCoSb for our subsequent measurements of their PE properties.

Initially, experimental PE measurements revealed the shear PE strain coefficients, d_{14}^{Expt}, of single-crystal TiNiSn, ZrNiSn, and TiCoSb to be 8, 38, and 33 pC/N, respectively (Fig. 1D). The PE strain coefficients for ZrNiSn and TiCoSb are very high for materials within the $F\bar{4}3m$ space group, exceeding the PE coefficient of commercial SiO_2 by an order of magnitude. Although the values are lower than those of ferroelectric PZT and PMN-PT materials (Fig. 1D), these narrow-bandgap HH compounds are notable for their potential PE performance, compared with that of other wide-bandgap semiconductors without externally electric field–induced polarization, such as GaSb (2.9 pC/N) (*16*) and SiO_2 (2.3 pC/N) (*19*).

To achieve the PE strain coefficient, we need to prepare single-crystal TiNiSn, ZrNiSn, and TiCoSb [111]-cut plates. On the basis of the noncentrosymmetric structure of the $F\bar{4}3m$ space group, HH crystals possess only one nonzero element in the PE matrix: shear PE strain coefficient d_{14}. The d_{14} signifies that the application of shear stress along the [010] direction on the (001) plane, which is orthogonal to the *Z* axis, leads to polarization along the [100] direction. The position relationship between [111]-cut plate (in a *X″Y″Z″* coordinate system) and raw HH crystal (in a *XYZ* coordinate system) is shown in Fig. 2A. The detailed processing of single-crystal HH [111]-cut plates is illustrated in fig. S4. The single-crystal TiNiSn, ZrNiSn, and TiCoSb [111]-cut plates are shown in Fig. 2B. As demonstrated in Fig. 2C, we assigned the measured values of the PE coefficient as positive or negative on the basis of the upperside and underside of the samples. The sign of the longitudinal PE strain coefficient, d_{eff}^{*}, at the upperside and underside of each crystal sample is opposite, indicating that the [111] lattice plates exhibit opposite charges when subjected to a perpendicular force, as a result of the PE characteristics of the HH crystals. Moreover, we calculated the averaged values of the longitudinal PE strain coefficients measured on the upperside and underside of the crystals (Table 1). The averaged absolute $\bar{d}_{eff}$ values for both sides are 5 ± 2,

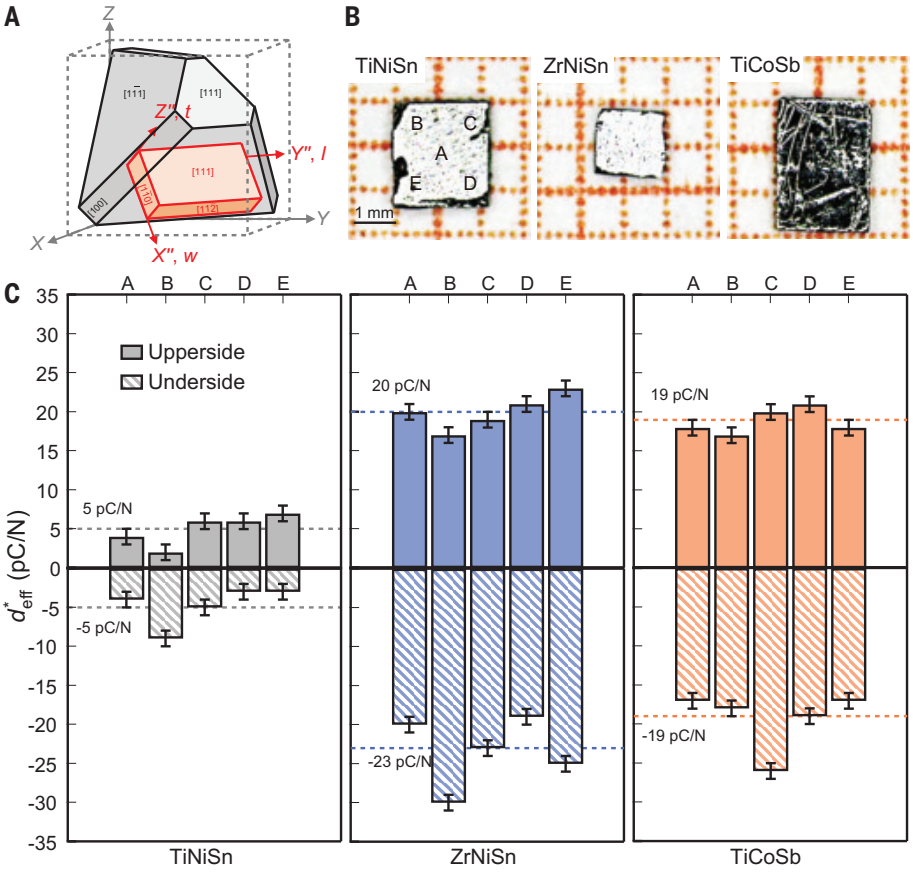


Fig. 2. Preparation and comparison of PE coefficients for the studied HH compounds. (**A**) Schematic illustration of the position relationship between [111]-cut plate (in a *X″Y″Z″* coordinate system) and raw HH crystal (in a *XYZ* coordinate system). (**B**) Photographs display the morphology of TiNiSn, ZrNiSn, and TiCoSb [111]-cut plates. Regions "A" to "E" were selected on both the upperside and underside surfaces of the samples, as indicated on the TiNiSn [111]-cut plate. (**C**) Histograms plot the d_{eff}^{*}, for the A to E regions of TiNiSn, ZrNiSn, and TiCoSb samples. The dashed lines indicate the average longitudinal PE strain coefficient $\bar{d}_{eff}$.

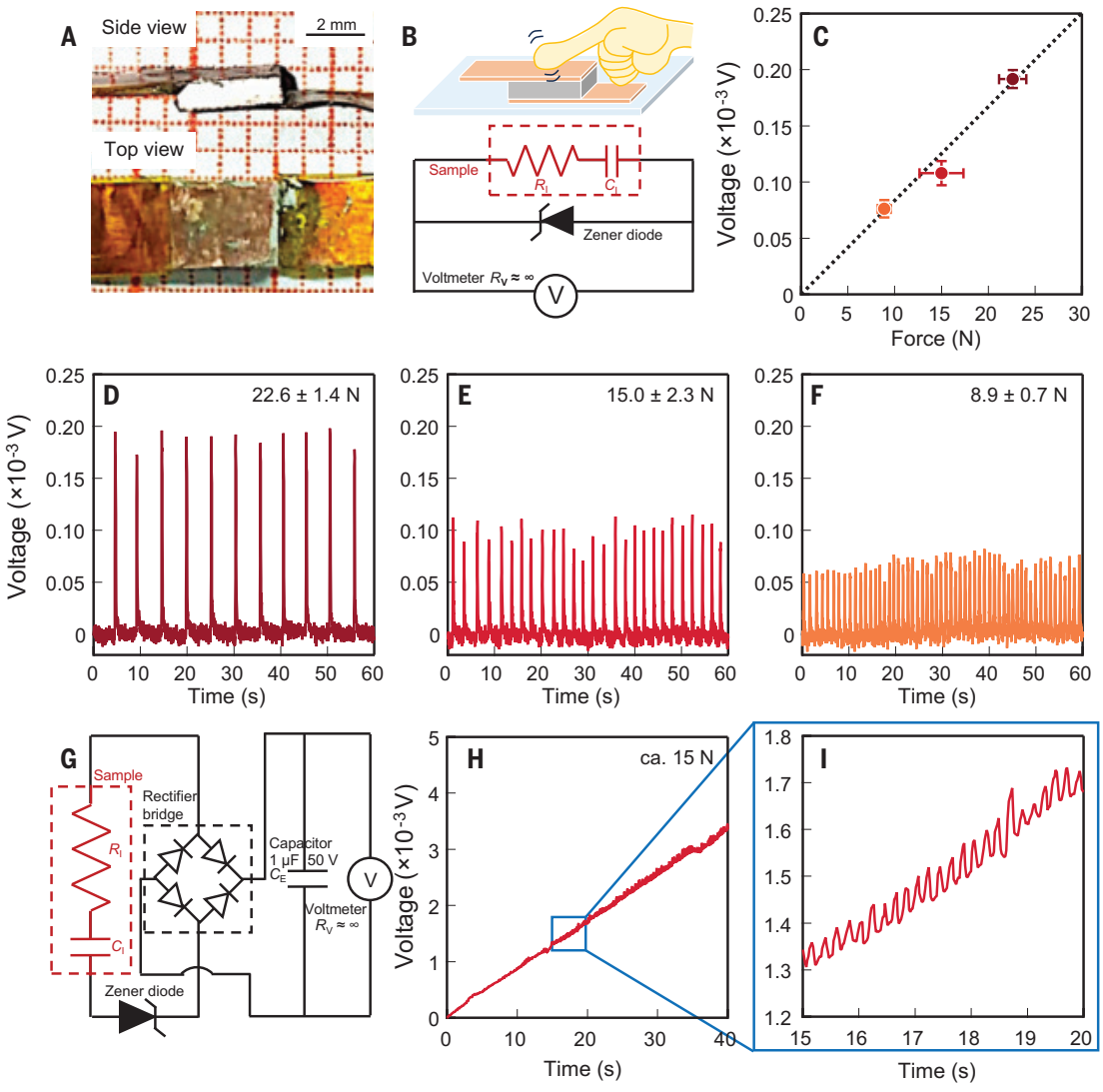


Fig. 3. The pressure sensitivity measurement of the TiCoSb [111]-cut plate. (**A**) Side and top view of the PE sensor demo for the TiCoSb sample. (**B**) Schematic diagrams of the state by finger tapping and the circuit diagram. (**C**) The force-dependent open-circuit voltage. (**D** to **F**) The time-dependent open-circuit voltage with (D) heavy force, (E) medium force, and (F) slight tap. (**G**) The circuit diagram of charging a capacitor with a typical bridge-rectifying circuit. (**H**) Time-dependent voltage in charging the capacitor with 1 μF and (**I**) enlarged plot of the charging curves.

22 ± 3, and 19 ± 2 pC/N for TiNiSn, ZrNiSn, and TiCoSb, respectively, indicating homogeneity in the compositions of the three samples. Among the three samples, the TiNiSn crystal had the smallest value of average longitudinal PE strain coefficient, $\bar{d}_{\text{eff}}$, whereas the ZrNiSn and TiCoSb crystals displayed nearly equal $\bar{d}_{\text{eff}}$ values. By using the relationship between d_{14} and $d'^{[111]}_{33}$ for the HH crystals (*39*), we could experimentally determine the d_{14}^{Expt} values for the single-crystal TiNiSn, ZrNiSn, and TiCoSb crystals: 8, 38, and 33 pC/N, respectively.

Theoretical PE coefficient of HHs

Using refined crystallographic data, we conducted first-principles calculations to estimate the PE coefficients for TiNiSn, ZrNiSn, and TiCoSb compounds. The calculated results—including lattice parameter, bandgap, elastic constant, and PE coefficient of TiNiSn, ZrNiSn, and TiCoSb compounds—are summarized in Table 1. We observed a good agreement between the theoretical and the experimental values of lattice parameters for the TiNiSn, ZrNiSn, and TiCoSb compounds.

Calculations of the strain response yield the elastic constant C_{44} and PE stress coefficient e_{14}, from which d_{14} is determined by using the relation $d_{14} = e_{14}/C_{14}$. The elastic constants for these compounds adhere to Born's criteria to demonstrate their mechanical stability ($C_{44} > 0$, $C_{11} + 2C_{12} > 0$, and $C_{11} - C_{12} > 0$) (*40*). The calculated shear PE strain coefficient d_{14} values of TiNiSn, ZrNiSn, and TiCoSb compounds are 10.0, 1.5, and 13.1 pC/N, respectively. The computational results of TiNiSn and ZrNiSn compounds are in agreement with those of Roy *et al.* (*35*).

Moreover, according to the formula of the coordinate rotation transformation, d_{14} for a [001]-cut plate in single-crystalline HH is $\sqrt{3}$ times the longitudinal PE strain coefficient $d'^{[111]}_{33}$ for an [111]-cut plate. The [001]- and [111]-cut plates are presented as the perpendicular cutting along the [001] and [111] directions of the crystal coordinate axis, respectively. The calculated $d'^{[111]}_{33}$ values of TiNiSn, ZrNiSn, and TiCoSb compounds are 5.8, 0.9, and 7.6 pC/N, respectively. These theoretical results suggest that TiNiSn and TiCoSb compounds have promising electromechanical properties, offering valuable insights for the experimental realization and further exploration of high-performance PE materials.

The experimentally measured d_{14} value is larger than the calculated value for ZrNiSn, whereas they are closer for TiNiSn and TiCoSb. Previous studies on HH compounds have shown that intrinsic point defects introduced by interstitial atoms at 4*d* sites and antisites can effectively alter HH compounds' electrical transport properties (*41*–*44*). Crystallographic defects are known to play a substantial role in the transport properties of HH compounds. For example, the as-prepared ZrNiSn exhibits *n*-type conduction because the defects influence

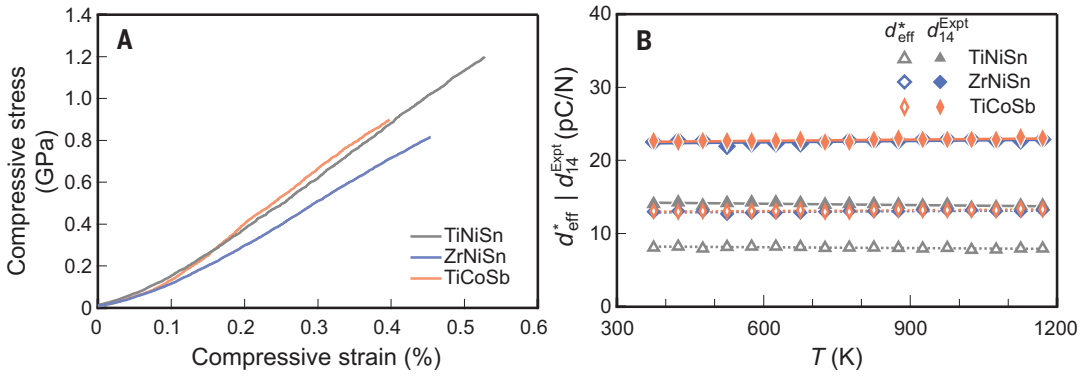


Fig. 4. Mechanical and thermal stability of TiNiSn, ZrNiSn, and TiCoSb compounds. (**A**) Compressive stress-strain curves for TiNiSn, ZrNiSn, and TiCoSb, illustrating their mechanical behavior under compression. (**B**) Temperature-dependent of d^*_{eff} (unfilled makers) and d^{Expt}_{14} (filled makers) for TiNiSn, ZrNiSn, and TiCoSb samples.

the electrical transport. Previous reports suggest that introducing defects into PE materials can form defect dipoles, leading to local lattice distortions and reduced symmetry, which may enhance their PE performance (*45*–*47*). The observed discrepancy between experimentally measured and theoretically calculated PE properties in ZrNiSn and TiCoSb may be attributed to the presence of intrinsic defects within the crystal structure. This highlights the potential of defect engineering in narrow-bandgap semiconductors to substantially enhance their PE properties. Exploring the relationship between crystallographic defects and PE performance is one attractive option for achieving superior PE properties through defect engineering.

Pressure sensitivity measurement

To further explore the potential application of HH materials in the field of piezoelectricity, we subjected a PE sensor consisting of a single-crystal TiCoSb [111]-cut plate. TiCoSb was selected for the pressure sensitivity measurement demonstration because of its relatively larger crystal size (Fig. 1B), compared with that of TiNiSn and ZrNiSn. We coated pure copper electrodes with Epo-tek H20E epoxy adhesive to connect to the single-crystal TiCoSb [111]-cut plate (Fig. 3A) (*39*). We investigated the resulting open-circuit voltage and charge distribution of the PE crystal when subjected to compression along the thickness direction. The pressing process can be divided into three distinct stages: (i) an initial stage with no pressure applied, (ii) an intermediate stage with maximum applied pressure, and (iii) a final stage of released pressure (schematically illustrated in fig. S5). During this process, the open-circuit voltage exhibited a specific behavior. It began at zero during the initial stage without pressure, reached its maximum value under the maximum applied pressure, and then returned to zero as pressure was gradually released.

We demonstrated the open-circuit voltage at room temperature for the single-crystal TiCoSb [111]-cut plates under three different pressure modes: [mode I] heavy pressure, [mode II] medium force, and [mode III] slight tapping without external resistors. The relationship between the open-circuit voltage and the applied pressure is illustrated in Fig. 3C. A fitted dashed line in Fig. 3C shows that the open-circuit voltage increases linearly with increasing pressure. As shown in Fig. 3, D to F, the open-circuit voltage decreased from 0.2×10^{-3} to about 0.05×10^{-3} V, with a decreasing perpendicular force from 22.6 ± 1.4 to 8.9 ± 0.7 N. For heavy and slow tapping, the obtained voltage was about 0.20×10^{-3} V (Fig. 3D). With medium force and patting, the obtained voltage was 0.10×10^{-3} V (Fig. 3E). For slight and fast tapping, a voltage response of about 0.05×10^{-3} V was achieved (Fig. 3F). Moreover, the output voltage of the TiCoSb [111]-cut plate at a force of 30 N was nearly unchanged for different loading durations of 10, 20, 30, and 60 s (fig. S6A). Under sustained pressure, the output voltage stabilized at a fixed value and remained unchanged as the duration increased. Upon removal of the pressure load, the voltage instantaneously dropped to zero. When the pressure load was fixed while the interval changes, the output voltage of the sample remained stable (fig. S6B). Furthermore, the output voltage increased with the increased applied load (fig. S6C).

Furthermore, we devised an experimental setup to assess the charging capability of the [111]-cut plate of TiCoSb crystal. The schematic illustration of the charging process is depicted in Fig. 3G, in which the sample is connected to a 1-μF, 50-V capacitor, arranged in series with a Zener diode and a standard bridge rectifying circuit. The charging profile of the capacitor subjected to continuous finger tapping and a voltage of up to 2.55 mV can be reached within 30 s (Fig. 3H). A zoomed-in view of the charging process within an interval of 5 s is provided in Fig. 3I, with the vibrating energy-harvesting circuit being connected in a continuum by a rectifier bridge. These outcomes underscore that TiCoSb, as a representative of narrow-bandgap HH semiconductors, can serve as a viable material for PE sensors.

Temperature-dependent PE coefficient

High-temperature PE materials have potential applications in aerospace, nuclear energy, automobile manufacturing, energy exploration, and high-temperature monitoring, which need to operate at low frequencies (10^2 to 10^3 Hz range) and high frequencies (10^6 to 10^9 Hz range) based on bulk and surface acoustic waves (*48*, *49*). PE materials such as GaN nitrides (*17*, *18*), (Al, Sc)N–based nitrides (*50*–*53*), and high-entropy oxides (*54*–*56*) exhibit high thermal stability under elevated temperatures, making them suitable for high-temperature PE applications. Moreover, efforts to advance these materials must address ongoing challenges, including undesired leakage current and reduced sensitivity (*57*). On one hand, a major limitation in ferroelectric-type PE materials is the occurrence of phase transition, which generally leads to the instability of PE properties with increasing temperature (*48*, *49*). On the other hand, the bipolar effect–induced rapid increase in the electrical conductivity of PE materials might have a negative effect on the sensitivity (*57*). PC-HH materials show great promise for high-temperature thermoelectric conversion applications (*25*, *30*) but may also have potential for high-temperature PE applications.

To evaluate the thermal stability of HH compounds at high temperature, we conducted differential scanning calorimetry (DSC) measurements from 324 to 1173 K for TiNiSn, ZrNiSn, and TiCoSb crystals (fig. S7). A small endothermic peak at 506.3(1) K in the TiNiSn and ZrNiSn (where the number in parentheses is the margin of error) resulting from the residual Sn flux appears in the chosen crystals. Within the tested temperature range, the DSC heating and cooling curves indicate that no obvious phase transitions or decomposition occur in these three crystals, demonstrating the excellent high-temperature thermal stability of the HH materials.

For further analysis of mechanical and temperature-dependent PE properties, we measured the elastic moduli and high-temperature longitudinal PE strain coefficients d^*_{eff} of the TiNiSn, ZrNiSn, and TiCoSb crystals. The elastic moduli were determined by measuring the room-temperature stress-strain curves during compression using dynamic mechanical analysis (DMA) (Fig. 4A). The resulting elastic moduli, C^*_{eff}, for single-crystal TiNiSn, ZrNiSn, and TiCoSb [111]-cut plates we obtained were 255.1(1), 210.7(1), and 273.5(1) GPa, respectively (Table 1). On the basis of the matrix transformation relationship of elastic constant for [001]-cut plate and [111]-cut plate in HH crystal (*39*), the calculated value $C'^{[111]}$ is comparable with the experimentally obtained value C^*_{eff}. These values, compared with those in Rogl *et al.* (171.6 and 187.7 GPa for the PC-TiNiSn and ZrNiSn, respectively) (*29*), indicate that single-crystal HHs also exhibit good mechanical properties.

Moreover, we further measured the temperature-dependent d^*_{eff} for the TiNiSn, ZrNiSn, and TiCoSb crystals under an argon atmosphere. As shown in Fig. 4B, the d_{14}^{Expt} values of all three samples remained stable from 300 to 1173 K. The average d_{14}^{Expt} values over this entire temperature range were found to be 14, 23, and 23 pC/N for the TiNiSn, ZrNiSn, and TiCoSb samples, respectively. This temperature-independent PE performance underscores the potential suitability of HH compounds for high-temperature PE applications.

To better understand the relationship between the electrical conductivity and PE performance of HH materials, we further measured the temperature-dependent electrical conductivity of TiNiSn, ZrNiSn, and TiCoSb crystals (fig. S8). The electrical conductivity of TiNiSn and ZrNiSn increases with temperature, showing a semiconducting behavior. For TiCoSb, the electrical conductivity even decreases with increasing temperature and shows metallic behavior. Overall, the bipolar effect–contributed electrical conductivity of the studied HH crystals shows a much weaker temperature dependence when compared with that of some conventional PE materials that can exhibit an increase of electrical conductivity by five orders of magnitude in this temperature range (*58*, *59*). This could explain why the measured PE coefficients of these HH compounds do not vary obviously with increasing temperature.

Conclusions

We observed the direct PE effect in PC and single-crystalline HH TiNiSn, ZrNiSn, and TiCoSb. The PE strain coefficients of single-crystal ZrNiSn and TiCoSb exhibit high values of approximately 38 and 33 pC/N at room temperature, respectively, which is notable for materials within the noncentrosymmetric nonpolar structures. The large PE response we observed in the PE sensor, demonstrated by using a single-crystal TiCoSb [111]-cut plate, underscores their potential applications. Additionally, the high-temperature mechanical stability highlights their suitability for use in advanced PE devices and sensors. HH narrow-bandgap semiconductors with noncentrosymmetric cubic structures position these crystals toward the forefront of potential PE materials. The successful experimental observation of PE properties in HH compounds establishes a foundation for their broader utilization in PE applications and facilitates the exploration of new PE materials within the narrow-bandgap semiconductors.

REFERENCES AND NOTES

1. A. Yoshikawa, H. Matsunami, Y. Nanishi, in *Wide Bandgap Semiconductors: Fundamental Properties and Modern Photonic and Electronic Devices*, K. Takahashi, A. Yoshikawa, A. Sandhu, Eds. (Springer, 2007), pp. 1–24.
2. Y. M. You *et al*., *Science* **357**, 306–309 (2017).
3. J. Yang *et al*., *Sci. Adv.* **5**, eaax1782 (2019).
4. L. Qiao *et al*., *Nat. Commun.* **15**, 805 (2024).
5. S. Yang *et al*., *Nat. Commun.* **12**, 1414 (2021).
6. L. Yang *et al*., *Nat. Commun.* **13**, 2444 (2022).
7. J. Li *et al*., *Science* **380**, 87–93 (2023).
8. K. Uchino, *Advanced Piezoelectric Materials: Science and Technology* (Woodhead Publishing, 2017).
9. M. Safaei, H. A. Sodano, S. R. Anton, *Smart Mater. Struct.* **28**, 113001 (2019).
10. H. Fu, R. E. Cohen, *Nature* **403**, 281–283 (2000).
11. X. Liu, X. Tan, *Adv. Mater.* **28**, 574–578 (2016).
12. C. Hu *et al*., *Sci. Adv.* **6**, eaay5979 (2020).
13. B. Narayan *et al*., *Nat. Mater.* **17**, 427–431 (2018).
14. F. Li *et al*., *Science* **364**, 264–268 (2019).
15. T. Okawa, *Applied Physics of Crystals* (Shokabo Co., Ltd., 1976).
16. I. B. Kobiakov, *Solid State Commun.* **35**, 305–310 (1980).
17. S. Muensit, "Piezoelectric coefficients of gallium arsenide, gallium nitride and aluminium nitride," thesis, Macquarie University (1998).
18. I. L. Guy, S. Muensit, E. M. Goldys, *Appl. Phys. Lett.* **75**, 4133–4135 (1999).
19. Y. Saigusa, in *Advanced Piezoelectric Materials*, K. Uchino, Ed. (Woodhead Publishing, ed. 2, 2017), pp. 197–233.
20. J. Chu, *Arden Sher, Physics and Properties of Narrow Gap Semiconductors* (Springer New York, 2008).
21. C. Rinaldi *et al*., *Nano Lett.* **18**, 2751–2758 (2018).
22. R. Shaltaf, E. Durgun, J. Y. Raty, Ph. Ghosez, X. Gonze, *Phys. Rev. B Condens. Matter Mater. Phys.* **78**, 205203 (2008).
23. M. Zhang *et al*., *Adv. Funct. Mater.* **34**, 2307864 (2024).
24. T. Graf, C. Felser, S. S. P. Parkin, *Prog. Solid State Chem.* **39**, 1–50 (2011).
25. K. Xia, C. Hu, C. Fu, X. Zhao, T. Zhu, *Appl. Phys. Lett.* **118**, 140503 (2021).
26. T. Zilber, S. Cohen, D. Fuks, Y. Gelbstein, *J. Alloys Compd.* **781**, 1132–1138 (2019).
27. C. Fu *et al*., *Adv. Sci.* **7**, 1902409 (2020).
28. F. Serrano-Sanchez *et al*., *Nanoscale* **14**, 10067–10074 (2022).
29. G. Rogl *et al*., *Acta Mater.* **107**, 178–195 (2016).
30. R. J. Quinn, J. W. G. Bos, *Mater. Adv.* **2**, 6246–6266 (2021).
31. P. A. Dowben, R. Skomski, *J. Appl. Phys.* **95**, 7453–7458 (2004).
32. Z. K. Liu *et al*., *Nat. Commun.* **7**, 12924 (2016).
33. H. Lin *et al*., *Nat. Mater.* **9**, 546–549 (2010).
34. G. H. Yu *et al*., *Rare Met.* **34**, 527–539 (2015).
35. A. Roy, J. W. Bennett, K. M. Rabe, D. Vanderbilt, *Phys. Rev. Lett.* **109**, 037602 (2012).
36. D. P. Rai *et al*., *J. Phys. Conf. Ser.* **765**, 012005 (2016).
37. A. Harzellaoui, O. Arbouche, K. Amara, *J. Comput. Electron.* **19**, 1365–1372 (2020).
38. B. A. Boukamp, M. T. N. Pham, D. H. A. Blank, H. J. M. Bouwmeester, *Solid State Ion.* **170**, 239–254 (2004).
39. Materials, methods, and supplementary text are available as supplementary materials.
40. B. Karki, G. Ackland, J. Crain, *J. Phys. Condens. Matter* **9**, 8579–8589 (1997).
41. Z. Li *et al*., *Mater. Today Phys.* **34**, 101072 (2023).
42. H. H. Xie *et al*., *CrystEngComm* **14**, 4467 (2012).
43. W. G. Zeier *et al*., *Nat. Rev. Mater.* **1**, 16032 (2016).
44. Y. Huang, K. Hayashi, Y. Miyazaki, *Chem. Mater.* **32**, 5173–5181 (2020).
45. Z. H. Zhao, Y. Dai, F. Huang, *Sustain. Mat. Technol.* **20**, e00092 (2019).
46. G. Huangfu *et al*., *Science* **378**, 1125–1130 (2022).
47. Y. Guan *et al*., *ACS Appl. Mater. Interfaces* **15**, 51421–51428 (2023).
48. D. Damjanovic, *Curr. Opin. Solid State Mater. Sci.* **3**, 469–473 (1998).
49. T. Stevenson *et al*., *J. Mater. Sci. Mater. Electron.* **26**, 9256–9267 (2015).
50. M. Akiyama *et al*., *Adv. Mater.* **21**, 593–596 (2009).
51. S. Fichtner, N. Wolff, F. Lofink, L. Kienle, B. Wagner, *J. Appl. Phys.* **125**, 114103 (2019).
52. M. R. Islam *et al*., *Appl. Phys. Lett.* **118**, 232905 (2021).
53. W. Zhu *et al*., *Appl. Phys. Lett.* **119**, 062901 (2021).
54. M. Zhang *et al*., *Mater. Des.* **200**, 109447 (2021).
55. Y. Sharma *et al*., *ACS Appl. Mater. Interfaces* **14**, 11962–11970 (2022).
56. M. Zhang *et al*., *Acta Mater.* **229**, 117815 (2022).
57. Y. Meng, G. Chen, M. Huang, *Nanomaterials* **12**, 1171 (2022).
58. J. P. Remeika, A. M. Glass, *Mater. Res. Bull.* **5**, 37–45 (1970).
59. S. Zhang, F. Yu, *J. Am. Ceram. Soc.* **94**, 3153–3170 (2011).

ACKNOWLEDGMENTS

The authors gratefully acknowledge the beamline scientists (beamline BL14B1 at the SSRF, Shanghai, China) for their assistance with the powder SR-XRD experiments. The authors acknowledge Y. Pei and Z. Chen from Tongji University for their assistance with the high-temperature electrical conductivity measurement. **Funding:** We acknowledge the support from the National Key Research and Development Program of China (2023YFB3809400), the National Natural Science Foundation (92163203, U23A20553, and 52471239), the Zhejiang Provincial Natural Science Foundation of China (LR25E020005), and the Fundamental Research Funds for the Central Universities (226-2024-00075). **Author contributions:** Conceptualization: Yi H., S.H., C.F., T.Z.; Methodology: Yi H., F. Lv, M.C.; Investigation: Yi H., F. Lv, C.F., Yu.H.; Computation: Yi H.; Data curation: Yi H., F. Lv, S.H., M.C., Y.W., Q.L.; Formal analysis: Yi H., F. Lv, S.H., Y.W., Yu.H., C.F. Funding acquisition: T.Z.; Supervision: T.Z.; Writing – original draft: Yi H.; Writing – review and editing: Yi H., F. Lv, S.H., M.C., C.F., Yu.H, F. Li, D.W., and T.Z. **Competing interests:** Authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ads9584
Materials and Methods
Supplementary Text
Figs. S1 to S10
Tables S1 to S6
References (*60*–*84*)

Submitted 5 September 2024; accepted 24 January 2025
10.1126/science.ads9584

PALEONTOLOGY

Mesozoic mammaliaforms illuminate the origins of pelage coloration

Ruoshuang Li[1], Liliana D'Alba[2,3], Gerben Debruyn[3], Jessica L. Dobson[3], Chang-Fu Zhou[4,5], Julia A. Clarke[6], Jakob Vinther[7], Quanguo Li[1]*, Matthew D. Shawkey[3]*

Pelage coloration, which serves numerous functions, is crucial to the evolution of behavior, physiology, and habitat preferences of mammals. However, little is known about the coloration of Mesozoic mammaliaforms that coevolved with dinosaurs. In this study, we used a dataset of melanosome (melanin-containing organelle) morphology and quantitatively measured hair colors from 116 extant mammals to reliably reconstruct the coloration of six Mesozoic mammaliaforms, including a previously undescribed euharamiyidan. Unlike the highly diverse melanosomes discovered in feathered dinosaurs, hairs in six mammaliaforms of different lineages and diverse ecomorphotypes showed uniform melanosome geometry, corresponding to dark-brown coloration consistent with crypsis and nocturnality. Our results suggest that the melanosome variation and color expansion seen in extant mammals may have occurred during their rapid radiation and diversification after the Cretaceous-Paleogene extinction.

Integumentary coloration, driven by both sexual and natural selection, is prevalent and serves numerous functions in living organisms, including thermoregulation, crypsis, aposematism, intraspecies recognition, mate selection, and communication (*1*–*5*). Color variability differs between vertebrate clades: Whereas tropical birds exhibit vivid and variable plumage coloration, extant mammals are characterized by relatively subdued shades of black, brown, gray, yellow, or red (*4*). Although limited in their color palette, owing in part to the use of only a single pigment type (melanin) and limited iridescence (*6*, *7*), extant mammals have attained distinctive appearances through the arrangement of those colors in discrete patterns, with adaptive significance recognized since the 19th century (*8*). Despite this importance, as well as long-standing interest from scientists and the public, our understanding of the origin and evolutionary history of mammalian coloration has lagged behind that of other groups, such as birds, because limited data exist on the color of extinct mammals known only from fossils, and no such attempt has been made for Mesozoic mammaliaforms (*9*).

Reconstructing the colors of fossil organisms was once thought to be impossible. However, the discovery that melanosomes, the melanin-containing organelles that produce colors in vertebrate integument, can be preserved in fossils as three-dimensional (3D) molds or imprints has revolutionized our understanding of integumentary coloration in deep time (*10*). Reconstruction of the colors of dinosaurs as far back as the Jurassic has elucidated the function and evolution of early feathers (*11*, *12*). The same is not true for mammal pelage, even though exquisite fossils with fur preservation were discovered more than a decade ago (*13*, *14*).

As in other vertebrates, mammalian melanosomes contain two chemical varieties of melanin: eumelanin, producing blacks and browns, and pheomelanin, producing yellows and reds (*15*). Colors of hair are produced by light absorption by a mixture of these two forms and have been associated with the morphology of the melanosomes themselves (*16*). Comparison of fossil and extant melanosomes can therefore enable reconstruction of fossil colors and, when coupled with broader analysis across the specimen, color patterns (*17*). We collected data on melanosome morphology (using scanning electron microscopy) and color (using microspectrophotometry) from hairs of 116 extant mammal species (n = 2615 melanosomes), including representatives of most clades in crown Mammalia, and produced a model that accurately predicts color from morphology. Whereas previous reconstructions have relied on binary, subjective color categories (*17*), we used spectrophotometric data to more precisely and objectively describe the spectrum of colors. We then applied this model to fossilized melanosomes in six phylogenetically and ecologically diverse taxa of the Yanliao and Jehol biotas from northeastern China, thereby reconstructing the pelage coloration of Mesozoic mammaliaforms. A previously undescribed fossil, intensively sampled and with the most complete fur preservation, is diagnosed by its distinctive dental pattern as a new species of Euharamiyida (*18*), adding to the known diversity in the Late Jurassic terrestrial ecosystem.

[1]State Key Laboratory of Biogeology and Environmental Geology, Frontiers Science Center for Deep-time Digital Earth, China University of Geosciences, Beijing, Beijing, China. [2]Evolutionary Ecology Group, Naturalis Biodiversity Center, Leiden, Netherlands. [3]Evolution and Optics of Nanostructures Group, Department of Biology, University of Ghent, Ghent, Belgium. [4]College of Earth Science and Engineering, Shandong University of Science and Technology, Qingdao, China. [5]Paleontological Institute, Shenyang Normal University, Shenyang, China. [6]Jackson School of Geosciences, University of Texas at Austin, Austin, TX, USA. [7]School of Biological Sciences, University of Bristol, Bristol, UK.
*Corresponding author. Email: liquanguo@cugb.edu.cn (Q.L.); matthew.shawkey@ugent.be (M.D.S.)

Systematic paleontology

Mammaliaformes Rowe, 1988 (*19*)
Euharamiyida Bi *et al.*, 2014 (*18*)
Arboroharamiyidae Zheng *et al.*, 2013 (*20*)
Arboroharamiya fuscus sp. nov.

Etymology. *fuscus* (Latin): dark, swarthy, dusky. The specific name *fuscus* refers to the dark coloration of pelage as reconstructed in this research.

Holotype. The holotype (CUGB-P1901, China University of Geosciences, Beijing, China) represents a partial skeleton exposed in ventral view (Fig. 1). The skull is partly preserved, teeth are dislocated and clustered around the skull, distal limbs and most of the phalanges are well preserved, and vertebrae (except caudal vertebrae) and integumentary tissues (including patagium for gilding) are preserved as impressions.

Locality and age. The specimen was collected from the Tiaojishan Formation at Nanshimen village, Gangou Township, Qinglong County, Hebei Province, China. The age of the strata is 158.58 million years within Oxfordian, Late Jurassic (*21*).

Diagnosis. Medium-sized euharamiyidan with a body mass estimated at 156 g. Dental formula presumed as I1-C0-P2-M2/i1-c0-p1-m2 (I, incisor; C, canine; P, premolar; M, molar; capital, upper teeth; lowercase, lower teeth) (Fig. 1). Similar to *Arboroharamiya jenkinsi* (*20*) but differing from other euharamiyidans in having larger teeth, P4 (ultimate upper premolar) being transversely wider than mesiodistally long, p4 (ultimate lower premolar) having simple distal heel with four minor cuspules distally, and molars bearing more cusps. Differs from *A. jenkinsi* in having M1 (first upper molar) subrhomboidal with longer mesial cusped ridge, m1 (first lower molar) narrower and with more cusps on both lingual and buccal rows; differs from *A. allinhopsoni* (*22*) in having four minor cuspules distally on the distal heel of p4 and more cusps on molars; differs from *Xianshou* (*18*) and *Vilevolodon* (*23*) in having larger tooth size and more cusps on cheek teeth. Further differs from *Mirusodens* (*24*), kermackodontids *Kermackodon* (*25*), and *Butlerodon* (*26*) in lacking heart-shaped P4, lacking secondary cusp row between the two primary cusp rows on upper molars, and lacking serrations on the hypertrophic a1 of p4. Further differs from shenshouids *Shenshou* (*18*) and *Qishou* (*27*) in lacking p4 submolariform with two cusp rows on the distal heel and having more cusps on cheek teeth with hypertrophic A1/a1; differs from *Maiopatagium* (*28*) in having enamel ridges and more conical cusps on cheek teeth with central basin closed mesially and distally. Further comparison and description of *A. fuscus* are included in the supplementary materials.

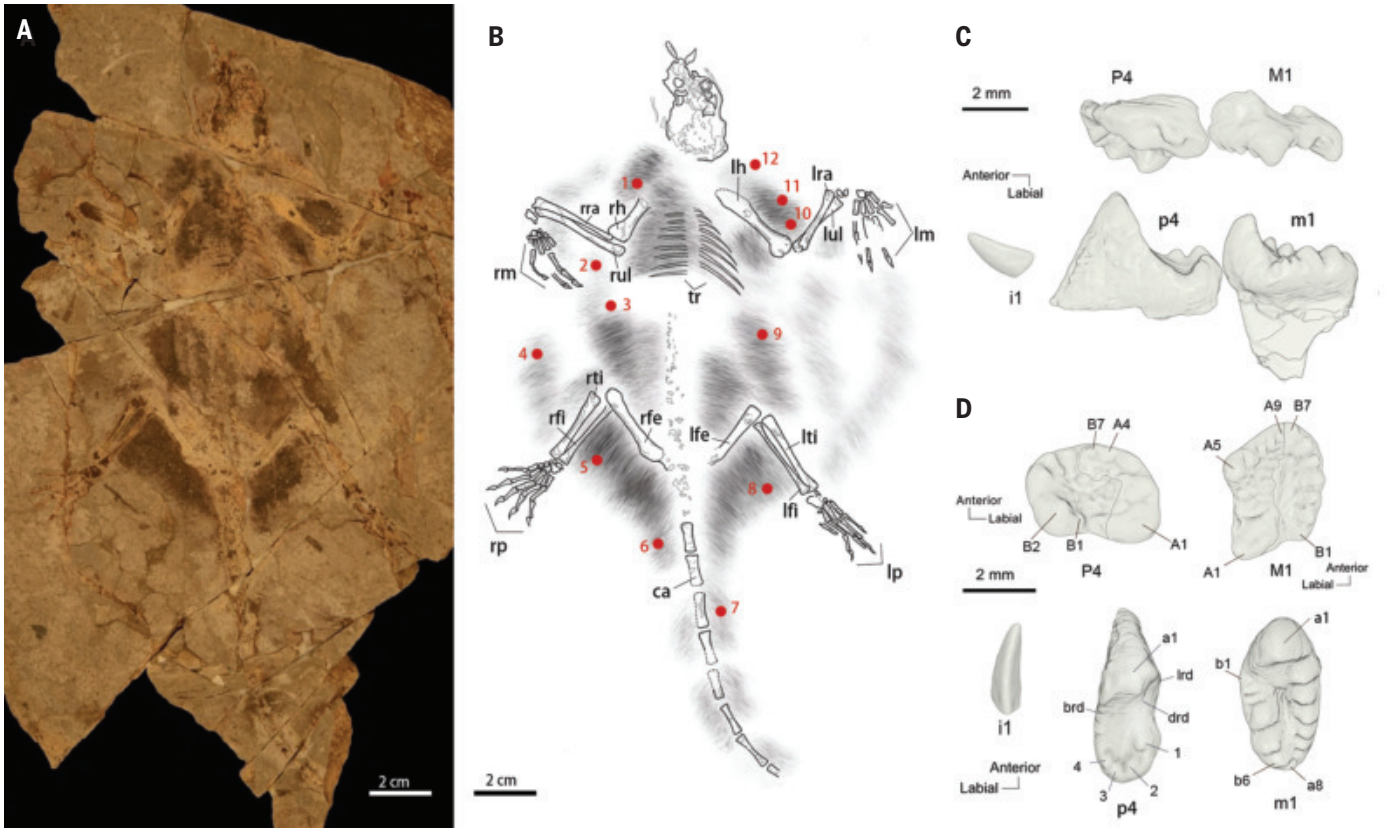


Fig. 1. The holotype specimen and teeth of *A. fuscus*. (**A**) The holotype specimen of *A. fuscus* (CUGB-P1901). (**B**) Line drawing of *A. fuscus*. Red dots represent sample locations for melanosome analyses. ca, caudal vertebrae; lfe, left femur; lfi, left fibula; lh, left humerus; lra, left radius; lm, left manus; lp, left pes; lti, left tibia; lul, left ulna; rfe, right femur; rfi, right fibula; rh, right humerus; rm, right manus; rp, right pes; rra, right radius; rti, right tibia; rul, right ulna; tr, thoracic ribs. (**C**) Upper and lower teeth in buccal view. P4, ultimate upper premolar; M1, first upper molar; i1, lower incisor; p4, ultimate lower premolar; m1, first lower molar. All teeth are from left dentition except M1, which is mirrored from the right first upper molar. (**D**) Occlusal views of teeth. All teeth are from left dentition except M1. The cusp tips on the buccal half of P4 and the lingual half of M1 were incomplete; cusp tips were reconstructed on the basis of preserved dentin and the outline of the tooth crown. brd, buccal ridge; drd, distal ridge; lrd, lingual ridge. Detailed description is included in the supplementary materials.

Integumentary impressions are preserved in the holotype specimen of *A. fuscus*, distributed as deep colored patches around the limb bones, torso, and caudal vertebrae (Fig. 1, A and B). Well-arrayed hairs are distinguishable in the matrix (Fig. 2, D and E, and fig. S2); these are thicker around the body and thinner in the outer area. There is no clear membrane edge preserved, but the extension of fine hair impressions in the area spanning between the wrists and ankles, along with thick hair impressions in the caudal area, suggests the location of the plagiopatagia and uropatagia (Fig. 2, C to E). X-ray fluorescence imaging (XRF) shows Ca and Cu enrichment correlated to hair impressions (Fig. 2C and fig. S4). Further examination under the scanning electron microscope revealed dense fossil melanosomes preserved both in 3D and in impression (Fig. 2, F and G, and fig. S3).

In addition to *A. fuscus*, we analyzed melanosomes of five other Mesozoic fossils preserved with integumentary impressions from the Yanliao and Jehol biotas of northeastern China (Fig. 2A and fig. S1; see supplementary materials, supplementary text section E). Three of the fossils [*A. fuscus*, *Megaconus mammaliaformis* (PMOL-AM00007), and *Vilevolodon diplomylos* (SDUST-V0010)] are placed in "Haramiyida," a taxon including the earliest gliders (*29*). Two fossils (SDUST-V0006 and SDUST-V0007) are assigned to Docodonta, a clade with species assessed to occupy diverse niches (e.g., fossorial, arboreal, or swimming) (*13*, *30*, *31*). The last fossil (SDUST-V0008) from the Lower Cretaceous is eutherian, placed in crown Mammalia. The phalangeal proportions and skeletal characteristics of the six fossils suggest that they belonged to distinct ecomorphotypes (Fig. 2B and fig. S5; supplementary materials, supplementary text section F). *A. fuscus* and *V. diplomylos* have patagia and prehensile claws, indicating an arboreal lifestyle and gliding ability. *M. mammaliaformis* and docodontan SDUST-V0006 are terrestrial and ground-living, whereas docodontan SDUST-V0007 shows heavy molar abrasion and thick bone walls, suggesting a ground-living habitat with fossorial ecology. Eutherian SDUST-V0008 has a pedal phalangeal index close to that of *Eomaia scansoria* (*32*), suggesting a scansorial or arboreal ecology (*32*, *33*). Among analyzed fossils, *M. mammaliaformis* (PMOL-AM00007), from the Middle Jurassic Tiaojishan Formation, is one of the oldest fossils reported with fur impression surrounding its body (*14*). Close examination shows that the fur of *M. mammaliaformis* was differentiated into thick guard hair and thinner underfur (*14*), but four other fossils from the Upper Jurassic (*A. fuscus*, *V. diplomylos*, SDUST-V0006, and SDUST-V0007) show no differences in hair diameter, with all hairs measuring between 0.02 and 0.04 mm (fig. S2). Although the hair impressions in eutherian SDUST-V0008 are scarce, a few of its distinguishable hairs also measure between 0.02 and 0.04 mm (fig. S2). Samples from hair impressions in all fossils show preservation of melanosomes (fig. S3).

With precautions to rule out influences of nonintegumentary tissues and shrinkage during diagenesis (see supplementary materials), we measured the geometry of fossil melanosomes

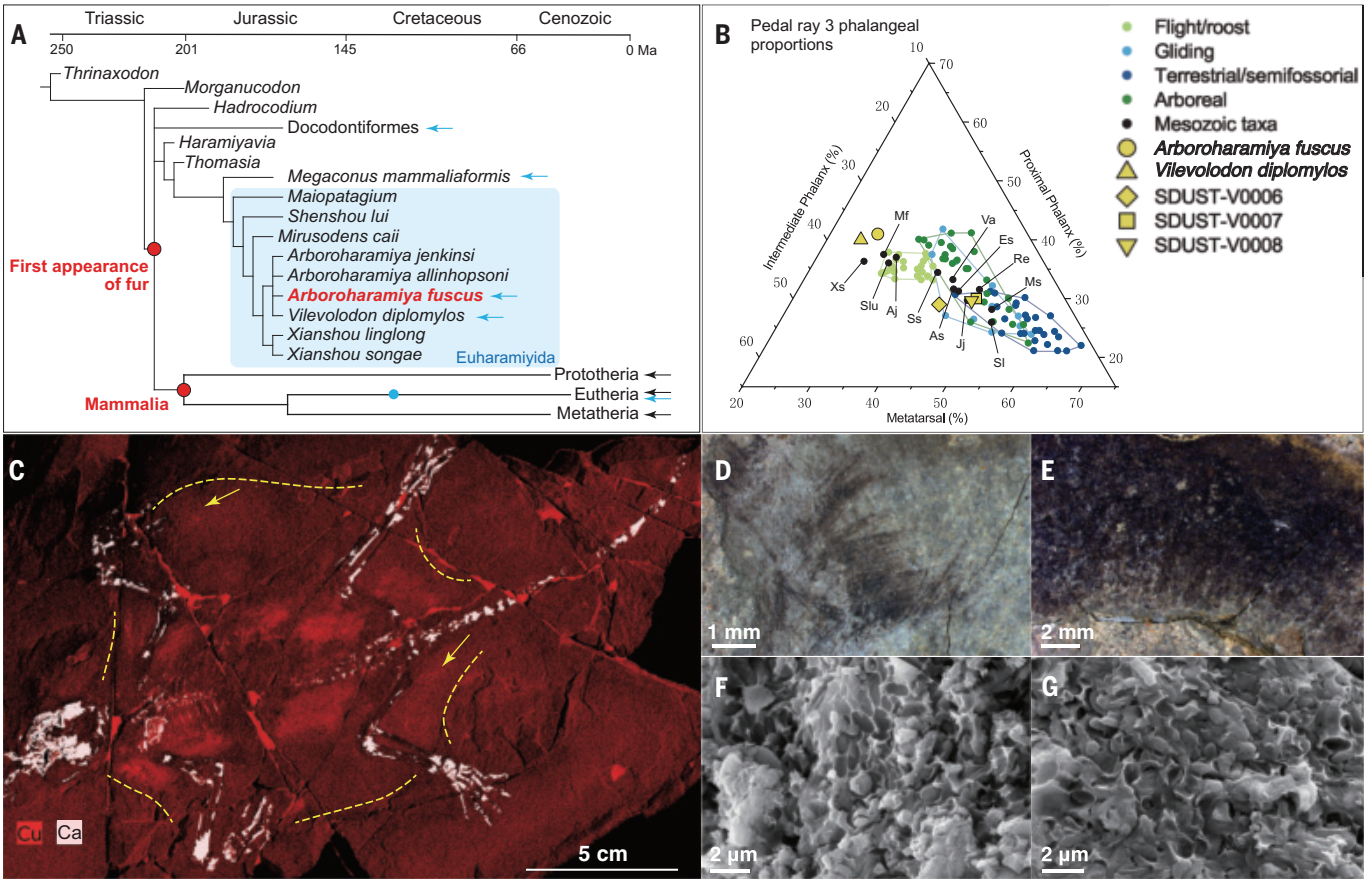


Fig. 2. Phylogenetic position of *A. fuscus*, phalangeal proportion analysis of five fossils, the Ca and Cu distribution, and hair and melanosome preservation in *A. fuscus*. (**A**) Phylogenetic position of *A. fuscus* in Euharamiyida (blue shaded rectangle); large-scale phylogeny tree is adopted from (*50*). Blue arrows, investigated fossils or the clade; black arrows, investigated extant mammalian clades; blue dot, the chronological position of SDUST-V0008. (**B**) Pedal ray 3 phalangeal proportion analysis of five fossils. Aj, *Arboroharamiya jenkinsi*; As, *Agilodocodon scansorius*; Es, *Eomaia scansoria*; Jj, *Jeholodens jenkinsi*; Mf, *Maiopatagium furculiferum*; Ms, *Maotherium sinensis*; Re, *Rugosodon eurasiaticus*; Slu, *Shenshou lui*; Sl, *Sinobaatar lingyuanensis*; Ss, *Sinodelphys szalayi*; Va, *Volaticotherium antiquus*; Xs, *Xianshou songae*. (**C**) Calcium distribution in *A. fuscus* is correlated with skeleton, and copper distribution is correlated with hair impressions. Dashed yellow lines represent the extension of patagium, and the two yellow arrows indicate magnified areas corresponding to (D) and (E). (**D**) Hair impressions preserved in the marginal area of the left plagiopatagia. (**E**) Dense hair impressions preserved in the caudal area. (**F**) Melanosomes in *A. fuscus* preserved as impressions. (**G**) Melanosomes in *A. fuscus* preserved as 3D molds.

in the six fossils, which exhibits high consistency both between and within fossil specimens and falls within the range of extant mammal melanosome measurements (Fig. 3). Extant mammal melanosomes are diverse in size and shape. Length varies within a range of ~1800 nm (346 to 2129 nm), and width varies within ~700 nm (219 to 921 nm). Thus, their aspect ratios range from 1.01 to 4.78 (i.e., almost spherical to highly tubular). By contrast, fossil melanosome geometries are confined to the upper end of extant melanosome size range, with lengths and widths varying by only ~500 and ~400 nm, respectively (Fig. 3), and all approximate an oval shape. Extreme melanosome shapes (either extremely round or elongated) as seen in extant mammals were not recovered from any of the fossils. Together, these data suggest that ancestral hair melanosomes are low-aspect-ratio (length/width < 2) with limited diversity, which is possibly linked with lower basal metabolic rates (fig. S7) (*34*). Our MCMCglmm analysis revealed a direct correlation between melanosome shape and hair color in extant mammals: Colorfulness (as measured by the first principal component of microspectrophotometric data) is negatively correlated with increasing length at a given width ($P < 0.005$) and positively correlated with width ($P < 0.005$) of melanosomes. That is, melanosomes of brightly colored hairs (such as red and orange) are more spherical, whereas melanosomes of dull-colored hairs (black, brown) are more elongated in shape. Furthermore, melanosome density, but not morphology, was significantly associated with the second principal component of color, which differentiates reds from yellows, suggesting higher melanosome densities in redder hairs than in less red hairs ($P = 0.005$). However, the melanosome density of fossil hair is not measurable, owing to the mixing of melanosomes from multiple hairs during fossilization. Surprisingly, no other variable, such as hair thickness or melanosome size variance, predicted coloration as they do in birds (see materials and methods), and microspectrophotometric data show that most individual mammalian hairs are brownish (even those perceived as black) and that true black or gray hairs are rare (fig. S6).

We reconstructed reflectance curves of fossil hairs for the assessed six taxa on the basis of the geometry of their melanosomes. All measurements from all sample locations on all fossils had predicted principal component 1 values corresponding to low reflectance and unsaturated coloration (Fig. 3 and table S5). XRF results of the fossils showed no evidence of Zn associated with pheomelanin (*35*) but high concentration of Cu associated with eumelanin in the integumentary impressions

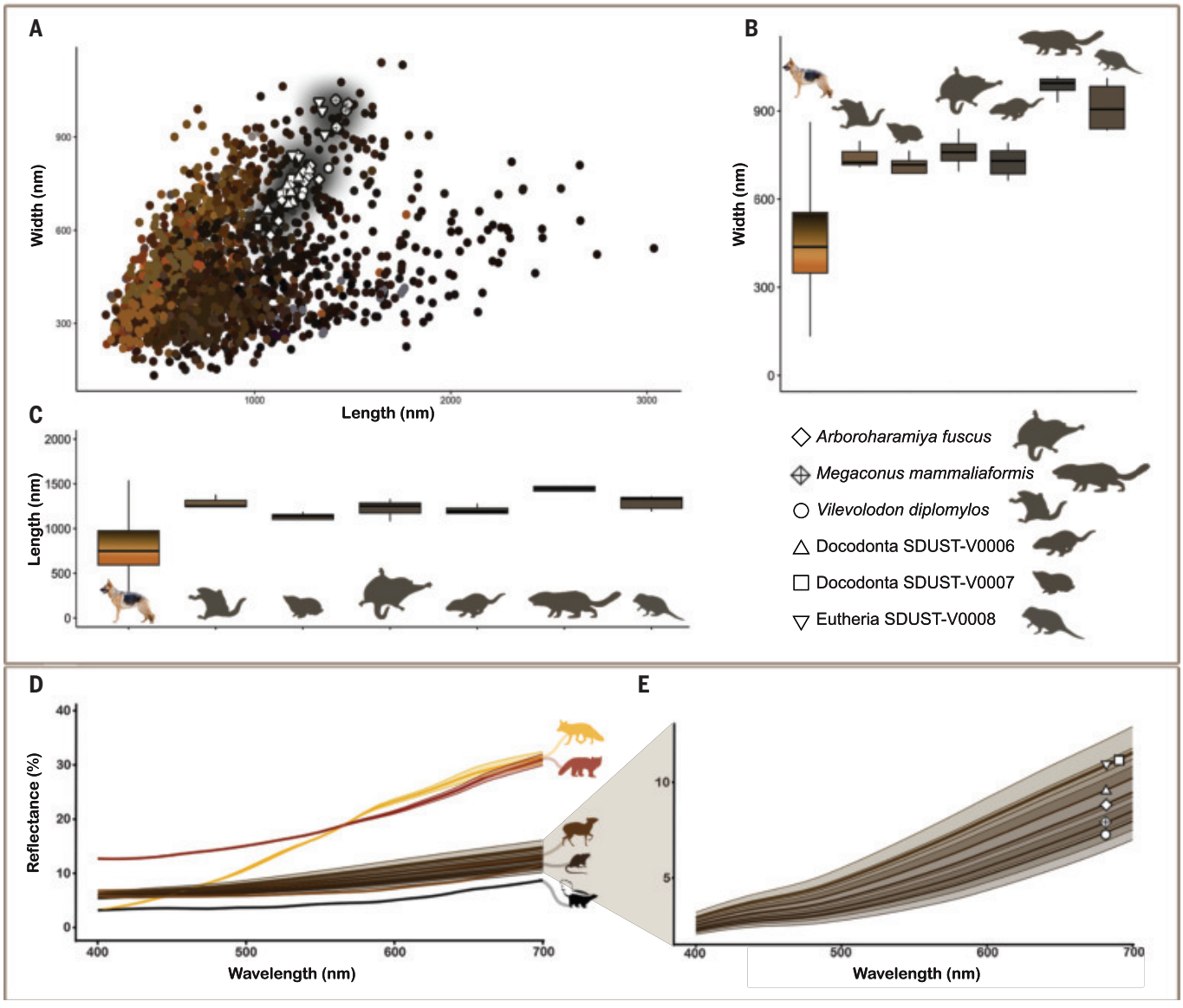


Fig. 3. Melanosome diversity and pelage coloration across extant mammals and Mesozoic mammaliaforms. (**A**) Scatterplot of individual melanosome measurements from 116 extant mammals and six fossils; dot color shows RGB values derived from spectral curves measured at specific hair locations. Extant hair melanosomes, *n* = 2615; *A. fuscus*, *n* = 760; *M. mammaliaformis*, *n* = 103; *V. diplomylos*, *n* = 395; docodontan SDUST-V0006, *n* = 289; docodontan SDUST-V0007, *n* = 407; eutherian SDUST-V0008, *n* = 205. (**B**) Width variation of melanosomes. (**C**) Length variation of melanosomes. (**D**) Measured reflectance spectra of selected extant mammals [*Vulpes vulpes* (orange), *Ailurus fulgens* (red), *Cephalophus dorsalis* (brown), *Pelomys fallax* (darker brown), and *Mephitis mephitis* (black)] and predicted reflectance curves for six Mesozoic mammaliaforms (browns), highlighting their constrained brown coloration and minimal variation. (**E**) Detailed view of the predicted reflectance curves for the six Mesozoic mammaliaforms.

(Fig. 2C and fig. S4) (*36*), supporting a dark coloration dominated by eumelanin. Together, these data indicate that the assessed six Mesozoic mammaliaforms had uniformly dark brown coloration with minor differences (Figs. 3 and 4).

Reconstructing the paleocolors of the six fossils provides physical evidence that pelage coloration in Mesozoic mammaliaforms was limited in diversity and darkly colored. The consistency in melanosome geometry of fossils suggests that, despite distinct variation in phylogeny and ecology, the melanin color system in Mesozoic mammaliaforms was largely invariant. This contrasts with the high diversity of melanosomes in feathered dinosaurs, early birds, and pterosaurs (*34*, *37*, *38*), indicating a different and unique evolutionary pattern of mammaliaform melanin color system after the sauropsid-synapsid split.

Samples taken from different integumentary locations of each fossil resulted in uniformly dark brown color, with no evidence of color patterns such as striping, spots, or countershading as seen in extant mammals, indicating that color patterns on the pelage might be a derived trait. Although the sex of fossils is difficult to determine, the highly confined reconstruction results suggest the absence of sexual dichromatism. The pelage coloration scheme in Mesozoic mammaliaforms was therefore unlikely for recognition or sexual display purposes but mostly for thermoregulation, for enhancing tactical sensory perception (*39*), for abrasion resistance, and for crypsis.

Most small Mesozoic mammaliaforms were likely nocturnal, as suggested by ancestral reconstruction of eye shape of extant mammals

(*40*, *41*) and phylogenetic inference of the genes involved in color perception and phototransduction (*42*). The compensatory sensory functions result from a combination of enhanced olfactory function, enhanced hearing by changes in the inner ear, and tactile sensation from fur (*39*, *43*). A nocturnal lifestyle may have enabled these species to avoid predation by diurnal carnivores and, furthermore, may have helped them to survive mass extinction (*44*). Our reconstruction results support the hypothesis that nocturnality might be ubiquitous in Mesozoic mammaliaforms, as dark-colored, uniformly dull pelage [which is mostly seen in nocturnal extant mammals (*4*)] is present in species occupying very different ecological niches. Additionally, highly melanized hairs could also be advantageous for thermoregulation and mechanical strength (*45*, *46*). Melanization increases heating speeds of materials (*45*), and darker hairs could help small mammaliaforms to reduce heat loss through insulation. Melanized materials are also stronger and abrasion-resistant (*46*), potentially enabling hairs to withstand wear and provide protection for the skin.

CREDIT: C. ZHAO

Fig. 4. Artistic reconstruction of pelage coloration of five Jurassic mammaliaforms. (Top to bottom) *A. fuscus* (CUGB-P1901), *V. diplomylos* (SDUST-V0010), *M. mammaliaformis* (PMOL-AM00007), and docodontans SDUST-V0006 and SDUST-V0007. The green eyes of *A. fuscus* are an artistic license of the tapetum lucidum, an intraocular reflecting structure that enhances visual sensitivity, indicating nocturnality. Eutherian SDUST-V0008 from the Lower Cretaceous is not included in this artistic reconstruction.

Extant mammals that inhabit ecological niches similar to the six fossils, such as fossorial moles (*Scalopus aquaticus*), mice and rats (*Lemniscomys griselda*, *Pelomys fallax*), and nocturnal bats (*Myotis myotis*, *Vespertilio murinus*), show a similar, very limited, color palette with mostly dark gray or brown colorations. However, diurnal or crepuscular species such as the Indian giant squirrel (*Ratufa indica*), red panda (*Ailurus fulgens*), and yellow-throated marten (*Martes flavigula*) exhibit vibrant purples, reds, and yellows, respectively. Daytime activity patterns may have enhanced the significance of sexual selection and intraspecific visual communication for the evolution of color and color patterns in extant groups. Melanosome diversity is also linked to the melanocortin system that affects not only melanin production but also numerous physiological and behavioral traits such as metabolic rate and appetite (*47*). Shifts in this system toward, for example, higher metabolic rates may have also led to the rapid evolution of melanosome shape within crown mammals (*48*). In any case, these changes of color employment through evolution likely developed under different pressures than for camouflage alone. After the Cretaceous-Paleogene event, mammals rapidly diversified into vacant niches previously occupied by dinosaurs (*49*), which could have simultaneously propelled the rapid radiation and diversification of pelage color strategies in new and diverse environments.

REFERENCES AND NOTES

1. I. C. Cuthill *et al.*, *Science* **357**, eaan0221 (2017).
2. D. Stuart-Fox, E. Newton, S. Clusella-Trullas, *Philos. Trans. R. Soc. Lond. Ser. B* **372**, 20160345 (2017).
3. N. Howell *et al.*, *Evolution* **75**, 2480–2493 (2021).
4. T. Caro, *Semin. Cell Dev. Biol.* **24**, 542–552 (2013).
5. C. R. Cooney *et al.*, *Nat. Commun.* **10**, 1773 (2019).
6. A. Roy, M. Pittman, E. T. Saitta, T. G. Kaye, X. Xu, *Biol. Rev. Camb. Philos. Soc.* **95**, 22–50 (2020).
7. H. K. Snyder *et al.*, *Biol. Lett.* **8**, 393–396 (2012).
8. T. Caro, *Trends Ecol. Evol.* **32**, 23–30 (2017).
9. M. E. McNamara *et al.*, *Trends Ecol. Evol.* **36**, 430–443 (2021).
10. J. Vinther, D. E. Briggs, R. O. Prum, V. Saranathan, *Biol. Lett.* **4**, 522–525 (2008).

11. Q. Li *et al*., *Science* **327**, 1369–1372 (2010).
12. D. Hu *et al*., *Nat. Commun.* **9**, 217 (2018).
13. Q. Ji, Z.-X. Luo, C.-X. Yuan, A. R. Tabrum, *Science* **311**, 1123–1127 (2006).
14. C.-F. Zhou, S. Wu, T. Martin, Z.-X. Luo, *Nature* **500**, 163–167 (2013).
15. L. D'Alba, M. D. Shawkey, *Physiol. Rev.* **99**, 1–19 (2019).
16. D. Grabolus, P. Wacławik, M. Zatoń-Dobrowolska, *Can. J. Anim. Sci.* **100**, 418–425 (2020).
17. J. Vinther, *Annu. Rev. Earth Planet. Sci.* **48**, 345–375 (2020).
18. S. Bi, Y. Wang, J. Guan, X. Sheng, J. Meng, *Nature* **514**, 579–584 (2014).
19. T. Rowe, *J. Vertebr. Paleontol.* **8**, 241–264 (1988).
20. X. Zheng, S. Bi, X. Wang, J. Meng, *Nature* **500**, 199–202 (2013).
21. Z. Yu *et al*., *Earth Planet. Sci. Lett.* **617**, 118246 (2023).
22. G. Han, F. Mao, S. Bi, Y. Wang, J. Meng, *Nature* **551**, 451–456 (2017).
23. Z.-X. Luo *et al*., *Nature* **548**, 326–329 (2017).
24. F. Mao, Z. Li, J. J. Hooker, J. Meng, *Zool. J. Linn. Soc.* **199**, 832–859 (2023).
25. P. M. Butler, J. J. Hooker, *Acta Palaeontol. Pol.* **50**, 185–207 (2005).
26. F. Mao, P. Brewer, J. J. Hooker, J. Meng, *J. Syst. Palaeontology* **20**, 1–37 (2022).
27. F.-Y. Mao, J. Meng, *Zool. J. Linn. Soc.* **186**, 529–552 (2019).
28. Q.-J. Meng *et al*., *Nature* **548**, 291–296 (2017).
29. D. M. Grossnickle *et al*., *Evolution* **74**, 2662–2680 (2020).
30. Q.-J. Meng *et al*., *Science* **347**, 764–768 (2015).
31. Z.-X. Luo *et al*., *Science* **347**, 760–764 (2015).
32. Q. Ji *et al*., *Nature* **416**, 816–822 (2002).
33. M. Chen, G. P. Wilson, *Paleobiology* **41**, 280–312 (2015).
34. Q. Li *et al*., *Nature* **507**, 350–353 (2014).
35. P. L. Manning *et al*., *Nat. Commun.* **10**, 2250 (2019).
36. R. A. Wogelius *et al*., *Science* **333**, 1622–1626 (2011).
37. K. K. Nordén *et al*., *Evolution* **73**, 15–27 (2019).
38. A. Cincotta *et al*., *Nature* **604**, 684–688 (2022).
39. T. B. Rowe, T. E. Macrini, Z.-X. Luo, *Science* **332**, 955–957 (2011).
40. M. P. Gerkema, W. I. L. Davies, R. G. Foster, M. Menaker, R. A. Hut, *Proc. Biol. Sci.* **280**, 20130508 (2013).
41. C. P. Heesy, M. I. Hall, *Brain Behav. Evol.* **75**, 195–203 (2010).
42. Y. Wu, H. Wang, E. A. Hadly, *Sci. Rep.* **7**, 46542 (2017).
43. T. B. Rowe, G. M. Shepherd, *J. Comp. Neurol.* **524**, 471–495 (2016).
44. E. Pennisi, *Science* **382**, 865–866 (2023).
45. M. D. Shawkey, A. M. Estes, L. M. Siefferman, G. E. Hill, *Proc. Biol. Sci.* **270**, 1455–1460 (2003).
46. R. H. C. Bonser, *Condor* **97**, 590–591 (1995).
47. A. L. Ducrest, L. Keller, A. Roulin, *Trends Ecol. Evol.* **23**, 502–510 (2008).
48. C. M. Eliason, J. A. Clarke, *Proc. Biol. Sci.* **285**, 20182014 (2018).
49. R. Maor, T. Dayan, H. Ferguson-Gow, K. E. Jones, *Nat. Ecol. Evol.* **1**, 1889–1895 (2017).
50. F. Mao *et al*., *Nature* **628**, 569–575 (2024).

ACKNOWLEDGMENTS

We thank A. Dong and C. Cao for CT scanning of the specimens; J. Yang and M. Yang for XRF scanning of the specimens; D. J. Tang and J. Hu for assisting with scanning electron microscopy analysis of the specimens; and C. Zhao for creating the artistic reconstruction image (Fig. 4). **Funding:** National Natural Science Foundation of China 42161134003 (Q.L., C.-F.Z., R.L.) ; National Natural Science Foundation of China 41872019 (Q.L., R.L.); Fundamental Research Funds for the Central Universities for the Frontiers Science Center for Deep-time Digital Earth, China University of Geosciences (Beijing) 2652023001 (Q.L., R.L.); Chinese "111" project B20011 (Q.L.); Taishan Scholar Program of Shandong Province tstp20240514 (C.-F.Z.); Flemish Research Funds FWO G0E8322N and G0A7921N (M.D.S., L.D.); Human Frontiers Science Program RGP0047 (M.D.S.); Air Force Office of Scientific Research FA9550-18-1-0477, FA9550-23-1-0622, and FA8655-23-2-7041 (M.D.S.). **Author contributions:** Conceptualization: L.D., J.V., J.A.C., Q.L., M.D.S.; Data curation: R.L., L.D., G.D., J.L.D., C.-F.Z., Q.L., M.D.S.; Funding acquisition: L.D., C.-F.Z., Q.L., M.D.S.; Project administration: R.L., G.D., J.L.D.; Supervision: L.D., C.-F.Z., Q.L., M.D.S.; Visualization: R.L., L.D., G.D.; Writing – original draft: R.L., G.D., J.L.D.; Writing – review & editing: R.L., L.D., G.D., J.L.D., C.-F.Z., J.A.C., J.V., Q.L., M.D.S. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The fossil specimen CUGB-P1901 is accessioned at the University of Geosciences, Beijing (CUGB), China. The fossil specimens SDUST-V0006, SDUST-V0007, SDUST-V0008, and SDUST-V0010 are accessioned at Shandong University of Science and Technology (SDUST), Qingdao, China. The fossil specimen PMOL-AM00007 is accessioned at the Paleontological Museum of Liaoning (PMOL) at Shenyang Normal University, Shenyang, China. Detailed information on the specimens is listed in table S7. All fossil specimens are accessible to researchers. All data and phylogenetic datasets are provided in the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ads9734
Materials and Methods
Supplementary Text
Figs. S1 to S7
Tables S1 to S7
References (*51*–*98*)
MDAR Reproducibility Checklist

Submitted 5 September 2024; accepted 28 January 2025
10.1126/science.ads9734

ATMOSPHERE

Sunlight drives the abiotic formation of nitrous oxide in fresh and marine waters

Elizabeth Leon-Palmero[1,2]*†, Rafael Morales-Baquero[1], Bo Thamdrup[2], Carolin Löscher[2], Isabel Reche[1,3]

Nitrous oxide (N_2O) is a potent greenhouse gas and the main stratospheric ozone-depleting agent, yet its sources are not well resolved. In this work, we experimentally show a N_2O production pathway not previously considered in greenhouse gas budgets, which we name photochemodenitrification. Sunlight induces substantial and consistent N_2O production under oxic abiotic conditions in fresh and marine waters. We measured photochemical N_2O production rates using isotope tracers and determined that nitrite is the main substrate and that nitrate can also contribute after being photoreduced to nitrite. Additionally, this N_2O production was strongly correlated to the radiation dose. Photochemodenitrification exceeded biological N_2O production in surface waters. Although previously overlooked, this process may contribute considerably to global N_2O emissions through its occurrence in fresh and marine surface waters.

Nitrous oxide (N_2O) is the main stratospheric ozone-depleting agent (*1*) and one of the strongest greenhouse gases—about 273 times as potent as carbon dioxide (*2*). Since preindustrial times, the concentration of atmospheric N_2O has risen by 23% (*2*), in a trend strongly correlated to the higher bioavailability of reactive nitrogen (N) in the environment, largely owing to industrial N fixation by the Haber–Bosch process (*3*). A considerable portion of the anthropogenic N applied to land as fertilizers enters rivers, estuaries, and continental shelves, which boosts the production and emission of N_2O in these aquatic systems (*4*–*6*). These emissions resulting from N inputs—estimated to 0.5 Tg of N $year^{-1}$ for the 2007 to 2016 period—exceed natural emissions from inland waters, estuaries, and coastal zones [0.3 Tg N $year^{-1}$ (*7*)]. The rate of increase in atmospheric N_2O during the past decade was even faster than predicted by the Intergovernmental Panel on Climate Change (IPCC) (*8*), which emphasizes the need for better identification of the sources of N_2O to reduce uncertainty in climate assessments and optimize mitigation strategies (*7*).

Ammonia oxidation and denitrification are the microbial processes that are assumed to control the N_2O budget in aquatic ecosystems. Ammonia-oxidizing archaea (AOA) and bacteria (AOB) release N_2O in well-oxygenated waters as a side product during ammonia oxidation to nitrite (NO_2^-) (*9*). N_2O is also an intermediate product during denitrification of NO_3^- and NO_2^- to N_2O and N_2, a process that is usually coupled to organic matter oxidation and occurs under oxygen-depleted conditions (*10*). Both microbial processes have been extensively studied in the water column and sediments of aquatic ecosystems. Abiotic processes, such as chemodenitrification, can also contribute to the production of N_2O in soils and marine sediments (*11*, *12*). During chemodenitrification, NO_3^- and NO_2^- are abiotically reduced to N_2O coupled to the oxidation of metals, such as Fe (II), or organic matter (*11*). However, the abiotic production of N_2O within the water column of fresh waters and marine ecosystems remains largely unexplored, particularly in surface waters. Processes occurring in the upper meters of the well-mixed layer hold the potential for a disproportionate contribution to N_2O emissions owing to their closeness to the atmosphere, which may facilitate the air-sea exchange of N_2O when oversaturated. In this work, we describe the discovery of a photochemical source of N_2O, which we named

[1]Departamento de Ecología, Universidad de Granada, Granada, Spain. [2]Nordcee, Department of Biology, University of Southern Denmark, Odense, Denmark. [3]Research Unit Modeling Nature (MNat), Universidad de Granada, Granada, Spain.
***Corresponding author. Email: el23@princeton.edu**
†Present address: Department of Geosciences, Princeton University, Princeton, NJ, USA.

photochemodenitrification. We detected this reaction, mediated by sunlight, that uses NO_2^- as substrate in two freshwater reservoirs and two coastal marine systems, demonstrating that it may represent an unrecognized but widespread process relevant to global budgets.

Photochemical formation of N_2O

In a previous study, we found recurrently higher N_2O emissions during the daytime versus during the nighttime during 24-hour flux measurement campaigns over the course of 2 years in two freshwater reservoirs in southeast Spain (Cubillas and Iznájar). Despite differences in magnitude among years and reservoirs, this pattern in N_2O emissions was consistent and significantly correlated to the solar cycle, with daytime emissions up to one order of magnitude greater than nighttime emissions (*13*). However, that pattern was inconsistent with classical microbial production through ammonia oxidation, which is inhibited by light (*14*), or through denitrification, which requires oxygen depletion (*15*, *16*). We hypothesized that sunlight may induce the photochemical production of N_2O in surface waters, boosting the fluxes during the daytime. To investigate this hypothesis, we first performed four incubation experiments, exposing surface water from Cubillas and Iznájar reservoirs to natural sunlight with inhibition of biological activity by the addition of $HgCl_2$ (experiments 1 to 4; table S1). We used ultraviolet (UV)–transparent quartz vials and included dark controls during the incubations. All experiments were conducted under natural, oxic conditions (more details are provided in the materials and methods and fig. S1). The incubation experiments performed during this study and the initial inorganic N pools are listed in table S1.

We detected a significant, consistent, and continuous increase in the N_2O concentration in the sunlight treatments in the four incubation experiments relative to the dark controls (Fig. 1; experiments 1 to 4). Photochemical N_2O production was detected in both the experiments performed using unfiltered (experiments 1 to 3) and filtered water (0.7 μm pore size; experiment 4), with rates varying from 1.01 ± 0.22 nmol N-N_2O liter^{-1} day^{-1} in experiment 4 to 66.6 ± 15.1 nmol N-N_2O liter^{-1} day^{-1} in experiment 2 (table S2). The photochemical N_2O production rate measured in experiment 4 was equivalent to $2.6 \times 10^{-9} \pm 5.5 \times 10^{-10}$ nmol-N W^{-1} photosynthetically active radiation (PAR) and to $2.3 \times 10^{-5} \pm 4.6 \times 10^{-6}$ nmol-N W^{-1} UVB received. In relation to the in situ N_2O concentration, the daily relative increase varied from 3.8% (initial concentration = 13.4 nmol liter^{-1}) in experiment 4 to 39.3% (initial concentration = 84.6 nmol liter^{-1}) in experiment 2 (table S2), thus contributing substantially to the N_2O pool.

Previous studies in rivers have reported higher N_2O emissions or concentrations during the daytime versus the nighttime (*17*–*20*), although the opposite has also been shown (*21*, *22*). These patterns have been attributed to changes in N cycling by biological activity and water temperature (*17*–*21*). Liu *et al.* (*23*) found that the seasonal pattern in N_2O fluxes in Antarctic lakes was positively related to daily radiation, concluding that the increase of daily radiation may favor photosynthesis and thus stimulate N_2O production from the decomposition of the algae. The experiments presented in this work indicate that solar radiation directly drives an abiotic production of N_2O that may represent an important source of N_2O to the atmosphere that has not been reported previously or considered in greenhouse gas budgets. Solar radiation, mostly in the UV band (280 to 400 nm), catalyzes many photochemical reactions in fresh and marine surface waters, including the decomposition of chromophoric and recalcitrant organic molecules (*24*); the reduction of metals, such as iron (*25*, *26*) or manganese (*27*); the photochemical production of climate-relevant gases, such as carbon dioxide (*28*), carbon monoxide (*29*), and methane (*30*); and the photolysis of NO_3^- and NO_2^- (*31*–*33*). The discovery presented in this work indicates that solar radiation can also catalyze the production of N_2O in surface waters, and its importance—not only for daily variability but also as a source of N_2O to the atmosphere—should be addressed.

Nitrite and sunlight explain N_2O photochemodenitrification

We hypothesized that NO_2^- may act as a substrate for photochemical N_2O production based on our data from experiments 1 to 4 (fig. S2). A marginally significant relationship between the in situ NO_2^- concentrations and the photochemical N_2O production rates was found (n = 4 observations, $P < 0.1$; fig. S2A). Additionally, we detected a significant decrease in the NO_2^- concentration to below detection (0.5 μmol liter^{-1}) in the sunlight treatments compared with the dark controls during experiment 4 ($P < 0.001$; fig. S2B). We tested our hypothesis by adding ^{15}N-NO_2^- or ^{15}N-NO_3^- to trace the formation of ^{15}N-N_2O [as $^{45}N_2O$ ($^{14}N^{15}N^{16}O$) and $^{46}N_2O$ ($^{15}N^{15}N^{16}O$); experiments 5 to 7], expanding our analysis to include two marine systems at different latitudes (Motril coast and Boknis Eck, experiments 6 and 7, respectively) in addition to a freshwater system (Cubillas reservoir, experiment 5; see table S1 for study site details and fig. S1 for the experimental setup). We included the treatment with ^{15}N-NO_3^- because NO_3^- is generally the most abundant form of dissolved inorganic N in these systems, and NO_2^- derived from its abiotic photolysis (*31*) may serve as a substrate for photochemical production of N_2O.

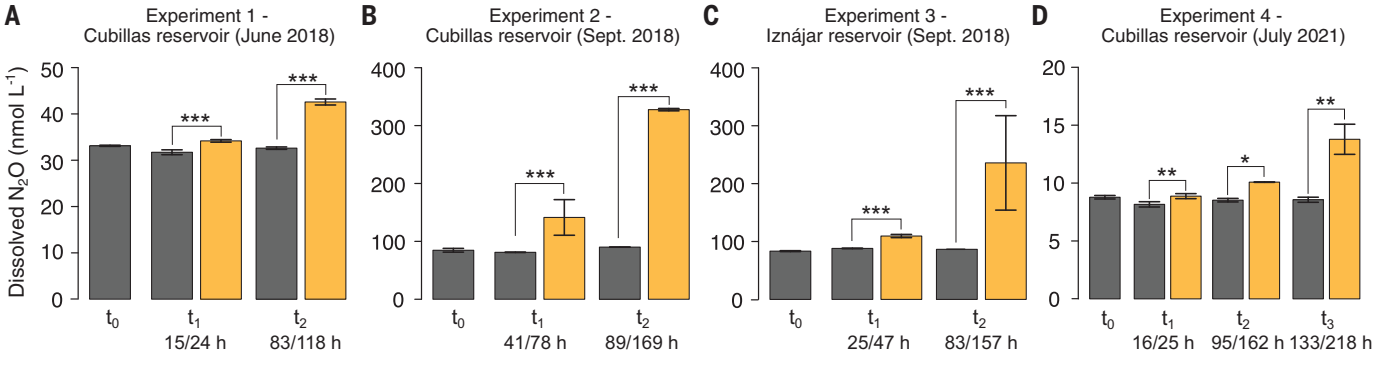


Fig. 1. Effect of sunlight on N_2O production. Dissolved N_2O concentrations over time (i.e., daylight time/total incubation hours) in the incubation experiments 1 to 4. (**A**) Cubillas in June 2018, unfiltered water. (**B**) Cubillas in September 2018, unfiltered water. (**C**) Iznájar in September 2018, unfiltered water. (**D**) Cubillas in July 2021, filtered water (0.7 μm pore size). $HgCl_2$ (1 mmol liter^{-1}) was added in all experiments to inhibit biological activity. Bars represent the mean values ± standard errors over the time course (t_0 to t_3), including dark controls (dark gray bars) and sunlight treatments (yellow bars). Sunlight exposure time/total incubation hours are shown (e.g., 15/24 hours). Note the different scales in the y axes. The significance of the sunlight treatments is included for each experiment: *P < 0.05; **P < 0.01; ***P < 0.001. Rates are presented in table S2. More statistical details are provided in table S3.

We detected a significant production of ^{15}N-N_2O from ^{15}N-NO_2^- in the sunlight treatments in the three study systems, with longer sunlight exposure yielding more N_2O (Fig. 2, A to C, and fig. S3). Significant production of ^{15}N-N_2O from ^{15}N-NO_3^- was also observed in experiments 5 and 6 (Fig. 2, D and E, and fig. S3), although the rate of N_2O production from ^{15}N-NO_3^- was at least one order of magnitude lower than that from ^{15}N-NO_2^-. These results support our hypothesis that NO_2^- is the direct and main substrate for this reaction. In turn, NO_3^- must be photoreduced to NO_2^- to later produce N_2O in a second step. On the basis of these findings, we named the reaction photochemodenitrification. The photochemical process generated both single- and double-labeled N_2O ($^{45}N_2O$ and $^{46}N_2O$) from both tracers (fig. S3). This is broadly consistent with the isotope pairing expected from the production of N_2O from a mixture of added $^{15}NO_2^-$ and unlabeled nitrite (^{14}N-NO_2^-) from either the initial in situ pool or formed, for example, from NO_3^- photolysis (fig. S4). The contribution of other forms of in situ organic or inorganic N compounds as sources of ^{14}N to the formation of $^{45}N_2O$ cannot be excluded (see the supplementary text for further explanation).

Photochemodenitrification depended linearly on the sunlight radiation received (Fig. 3 and table S5). We found similar radiation normalized rates (i.e., the slope in the increase in N_2O per radiation dose) across experiments, whether based on UVB or PAR (Fig. 3 and table S5). In experiment 4, in which we did not add any tracer (Fig. 3A), the slopes for PAR ($S_{PAR} = 5.9 \times 10^{-6} \pm 1.2 \times 10^{-6}$) and UVB radiation ($S_{UVB} = 0.051 \pm 0.010$) were not significantly different from the slopes calculated for the $^{15}NO_2^-$ additions in experiments 5 to 7 (Fig. 3B; $S_{PAR} = 6.3 \times 10^{-6} \pm 2.4 \times 10^{-7}$ and $S_{UVB} = 0.054 \pm 0.002$). In agreement with our previous results, the slopes were one order of magnitude lower for the $^{15}NO_3^-$ additions (Fig. 3C).

Rates of photochemical production of ^{15}N-N_2O in the ^{15}N-NO_2^- addition experiments varied from 0.03 ± 0.00 nmol-N liter^{-1} day^{-1} in Boknis Eck (experiment 7) to 2.03 ± 0.18 nmol-N liter^{-1} day^{-1} in the Cubillas reservoir (experiment 5; table S6). Photochemical production rates from ^{15}N-NO_3^- were similar in the Motril coast (0.26 ± 0.02 nmol-N liter^{-1} day^{-1}) and the Cubillas reservoir (0.29 ± 0.04 nmol-N liter^{-1} day^{-1}) (experiments 5 and 6; table S6). With both ^{15}N-NO_2^- and ^{15}N-NO_3^-, the photochemical N_2O production rates measured in the Cubillas reservoir (experiment 5) and the Motril coast (experiment 6) were larger than in Boknis Eck (experiment 7; Fig. 2 and table S6). The higher substrate availability and higher radiation doses in experiments 5 and 6 versus those in experiment 7 could have caused these differences. We used higher concentrations of ^{15}N-labeled tracers in experiments 5 and 6 compared with experiment 7 (i.e., 2 to 5 μmol liter^{-1} versus 0.5 μmol liter^{-1}) to reflect the natural differences in NO_3^- and NO_2^- concentrations found in these systems (table S1). In addition, the lower production detected in experiment 7 compared with experiments 5 and 6 may also be related to the lower radiation received in this experiment (Fig. 3), which was incubated in Odense, Denmark (~55°N, 57-m altitude), whereas the others were incubated in Granada, Spain (~37°N, 738-m altitude). Despite the lower radiation dose received, which determined the lower ^{15}N-N_2O production, the radiation-normalized rates were similar in the three systems at different latitudes (Fig. 3 and table S5). UV radiation, which may be causing this reaction, tends to increase as we approach the equator and ascend in altitude. Sunlight hours per day also change with the time of the year and latitude, but this variation was accounted for by considering the exact number of sunlight hours in the rate calculations. In addition to the experimental evidence, we also found that the photochemical N_2O production rates were a function of in situ NO_2^- concentration ($P < 0.05$), following an exponential fit [N-N_2O production = $0.32e^{0.23(\text{nitrite})}$ (where e is Euler's number), adjusted coefficient of determination (R^2) = 0.65; Fig. 4]. This finding supports the hypothesis that NO_2^- plays an active role in the photochemical formation of N_2O.

We performed a final experiment to rule out the potential impact of the biocide $HgCl_2$ on the photochemical production of N_2O (experiment 8; fig. S5 and table S7). Water was gently prefiltered once (1.0 μm), and sterile filtered twice (0.2 μm) to remove the organisms present in the sample. Then, parallel treatments were incubated with and without $HgCl_2$. We detected a significant increase in the ^{15}N-N_2O concentration after sunlight exposure in both treatments. N_2O production in sunlight conditions was 54% larger ($P < 0.05$) in the treatment without $HgCl_2$ than in the treatment with $HgCl_2$, and no production was detected in dark conditions. Although unlikely, we cannot exclude that the difference may be the result of biological reactions by cells smaller than 0.2 μm. However, if that were the case, we would likely have detected production in the dark controls without $HgCl_2$, but we did not detect any production in those controls. Based on these results, it is very unlikely that $HgCl_2$ causes the photochemical N_2O production. If anything, $HgCl_2$ addition may reduce photochemical N_2O production rather than increase it.

Our data show experimental (experiments 1 to 7; Figs. 1 and 2) and in situ evidence (Fig. 4) that

Table 1. Summary of the rates of photochemodenitrification and biological N_2O production from ammonia oxidation. When derived from ^{15}N labeling experiments, the total rates of N_2O production were obtained by adding the rates from ^{15}N-NO_2^- and from ^{15}N-NO_3^-. The incubation experiments to measure biological N_2O production were performed in darkness, except experiment 7, which included darkness and natural day-night cycle treatments. More details are presented in tables S2, S6, and S8. Dashes indicate not applicable.

Site (experiment)	Total N_2O from photochemodenitrification (nmol N-N_2O liter^{-1} day^{-1})		N_2O from ammonia oxidation (nmol N-N_2O liter^{-1} day^{-1})	
Cubillas and Iznájar reservoirs (experiments 1 to 4 and 9)	June 2018 (exp. 1, Cubillas)	2.96 (±0.20)	July 2018 (exp. 9)	0.83 (±0.00): Iznájar; 2.08 (±0.15): Cubillas
	September 2018 (exp. 2, Cubillas)	66.56 (±15.11)	September 2018 (exp. 9)	0.06 (±0.00), 0.47 (±0.12): Cubillas; 0.93 (±0.25): Iznájar
	September 2018 (exp. 3, Iznájar)	47.86 (±17.44)		
	July 2021 (exp. 4, Cubillas)	1.01 (±0.22)	–	–
Cubillas reservoir (experiment 5)	July 2021	2.32 (±0.22)	–	–
Motril coast (experiment 6)	July 2021	1.41 (±0.06)	–	–
Boknis Eck (experiment 7)	July 2022	0.03 (±0.00)	July 2022	0.015 (±0.002) (darkness); 0.007 (±0.002) (day-night cycle)

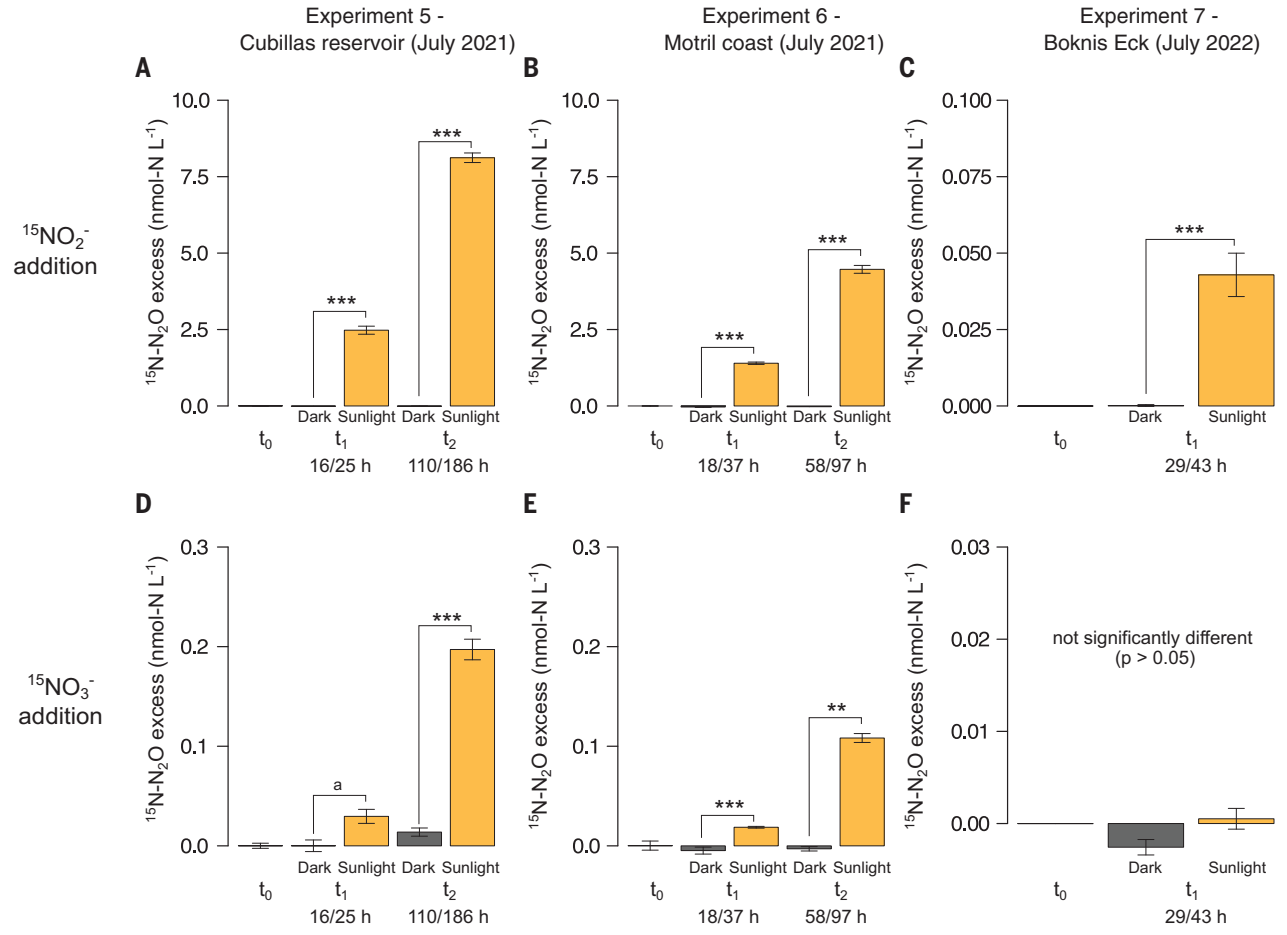


Fig. 2. Photochemical production of ^{15}N-N_2O from ^{15}N-NO_2^- and ^{15}N-NO_3^-. (**A** to **F**) ^{15}N-N_2O excess (in nanomoles of N per liter) calculated with respect to t_0, over incubation time (sunlight hours/total hours) in the experiments in the Cubillas reservoir (filtered water, 0.7 μm pore size) [(A) and (D)], the Motril coast (filtered water, 0.7 μm pore size) [(B) and (E)], and Boknis Eck (unfiltered water) [(C) and (F)], from ^{15}N-NO_2^- [(A) to (C)] and ^{15}N-NO_3^- [(D) to (F)]. $HgCl_2$ (1 mmol liter^{-1}) was added in all experiments to inhibit biological activity. Bars represent the mean values ± standard errors over the time course (t_0 to t_2), including dark controls (dark gray bars) and sunlight treatments (yellow bars). Note the different scales in the *y* axes. The significance of the sunlight treatments is included for each experiment: *a* indicates $P < 0.1$; $^{**}P < 0.01$; $^{***}P < 0.001$. Photochemical production rates as a function of the bottle area, volume, and radiation doses are presented in table S6, and the statistical details are provided in table S4.

NO_2^- and NO_3^- are substrates for the photochemical production of N_2O and that this process significantly depends on sunlight exposure (Fig. 3). NO_2^- and especially NO_3^- are abundant compounds in eutrophic aquatic systems, such as many freshwater bodies and wetlands, as well as coastal areas and estuaries, which are strongly influenced by N loading from their watersheds (*34*, *35*) and act as global hotspots for N_2O production (*4*, *6*, *36*). For instance, NO_2^- and NO_3^- concentrations reached up to 11 μmol liter^{-1} and >300 μmol liter^{-1}, respectively, in the surface water column of Mediterranean reservoirs (*34*) and >2 μmol liter^{-1} in the surface waters of the Chesapeake Bay estuary (*37*). In the open ocean, inorganic N concentrations in surface waters vary from region to region. They are typically low in oligotrophic tropical and subtropical regions, where nutrient uptake by phytoplankton is high and upwelling is less common. However, concentrations are higher in upwelling zones and polar regions, where annual average surface values of NO_3^- are ~5 to 25 μmol liter^{-1} (*38*).

Therefore, photochemical N_2O production rates are expected to be substantial in regions with greater availability of inorganic N and lower in oligotrophic waters, which generally correlates with patterns in N_2O concentration and biological N_2O production (*4*, *6*, *7*, *36*, *39*). We hypothesize that other biological reactions, such as ammonia oxidation, and/or abiotic processes, such as the photodegradation of dissolved organic matter, which release NO_2^- and other forms of inorganic N (*40*), may enhance photochemical N_2O production. If production and consumption are tightly coupled, they might create a cryptic cycle in which the NO_2^- remains undetected. This effect may be particularly important in oligotrophic waters. The contribution of other sources of NO_2^- or forms of N may contribute to explain the results found in Boknis Eck (experiment 7), in which a higher production of $^{45}N_2O$ compared with $^{46}N_2O$ was detected, despite in situ NO_2^- and NO_3^- concentrations being below detection levels. This result suggests that other forms or sources of organic or inorganic in situ N compounds may contribute ^{14}N to the formation of $^{45}N_2O$ (fig. S3 and supplementary text).

The photochemical reduction of NO_2^- to N_2O during photochemodenitrification may be coupled to the oxidation of organic matter or metals, such as Fe or Mn, similar to chemodenitrification (*11*, *12*, *41*). These metals experience photoreduction in surface waters (*25*, *27*, *42*). Alternatively, N_2O may also be produced through UV light–catalyzed photolysis (*33*, *43*, *44*). Additional research is needed to elucidate the ultimate mechanism of this process.

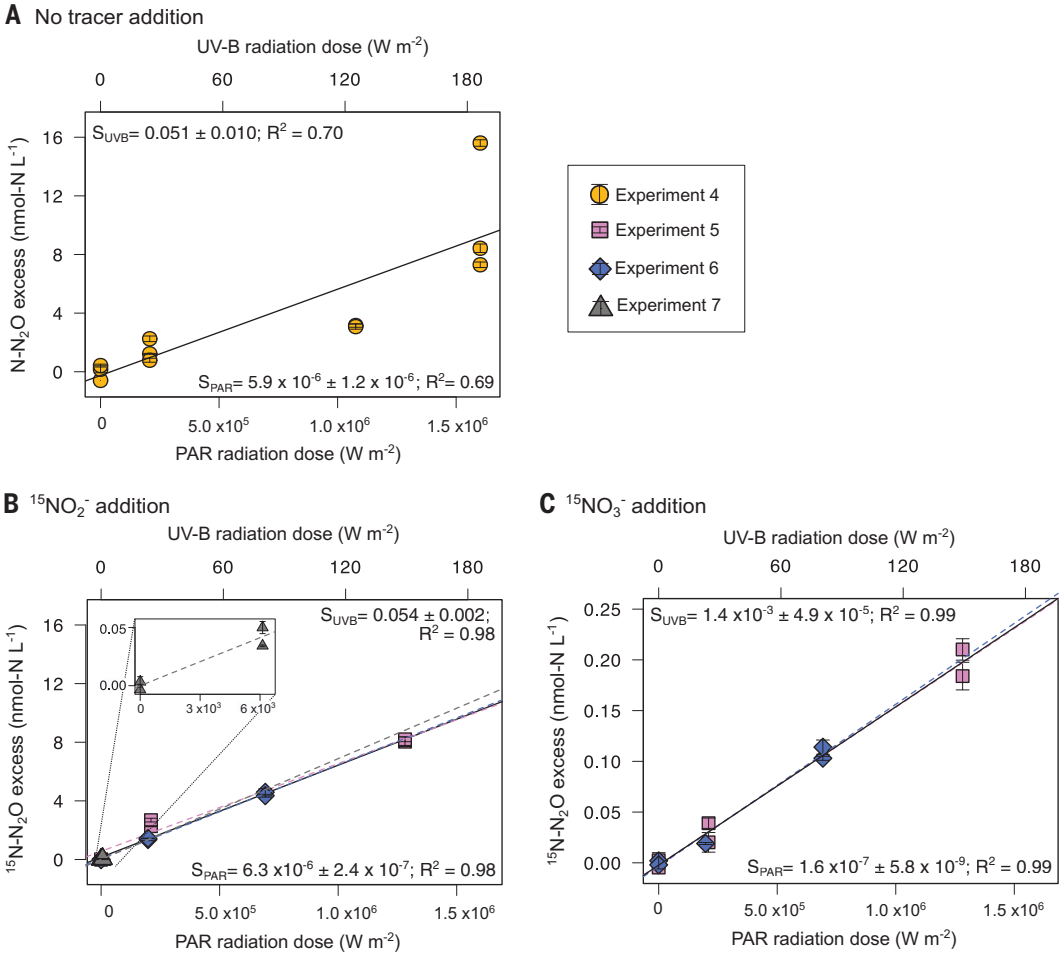


Fig. 3. Photochemical production of N_2O as function of UVB and PAR dose. Linear increases in the concentrations of $^{(15)}N$ in the N_2O pool relative to the initial abundance (N_2O excess, in nanomoles of N per liter) in experiments 4 to 7. (**A**) No tracer addition. (**B**) $^{15}NO_2^-$ addition. (**C**) $^{15}NO_3^-$ addition. The combined slopes (S_{PAR} and S_{UVB}) are shown in the figure, and the individual slopes for each experiment are provided in table S5. Note the zoom-in plot in (B) and the different scales in the *y* axes. We did not measure UVB or PAR during experiments 1 to 3.

Photochemical processes rely mostly on UV radiation, including UVB (280 to 315 nm) and UVA (315 to 400 nm), which is rapidly attenuated in the water column. For instance, NO_3^- photolysis occurs in the UVB band, centered at 302 nm, whereas NO_2^- photolysis occurs in the UVB and UVA bands, with an absorption maximum at 354 nm (*33*). Longer wavelengths carry less energy and penetrate deeper into the water column. Vanderploeg *et al.* found that 1% of the UV radiation penetrated from 2 to 4 m for UVB (305 nm) and between 11 and 22 m for UVA (395 nm) in Lake Michigan (*45*). By contrast, there is still about 10% of the surface radiation at depths between 6 and 31 m (UVB) and between 15 and 104 m (UVA) in the oligotrophic ocean (*46*). In more eutrophic water bodies, sunlight attenuation is typically high; yet, the likely greater availability of inorganic N in these systems may compensate for the total depth-integrated photochemical production rates (Fig. 4). Solar radiation also changes with latitude and altitude. Therefore, N_2O production by photochemodenitrification may present a latitudinal pattern in the global ocean and fresh waters related to the differences in solar radiation and influenced by NO_2^- and NO_3^- concentrations around the globe.

Photochemodenitrification versus biological N_2O production

The N_2O production rates from photochemodenitrification exceeded the biological N_2O production from ammonia oxidation, which has been classically considered the key N_2O production process in surface waters (Table 1). Biological production of ^{15}N-N_2O from ammonia oxidation varied from 0.06 ± 0.00 nmol N-N_2O liter^{-1} day^{-1} in Iznájar to 2.08 ± 0.15 nmol N-N_2O liter^{-1} day^{-1} in Cubillas (experiment 9; Table 1). These biological rates were determined by incubating the samples in darkness. In Boknis Eck, biological production of ^{15}N-N_2O reached up to 0.015 ± 0.002 nmol N-N_2O liter^{-1} day^{-1} when the samples were incubated in darkness and 0.007 ± 0.002 nmol N-N_2O liter^{-1} day^{-1} when the samples were incubated in natural day-night conditions (Table 1). We did not find a significant relationship between N_2O production rates by ammonia oxidation and the in situ N_2O concentration in the study reservoirs (fig. S6). However, we found that in situ N_2O concentration was a function of the photochemical N_2O production rates ($n = 5$, $P < 0.01$; fig. S6). Therefore, the production of N_2O by photochemodenitrification may contribute substantially to sustaining the recurrent surface N_2O supersaturation found in fresh waters (*34*) and potentially in marine ecosystems, including coastal areas (*37*) and the open ocean (*47*), where the production of N_2O from ammonia oxidation in surface waters is typically low.

The rates of N_2O photochemodenitrification detected were higher than the N_2O production by ammonia oxidation in coastal and open ocean regions (*37*, *47*) but lower than those found for N_2O production by denitrification in rivers (*48*). They were similar to the N_2O production rates by phytoplankton cultures (*49*). Moreover, ammonia oxidation rates are typically measured in the dark, which eliminates possible contributions with photochemical processes. Because ammonia oxidizers are sensitive to light (*14*), this may overestimate the biological N_2O production rates in surface waters, as illustrated by our finding of higher rates in darkness compared with natural day-night conditions (experiment 7; Table 1). Still, the photochemical N_2O production rate was at least twice the biological N_2O production in darkness.

N_2O production by photochemodenitrification represents a substantial contribution to the N_2O pool in freshwater reservoirs and may also represent a considerable contribution in coastal marine waters. The fact that N_2O

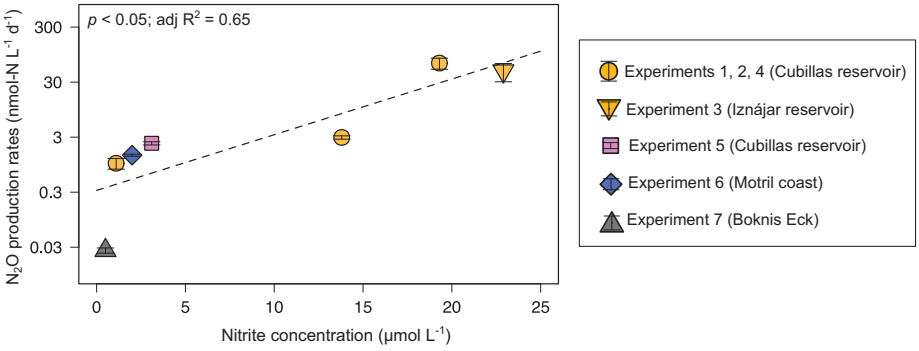


Fig. 4. Total N_2O photochemical production rates as a function of the in situ nitrite concentration. N_2O photochemical production rates (in nanomoles of N per liter per day) ± standard errors are represented. Note the log scale in the *y* axis. The N_2O photochemical production rate measured in the Cubillas reservoir in June 2018 (sampling performed on 22 June) was related to the nitrite concentration measured in the Cubillas reservoir on 4 July 2018. In experiments 5 to 7, we added the nitrite used as tracer to the in situ nitrite concentration.

production by photochemodenitrification exhibits its maximum at the very surface of the water, which is in direct contact with the atmosphere, suggests that this reaction may have a disproportionate effect on N_2O fluxes. This is because the newly formed N_2O may diffuse more quickly to the atmosphere when oversaturated compared with the N_2O produced and stored deeper in the water column. Our findings could also imply that the emission of N_2O from surface waters may be larger than previously estimated, particularly by biogeochemical models (*36*, *50*). The recent global synthesis on N_2O fluxes by Resplandy *et al.* (*36*) showed that global ocean biogeochemical model emission estimates for coastal waters are lower than those based on observations (*39*). This may indicate that these biogeochemical models underestimate emissions because they do not account for all N_2O sources. We suggest that this underestimation could be, at least partially, related to photochemical N_2O production.

The N_2O production pathway presented in this work (fig. S7) represents a previously unrecognized but potentially relevant source of N_2O for future emission budgets. This reaction may be particularly important for eutrophic freshwater bodies, coastal areas, and upwelling marine regions, which are global N_2O emission hotspots (*4*, *6*, *7*, *36*). In the coming decades, N_2O emissions may be further amplified by the projected increase in global N export to fresh waters and coastal environments resulting from the growing population (*35*). This increase may lead to higher rates of photochemical N_2O formation. Therefore, it is crucial to include this pathway in future assessments. This discovery represents a breakthrough in the study of global sources of N_2O, helping to reduce the current uncertainties in the bottom-up estimates highlighted in the last global N_2O inventories (*7*) and to explain accelerating emissions, which exceed some of the highest projected scenarios (*8*).

REFERENCES AND NOTES

1. A. R. Ravishankara, J. S. Daniel, R. W. Portmann, *Science* **326**, 123–125 (2009).
2. Intergovernmental Panel on Climate Change (IPCC), *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2023).
3. N. Gruber, J. N. Galloway, *Nature* **451**, 293–296 (2008).
4. S. P. Seitzinger, C. Kroeze, R. V. Styles, *Chemosphere Glob. Change Sci.* **2**, 267–279 (2000).
5. E. León-Palmero, R. Morales-Baquero, I. Reche, *Environ. Res. Lett.* **15**, 044012 (2020).
6. Y. Li *et al.*, *Nat. Commun.* **15**, 942 (2024).
7. H. Tian *et al.*, *Nature* **586**, 248–256 (2020).
8. R. L. Thompson *et al.*, *Nat. Clim. Change* **9**, 993–998 (2019).
9. B. B. Ward, "Nitrification" in *Encyclopedia of Ecology*, vol. 2, B. Fath, Ed. (Elsevier, ed. 2, 2013), pp. 351–358.
10. B. B. Ward, *Science* **341**, 352–353 (2013).
11. X. Zhu-Barker, A. R. Cavazos, N. E. Ostrom, W. R. Horwath, J. B. Glass, *Biogeochemistry* **126**, 251–267 (2015).
12. S. D. Wankel *et al.*, *Nat. Commun.* **8**, 15595 (2017).
13. E. León-Palmero, R. Morales-Baquero, I. Reche, bioRxiv 2023.10.24.563772 [Preprint] (2023).
14. S. N. Merbt *et al.*, *FEMS Microbiol. Lett.* **327**, 41–46 (2012).
15. P. Bonin, M. Gilewicz, J. C. Bertrand, *Can. J. Microbiol.* **35**, 1061–1064 (1989).
16. T. Dalsgaard *et al.*, *mBio* **5**, e01966 (2014).
17. J. A. Harrison, P. A. Matson, S. E. Fendorf, *Aquat. Sci.* **67**, 308–315 (2005).
18. T. J. Clough, L. E. Buckthought, F. M. Kelliher, R. R. Sherlock, *Glob. Change Biol.* **13**, 1016–1027 (2007).
19. L. Yang, W. Yan, P. Ma, J. Wang, *J. Geogr. Sci.* **21**, 820–832 (2011).
20. H. M. Baulch *et al.*, *Freshw. Biol.* **57**, 509–525 (2012).
21. M. S. Rosamond, S. J. Thuss, S. L. Schiff, R. J. Elgood, *J. Environ. Qual.* **40**, 256–270 (2011).
22. R. L. Woodrow *et al.*, *Limnol. Oceanogr. Lett.* **9**, 276–285 (2024).
23. Y. Liu *et al.*, *Atmos. Environ.* **45**, 1464–1475 (2011).
24. I. Reche, M. L. Pace, J. J. Cole, *Ecosystems* **3**, 419–432 (2000).
25. L. Emmenegger, R. Schönenberger, L. Sigg, B. Sulzberger, *Limnol. Oceanogr.* **46**, 49–61 (2001).
26. K. Barbeau, E. L. Rue, K. W. Bruland, A. Butler, *Nature* **413**, 409–413 (2001).
27. W. G. Sunda, S. A. Huntsman, G. R. Harvey, *Nature* **301**, 234–236 (1983).
28. B. Koehler, T. Landelius, G. A. Weyhenmeyer, N. Machida, L. J. Tranvik, *Global Biogeochem. Cycles* **28**, 696–711 (2014).
29. L. Conte, S. Szopa, R. Séférian, L. Bopp, *Biogeosciences* **16**, 881–902 (2019).
30. Y. Li, C. G. Fichot, L. Geng, M. G. Scarratt, H. Xie, *Geophys. Res. Lett.* **47**, e2020GL088362 (2020).
31. O. C. Zafiriou, M. B. True, *Mar. Chem.* **8**, 33–42 (1979).
32. O. C. Zafiriou, M. B. True, *Mar. Chem.* **8**, 9–32 (1979).
33. J. J. Jankowski, D. J. Kieber, K. Mopper, *Photochem. Photobiol.* **70**, 319–328 (1999).
34. E. León-Palmero, R. Morales-Baquero, I. Reche, *Limnol. Oceanogr.* **68**, 1734–1749 (2023).
35. A. H. W. Beusen *et al.*, *Glob. Environ. Change* **72**, 102426 (2022).
36. L. Resplandy *et al.*, *Glob. Biogeochem. Cycles* **38**, e2023GB007803 (2024).
37. W. Tang *et al.*, *Limnol. Oceanogr.* **67**, 2101–2116 (2022).
38. EU Copernicus Marine Service information (CMEMS), "Global Ocean Biogeochemistry Analysis and Forecast" (Marine Data Store, 2025); https://doi.org/10.48670/moi-00015.
39. S. Yang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 11954–11960 (2020).
40. R. J. Kieber, A. Li, P. J. Seaton, *Environ. Sci. Technol.* **33**, 993–998 (1999).
41. J. Heil *et al.*, *Geochim. Cosmochim. Acta* **139**, 72–82 (2014).
42. M. J. A. Rijkenberg *et al.*, *Mar. Chem.* **93**, 119–129 (2005).
43. R. G. Zepp, J. Hoigne, H. Bader, *Environ. Sci. Technol.* **21**, 443–450 (1987).
44. T. Liu, J. Deng, C. Yang, M. Liu, Y. Liu, *Separ. Purif. Tech.* **304**, 122364 (2023).
45. H. A. Vanderploeg *et al.*, *J. Great Lakes Res.* **50**, 102291 (2024).
46. S. Overmans *et al.*, *J. Geophys. Res. Oceans* **127**, e2021JC017654 (2022).
47. X. S. Wan *et al.*, *Nat. Geosci.* **16**, 29–36 (2022).
48. J. J. Beaulieu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 214–219 (2011).
49. A. R. McLeod, T. Brand, C. N. Campbell, K. Davidson, A. D. Hatton, *J. Geophys. Res. Biogeosci.* **126**, e2021JG006345 (2021).
50. J. Wang *et al.*, *Environ. Sci. Technol.* **57**, 13506–13519 (2023).

ACKNOWLEDGMENTS

We thank J. Forja and T. Ortega for helping with gas chromatography analysis at the Department of Physical Chemistry of the University of Cádiz and G. Sánchez and L. Alados-Arboledas from the Applied Physics Department of the Andalusian Inter-University Institute for Earth System Research (IISTA-CEAMA) for providing the sunlight radiation data during the 2021 experiments. We also thank L. Bristow for measuring samples of experiments 7 and 8 and H. W. Bange for organizing the monthly Boknis Eck sampling campaign and providing the inorganic N measurements. We thank B. B. Ward for hosting E.L.-P. in 2018 and allowing her to use the mass spectrometer for experiment 9 measurements and for insightful comments and support during the review process. We also express our gratitude to the members of the Ward laboratory at Princeton University for their thoughtful comments on this manuscript. **Funding:** This study was supported by MNat - Consejería de Universidad, Investigación e Innovación, Junta de Andalucía, Spain, grant QUAL21-011 (I.R.); Spanish Ministry of Science, Innovation and Universities, grants RTI2018- 098849-B-I00 (I.R. and R.M.-B.) and MICIU/AEI/10.13039/501100011033 (I.R.); the European Regional Development Fund (ERDF), grant PID2022.1378650B.100 (I.R.); Spanish Ministry of Education, Culture and Sport, PhD grant FPU014/02917 (E.L.-P.); AEET-SIBECOL project for young researchers, grant PHOTON 2020 (E.L.-P.); Danmarks Frie Forskningsfond, grants 1026-00428B (C.L.) and 1127-00362B (B.T.); Carlsbergfondet, grant CF18-1071 (B.T.); and European Commission, Marie Skłodowska-Curie postdoctoral fellowship HORIZON-MSCA-2021-PF-01, grant 101066750 (E.L.-P.). **Author contributions:** Conceptualization: E.L.-P., I.R., R.M.-B., C.L., B.T.; Funding acquisition: I.R., R.M.-B., C.L., E.L.-P.; Investigation: E.L.-P., I.R., R.M.-B., B.T.; Methodology: E.L.-P., I.R., R.M.-B., B.T., C.L.; Project administration: I.R., E.L.-P., C.L.; Supervision: I.R., R.M.-B., C.L.; Visualization: E.L.-P.; Writing – original draft: E.L.-P.; Writing – review & editing: I.R., R.M.-B., B.T., C.L., E.L.-P. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adq0302
Materials and Methods
Supplementary Text
Figs. S1 to S7
Tables S1 to S8
References (*51*–*60*)

Submitted 3 May 2024; accepted 4 February 2025
10.1126/science.adq0302

BIOGEOGRAPHY

Island geography drives evolution of rattan palms in tropical Asian rainforests

Benedikt G. Kuhnhäuser[1,2]*, Christopher D. Bates[3], John Dransfield[1], Connie Geri[4], Andrew Henderson[5], Sang Julia[6], Jun Ying Lim[7], Robert J. Morley[8,9], Himmah Rustiami[10], Rowan J. Schley[11], Sidonie Bellot[1]†, Guillaume Chomicki[12]†, Wolf L. Eiserhardt[1,13]†, Simon J. Hiscock[2]†, William J. Baker[1,13]*†

Distributed across two continents and thousands of islands, the Asian tropics are among the most species-rich areas on Earth. The origins of this diversity, however, remain poorly understood. Here, we reveal and classify contributions of individual tropical Asian regions to their overall diversity by leveraging species-level phylogenomic data and new fossils from the most species-rich Asian palm lineage, the rattans and relatives (Arecaceae, Calamoideae). Radiators (Borneo) generate and distribute diversity, incubators (Indochina, New Guinea, and Sulawesi) produce diversity in isolation, corridors (Java, Maluku, Sumatra, and the Thai-Malay Peninsula) connect neighboring regions, and accumulators (Australia, India, Palawan, and the Philippines) acquire diversity generated elsewhere. These contrasting contributions can be explained by differences in region size and isolation, elucidating how the unique island-dominated geography of the Asian tropics drives their outstanding biodiversity.

The exceptional diversity of tropical rainforests, housing almost half of all plant species in <7% of Earth's land surface, has attracted the attention of biologists for more than a century (*1*, *2*). However, the biotic assembly, defined here as the build-up of species diversity through dispersal and speciation, of this extraordinary biodiversity remains poorly understood due to the scarcity of genetic data of native lineages and insufficient conceptual frameworks (*3*). Existing studies of biotic assembly often focus on dispersal by characterizing biogeographic regions either as sources or sinks of biodiversity, but do not take differences in within-region speciation sufficiently into account (*4*, *5*). An expanded framework that considers both dispersal and speciation as key mechanisms through which individual biogeographic regions assemble their species diversity is therefore needed.

Among the world's tropical realms, the Asian tropics are particularly diverse, with ~50,000 plant and 7000 vertebrate species, surpassing the biodiversity of the American and African tropics by 1.5- and 3-fold when area is considered (*6*). The Asian tropics are also geographically and geologically the most complex tropical realm, spreading across a vast distance of >8000 km that ranges from India to Australia. They are dominated by the world's largest island group, the Indo-Australian Archipelago, which comprises >20,000 islands and includes the two largest tropical islands on Earth, Borneo and New Guinea (*7*). Islands on the continental shelves of Asia and Australia (Sunda and Sahul, respectively) are surrounded by shallow seas and were regularly connected to their adjacent mainland during repeated low sea-level stands in the Quaternary [2.58 million years ago (Ma) to the present]. By contrast, the islands between Sunda and Sahul are surrounded by deep seas and have remained separated even when sea levels were lowest (*4*). This complex island landscape started forming around 100 Ma (fig. S1 and tables S1 and S2), when Gondwanan plate fragments collided with Southeastern Asia, forming Borneo, Java, and West Sulawesi, and was further shaped by the collision of India with Asia between 50 and 40 Ma (*8*). From around 30 Ma, Asia converged with the northward-moving Australian plate and the westward-moving Pacific plate, resulting in the rapid uplift of entire islands, including Java and New Guinea (*8*, *9*). However, a direct land connection between Sunda and Sahul was never established. Several islands and island groups, such as Sulawesi, the Philippines, and Maluku, attained their present geological configuration only in the past 5 million years (*8*). Although it is well established that this complex geographic setting and geological history of the Asian tropics has played an important role in shaping their extant biodiversity (*10*–*13*), the contributions of individual tropical Asian regions to their overall diversity remain unclear.

Research on the origins of tropical Asian biodiversity has largely focused on biotic interchanges between Sunda and Sahul, revealing the existence of major biogeographic boundaries such as Wallace's Line and the predominance of eastward dispersal of rainforest lineages (*4*, *14*–*17*). However, studies have often been constrained by low spatial or phylogenetic resolution of the lineages investigated. For a more detailed understanding of the biotic assembly of tropical Asian biodiversity through time and space, high-resolution information about the evolutionary history of lineages that are species rich throughout the Asian tropics is needed, yet such data remain scarce (*2*, *3*, *18*–*20*). Rapid advances in DNA sequencing of natural history collections (*21*, *22*) now allow us to address this knowledge gap by investigating the evolution of tropical Asian lineages at the species level.

Here, we used the rattan palms and relatives (Arecaceae, Calamoideae) to understand biotic assembly of the Asian tropics. The calamoid palms are an ideal model group for this purpose because they (i) are highly diverse across the Asian tropics, (ii) have a rich fossil record spanning the geological history of the area, (iii) are taxonomically and geographically well-known, and (iv) depend on generalist seed dispersers, in common with many other tropical Asian plant lineages (*23*–*31*). The Calamoideae includes 499 Asian species (~1% of tropical Asian plant diversity) in two sister tribes, Calameae and Eugeissoneae (*23*). It contains the rattans, which are spiny, climbing palms from Asia and Africa that are the source of cane for the rattan furniture industry. The Asian rattans encompass the most diverse genus of palms *Calamus*, with 414 species, as well as several smaller genera, all within tribe Calameae. However, tree palms and stemless shrubs (e.g., ~10% of *Calamus* species) also occur in tribe Calameae and its sister tribe, Eugeissoneae (*23*, *32*). The Asian calamoid palms (hereafter "rattans" for brevity) are characteristic components of a wide range of tropical Asian habitats from sea level to 3000 m elevation, including coastal swamps, lowland rainforests, and montane forests. The rattans usually have single-seeded, small, fleshy fruits that are disseminated by a wide range of animals including birds (e.g., hornbills, fruit pigeons, great argus pheasants, and cassowaries), primates (e.g., gibbons, orangutans, langurs, and macaques), small mammals (e.g., palm civets), and possibly reptiles (*29*–*31*), suggesting that all rattans are involved in an unspecialized seed dispersal mutualism.

By building a comprehensive time-calibrated, species-level phylogenetic tree, we infer the

[1]Royal Botanic Gardens Kew, Richmond, Surrey, UK. [2]Department of Biology, University of Oxford, Oxford, UK. [3]PT Horizon Geoconsulting, Jakarta Selatan, Indonesia. [4]Sarawak Forestry Corporation, Jalan Sungai Tapang, Kota Sentosa, Kuching, Sarawak, Malaysia. [5]The New York Botanical Garden, Bronx, NY, USA. [6]Sarawak Herbarium, Forest Department Sarawak, Jalan Datuk Amar Kalong Ningkan, Kuching, Malaysia. [7]Centre for Nature-based Climate Solutions and Department of Biological Sciences, National University of Singapore, Singapore, Singapore. [8]Palynova Ltd, Littleport, UK. [9]Earth Sciences Department, Royal Holloway, University of London, Egham, Surrey, UK. [10]Herbarium Bogoriense, Research Center for Biosystematics and Evolution, National Research and Innovation Agency (BRIN), Cibinong, Bogor, Indonesia. [11]Department of Geography, University of Exeter, Exeter, UK. [12]Department of Biosciences, Durham University, Durham, UK. [13]Department of Biology, Aarhus University, Aarhus, Denmark.
***Corresponding author. Email: b.kuhnhaeuser@kew.org (B.G.K.); w.baker@kew.org (W.J.B.)**
†These authors contributed equally to this work.

biogeographic history of rattans through time and space. We propose a classification of regional contributions to overall rattan diversity across the Asian tropics through speciation and dispersal, which expands the existing source-sink framework (*4*, *5*) and may serve as a general biogeographic framework for understanding the evolution of other lineages and biota. This classification reveals distinct roles among biogeographic regions, namely as radiators (high speciation and emigration), incubators (high speciation but low emigration), corridors (low speciation but high emigration), and accumulators (low speciation and emigration). Finally, we investigated the influence of variation in region size and isolation on the revealed patterns of dispersal and speciation, demonstrating that regional differences can be explained by simple biogeographic rules and may thus apply more broadly to the tropical Asian biota.

A phylogenomic and fossil framework for biogeographic analyses

Leveraging natural history collections using targeted sequencing of hundreds of nuclear genes (tables S3 and S4), we constructed a densely sampled phylogenetic tree of the rattans (*33*). After strict data filtering (removal of samples with <100 genes) of a preliminary analysis with near-complete (95.5%) taxon sampling (fig. S2 and table S5), our final phylogenetic tree included 360 (72.1%) of 499 species and 354 single-copy nuclear genes (Fig. 1A and table S6). The resolved phylogenetic relationships are well supported (93% of nodes at or above genus level supported with local posterior probability of 1; fig. S3).

To time-calibrate the phylogenetic tree, we assessed the fossil record of the Asian rattans. The lineage has a well-documented Cenozoic (66.0 Ma to present) fossil record (*25*), and we substantially extend this record here into the beginning of the Late Cretaceous by describing two new fossil taxa from the Cenomanian (100.5 to 93.9 Ma) and Turonian (93.9 to 89.8 Ma) of New Guinea (Fig. 1, B and C; figs. S4 to S7; and supplementary text). The new taxa, *Dicolpopollis cenomanicus* Morley & Bates sp. nov. and *D. novaguineensis* Morley & Bates sp. nov., corroborate previous unverified reports of *Dicolpopollis* from the same area and geological age (*34*, *35*). They predate the oldest currently accepted records of *Dicolpopollis* by >20 million years (*24*, *36*). Their morphology compares closely with the extant rattan subtribes Calaminae and Plectocomiinae (Fig. 1, D and E, and supplementary text), making them the oldest fossils that can be placed with confidence within a subfamily of the Arecaceae (*36*), and indicating that previous crown node age estimates of the Calamoideae of 12.3 to 80.2 Ma (mean ages; table S7) were too young.

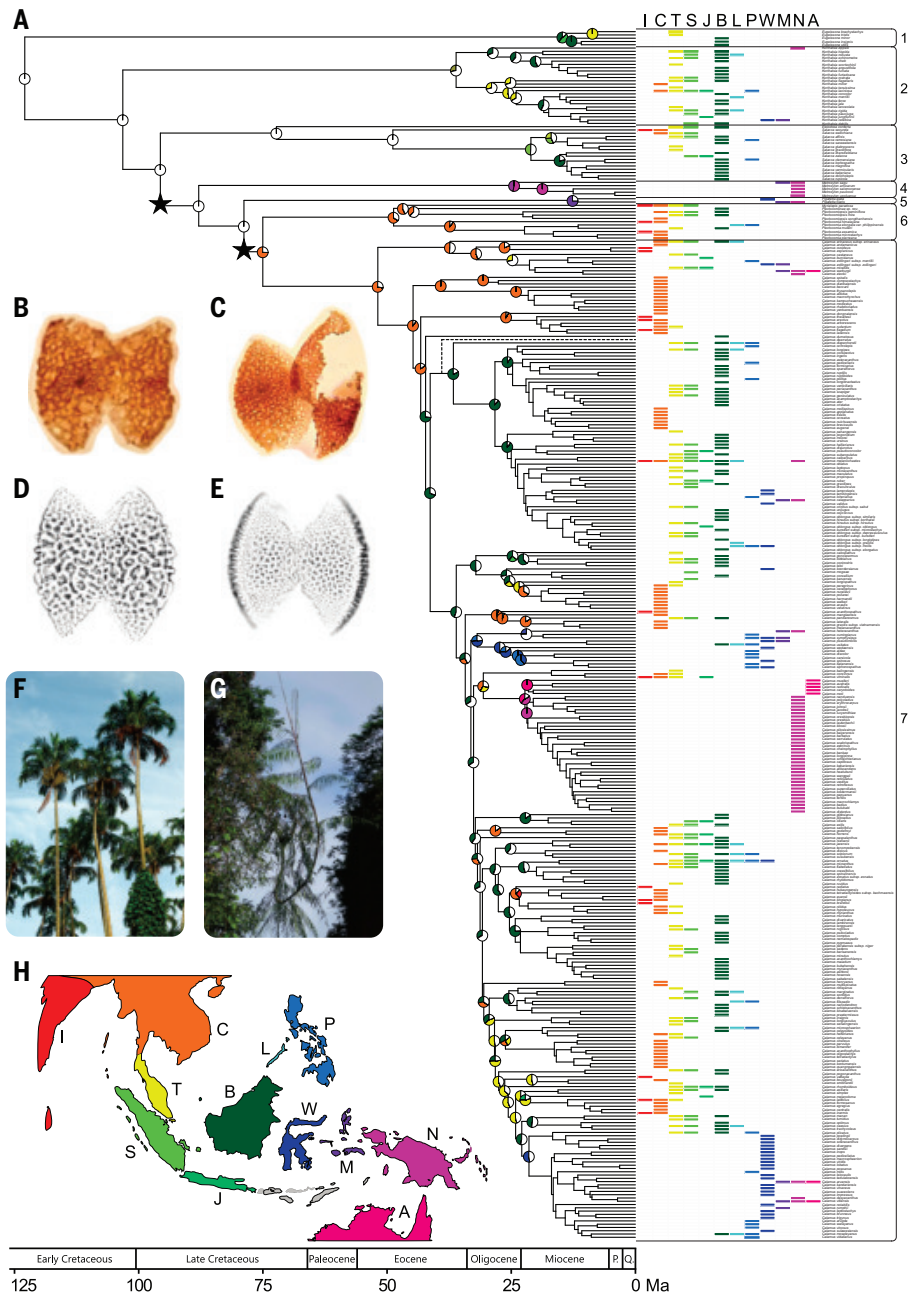


Fig. 1. Spatiotemporal evolution of the Asian rattan palms. (**A**) Ancestral range estimates based on a fossil-calibrated phylogenetic tree of 360 Asian rattan species inferred using 354 single-copy nuclear genes. The occurrence of extant rattans in 12 biogeographic regions is indicated. A, Australia; B, Borneo; C, Indochina; I, India; J, Java; L, Palawan; M, Maluku; N, New Guinea; P, Philippines; S, Sumatra; T, Thai-Malay Peninsula; W, Sulawesi. Pie charts represent ancestral range estimates of selected nodes, with ranges with probability <0.25 combined and shown in white (for estimates of all nodes, see fig. S12). Black stars indicate the two different phylogenetic placements of *Dicolpopollis* fossils in our dating analyses (table S8). Node ages are mean divergence time estimates of dating analysis "MetCal 127.4" (fig. S10). The timescale shows geological epochs or periods. P., Pliocene; Q., Quaternary. Numbered boxes mark clades: 1, tribe Eugeissoneae; 2-7, subtribes of tribe Calameae. 2, Korthalsiinae; 3, Salaccinae; 4, Metroxylinae; 5, Pigafettinae; 6, Plectocomiinae; 7, Calaminae. The dashed line in the phylogenetic tree indicates the position of *Calamus deerratus*, which was not included in the biogeographic analyses because it occurs outside of the Asian tropics (*33*). (**B** to **E**) Fossil and extant *Dicolpopollis*-type rattan pollen. (B and C) New fossil taxa formally described here (figs. S4 to S7 and supplementary text). Photos: C.D.B. (B) *D. cenomanicus* sp. nov. Morley & Bates (Cenomanian; 100.5 to 93.9 Ma). (C) *D. novaguineensis* sp. nov. Morley & Bates (Cenomanian and Turonian; 100.5 to 89.8 Ma). (D and E) Pollen of extant rattan species. Photos: RBG Kew. (D) *Calamus deerratus*. (E) *Calamus formicarius*. (**F** and **G**) Extant Asian rattan species. (F) *Pigafetta filaris*. Photo: W.J.B. (G) *Plectocomiopsis geminiflora*. Photo: B.G.K. (**H**) Delimitation of biogeographic regions in (A).

Recent diversification explains extant rattan diversity

Using the phylogenetic tree and fossil evidence, we conducted four dating analyses and compared 66 biogeographic models to infer the divergence times and ancestral ranges of the Asian rattans (tables S8 and S9 and figs. S8 and S9). The age estimates of these analyses were highly congruent (figs. S10 and S11), indicating an Early Cretaceous origin (stem node age) of the lineage at 126 (range, 122 to 232) Ma (mean age of analysis "MetCal 127.4" and 95% confidence interval across all analyses), soon followed by divergence into tribes Calameae and Eugeissoneae at 123 (118 to 214) Ma (fig. S10 and table S10). The earliest ancestral range that could be inferred unambiguously was of the most recent common ancestor of subtribes Calaminae and Plectocomiinae (Fig. 1A and fig. S12), which diverged in Indochina at 75 (69 to 106) Ma (table S10) and eventually gave rise to 436 (87%) extant rattan species. After initial diversification in Indochina, the rattans colonized Borneo at 42 (38 to 69) Ma, coinciding with the onset of perhumid tropical climates on the island (*37*). During the convergence of the Sunda and Sahul continental shelves, the lineage began to spread across the entire Asian tropics, starting with dispersal from Borneo to Sulawesi at 32 (29 to 53) Ma (Fig. 1A). This finding corroborates the hypothesis that the Sunda-Sahul convergence facilitated a major biotic exchange between both areas (*3*). India and Australia, which have had fossil-documented rattan floras since the Paleocene and Late Cretaceous, respectively (*35*, *38*) (Fig. 1, B and C), were inferred to have been colonized by extant rattan species only during the Miocene (23 to 5.3 Ma). This implies that the early rattan floras of India and Australia went extinct and did not directly contribute to the current rattan diversity in these regions. Consistent with this, a lineages-through-time analysis indicated that 90% of extant Asian rattans diverged within the past 31 (28 to 51) Ma (fig. S11 and table S11), corroborating a fossil-inferred steady increase in the diversity of Calaminae throughout the Cenozoic (*25*) and suggesting that recent diversification can explain the present high species richness of rattan palms.

An extended classification of regional roles reveals biotic assembly of rattan diversity

To elucidate how the Asian tropics may have attained their present species richness across their complex geographic landscape, we inferred speciation and dispersal histories of the rattan palms using biogeographic stochastic mapping (Fig. 2). Our analyses indicated that most (71.4%) of their extant diversity originated from speciation within individual regions of the Asian tropics, including major islands and island groups (Fig. 3A). The remaining diversity resulted either from dispersal between regions directly followed by speciation (22.6%), such as after the colonization of isolated oceanic islands, or from vicariance (6.0%), i.e., the division of previously continuous ranges through newly formed dispersal barriers.

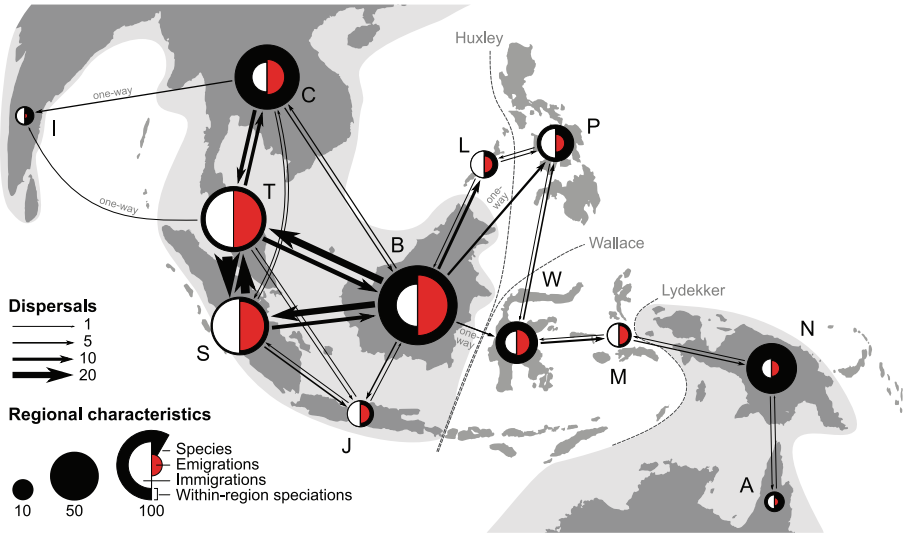


Fig. 2. Dispersal and speciation of rattan palms across the Asian tropics. Dispersal and speciation were inferred from ancestral range estimates (Fig. 1) using biogeographic stochastic mapping with 1000 iterations. Arrows denote dispersals between regions for which at least one dispersal was inferred (table S13), with width corresponding to the number of dispersals. Unidirectional dispersals are marked with the word "one-way." Areas of black circles correspond to the number of species per region, areas of white half circles show the sum of dispersals into each region (immigrations), and red half circles show the sum of dispersals out of each region (emigrations). Within-region speciation events were calculated as the difference between number of species and immigrations. The continental shelves of Asia (left) and Australia (right) are shaded in light gray. Dashed lines represent selected biogeographic boundaries that have been proposed between Asia and Australia (*44*). For region abbreviations, see Fig. 1.

Within the Asian tropics, we found strong differentiation in regional contributions to the overall species richness of rattan palms (Fig. 3B). Considering both within-region speciation [(S), the difference between species number and immigration (dispersal into the region)] and emigration (E; dispersal to other regions) relative to the regional species pool sizes (table S12), we defined four roles and classified regions accordingly: radiator (high speciation and emigration; $S \geq 0.5$, $E \geq 0.5$), incubator (high speciation but low emigration; $S \geq 0.5$, $E < 0.5$), corridor (low speciation but high emigration; $S < 0.5$, $E \geq 0.5$), and accumulator (low speciation and emigration; $S < 0.5$, $E < 0.5$). This classification expands and differentiates the widely used binary characterisation of biogeographic regions as either sources (high emigration) or sinks (high immigration) of biodiversity (*4*, *5*) by recognizing that immigration and emigration are not mutually exclusive and by taking within-region speciation into account. Radiators can be regarded as sources, accumulators as sinks, corridors as both sources and sinks, whereas incubators are neither sources nor sinks.

The rattan floras of Borneo, Indochina, New Guinea, and Sulawesi were mainly ($S \geq 0.5$) derived from within-region speciation (Fig. 3C and table S12). These four regions are established centers of biodiversity (*4*, *19*, *39*–*41*), yet the relative importance of immigration and within-region speciation in the assembly of their high species richness has, to our knowledge, not been quantified previously. Borneo was resolved as the only radiator, signifying a key role both in generating new species and distributing them across the archipelago (Fig. 3B). It acted as the only link connecting Indochina, the Thai-Malay Peninsula, Sumatra, and Java to its west, with Palawan, the Philippines, and Sulawesi to its east (Fig. 2), and had the highest numbers of within-region speciation and emigration events of all regions (table S12). This finding is consistent with the previous characterization of Borneo as a key source region of Southeast Asian biodiversity (*4*), and further suggests that the island may stand out as the single most influential region in shaping the extant diversity of rattans in the Asian tropics. By contrast, New Guinea, Indochina, and Sulawesi were identified as incubators, implying that these regions generated high diversity in relative isolation with a limited influence on other regions through emigration. The low proportion of emigration from Indochina contrasts with previous inferences performed at lower phylogenetic resolution (*4*),

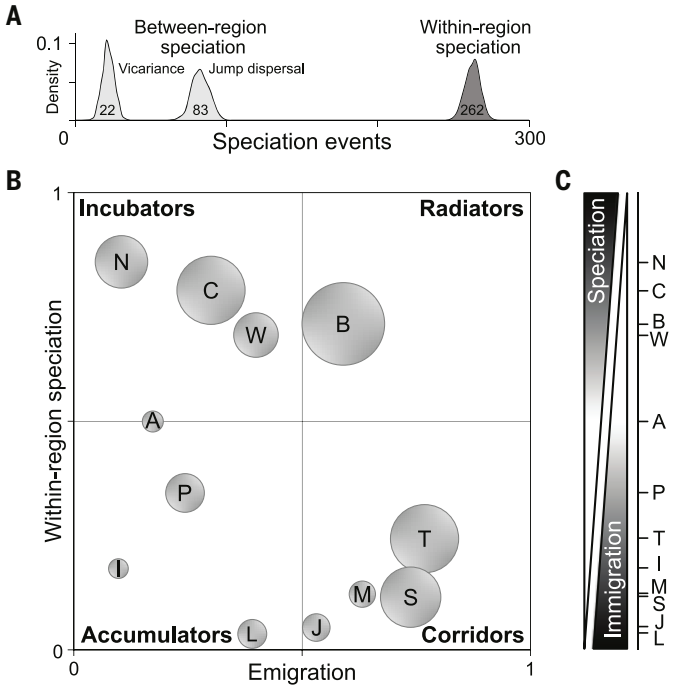


Fig. 3. Speciation, biotic assembly, and regional contributions to the overall diversity of rattans in the Asian tropics. Analyses are based on biogeographic inferences of 360 species of rattan palms (Fig. 1). (**A**) Modes of speciation. Within-region speciation accounted for most speciation events, whereas between-region speciation resulting from geographic separation, such as vicariance or "jump dispersal" (colonization of a new region directly followed by speciation), played a subordinate role. Numbers indicate speciation events. (**B**) Regional roles in shaping the overall diversity of the Asian tropics. Roles were defined based on the proportions of within-region speciation (S) and emigration (E) relative to the regional number of species (table S12). Radiators (S ≥ 0.5, E ≥ 0.5) play a key role in generating new species and distributing species across the archipelago, incubators (S ≥ 0.5, E < 0.5) generate high diversity in isolation, corridors (S < 0.5, E ≥ 0.5) produce few new species but are important in linking other regions, and accumulators (S < 0.5, E < 0.5) are mostly composed of immigrant species. Note that the proportion of immigration is the inverse of the proportion of within-region speciation. Circled areas correspond to the relative number of species per region. (**C**) Biotic assembly of regions. Region positions indicate the relative contributions of immigration and within-region speciation in the biotic assembly of these regions. For region abbreviations, see Fig. 1.

highlighting the need for further research to clarify the contribution of Indochina to the extant species diversity of the Asian tropics.

The remaining regions had low (S < 0.5) proportions of within-region speciation and were dominated by immigration (Fig. 3C and table S12). The Thai-Malay Peninsula, Sumatra, Java, and Maluku were revealed as corridors, indicating an important function in connecting neighboring regions despite a small direct contribution to overall species diversity through within-region speciation (Fig. 3B). Our inferences did not provide evidence for a dispersal barrier between Borneo and Sumatra, which has been proposed based on floristic distribution patterns (*42*). India, Palawan, the Philippines, and Australia were recovered as accumulators, suggesting that they had a relatively low influence on the extant rattan diversity of the Asian tropics and currently largely serve as repositories of species diversity generated elsewhere.

We inferred a total of 262 dispersal events across the Asian tropics (table S13), most of which (68.6%) were range expansions. On the Sunda shelf, we found a dense web of 167 dispersals closely linking the floras of Borneo, Sumatra, the Thai-Malay Peninsula, and Indochina (Fig. 2 and table S13). By contrast, regions east of Sunda were connected by a linear string of only 20 dispersals between Sulawesi, Maluku, New Guinea, and Australia, validating the hypothesized function of Sulawesi and Maluku as stepping stones between Sunda and Sahul (*3*). Dispersals from Sunda to Sahul were 2.5-fold more frequent than dispersals in the opposite direction, which is consistent with established patterns of biotic exchange across the archipelago in other wet tropical plant and animal lineages (*17*, *43*).

Regional variation in biotic assembly correlates with area and isolation

To test whether the inferred biodiversity patterns (Figs. 2 and 3) could be explained by the geography of the Asian tropics, we investigated the correlation of speciation and dispersal with region size and isolation (Fig. 4) (*33*). The number of dispersals decreased with increasing distance between regions (R^2 = 0.41, P = 0.001, n = 66) (Fig. 4A), and dispersal almost exclusively occurred between directly neighboring regions (Fig. 2), reflecting the well-established distance dependence of dispersal in plant and animal lineages (*11*, *13*). The only cases of dispersals bypassing intermediate regions were found on the Sunda shelf (Fig. 2), where direct land dispersal between the concerned regions was possible during most of the Cenozoic (*4*, *37*). Water barriers were also found to impede dispersal, with more dispersals inferred between regions that were continuously or regularly connected by land than between regions permanently separated by water (P < 0.001) (Fig. 4B). This isolating effect of permanent water barriers is reflected by the multitude of biogeographic boundaries that have been proposed between Sunda and Sahul (Fig. 2), such as Wallace's Line (*14*, *44*, *45*). Moreover, our results for rattans, which are dependent on dispersal by birds and mammals (*23*, *29*), indicate that none of these biogeographic boundaries are absolute but that all have been repeatedly crossed both by plants and animals.

The numbers of species and within-region speciation events increased with region size (R^2 = 0.52, P = 0.028, n = 9; R^2 = 0.70, P = 0.005, n = 9) (Fig. 4, C and D). This can be explained by the area dependency of topographic and habitat diversity, population size, and potential for within-region geographic isolation of populations, all of which are linked to speciation (*12*). The proportion of within-region speciation also increased with the distance to the next closest region (R^2 = 0.57, P = 0.019, n = 9) (Fig. 4E), which may be due to geographic isolation promoting speciation by reducing gene flow (*10*, *13*, *46*) and is consistent with the negative relationship between dispersal and distance observed here (Fig. 4A).

Overall, these results (Fig. 4) are consistent with island biogeographic theory (*10*, *13*, *47*). This indicates that biotic assembly of rattan diversity through speciation and dispersal (Figs. 2 and 3) was shaped by simple biogeographic principles acting across the complex island landscape of the Asian tropics (Fig. 4).

Conclusions

The evolutionary origins of the Asian tropics have long remained unclear because of their geographic complexity and the scarcity of phylogenetic data (*2*, *3*, *18*–*20*). Our species-level analyses of the rattan palms (Figs. 1 and 2) indicate that their diversity in the Asian tropics originated through the interaction of differentiated regional contributions to overall diversity determined by each region's geographic characteristics (Figs. 3 and 4). The regional roles revealed for rattans may therefore be

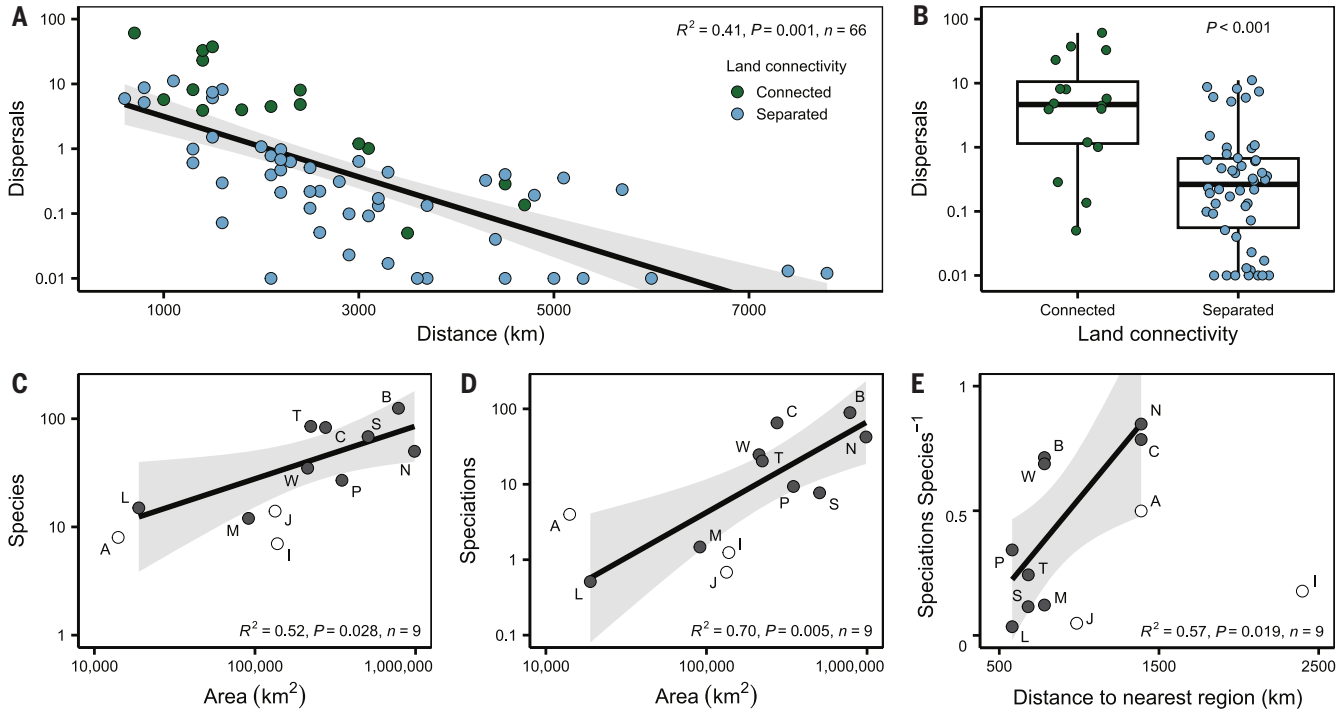


Fig. 4. Geographic correlates of dispersal and speciation of rattan palms in the Asian tropics. Dispersal and within-region speciation were inferred from ancestral range estimates (Fig. 1) averaged across 1000 iterations using biogeographic stochastic mapping. Regression lines are shown in black with 95% confidence intervals in gray. (**A**) Sum of bidirectional dispersals between region pairs by centroid distance between regions. Regression analysis using Mantel test to account for autocorrelation. (**B**) Dispersals by land connectivity. "Connected" indicates regions continuously or regularly connected by land; "Separated" indicates regions permanently separated by deep seas. Means were compared using Mann-Whitney *U* test. Boxplot shows the mean (middle horizontal line) and first and third quartiles (box boundaries), with vertical lines on each side extending from the box boundaries up to the 1.5 interquartile range and superimposed with individual data points. Because all region pairs with a distance ≥5000 km are separated (Fig. 4A), we conducted additional analyses excluding these region pairs, with differences between connected and separated regions remaining significant at $P < 0.001$. (**C**) Species by region size. Because rattan palms are constrained to wet tropical climates (*23*), we used the area per region with perhumid or seasonal tropical climates as a proxy for region size (*33*). (**D**) Within-region speciation events by region size. (**E**) Proportion of within-region speciation events relative to regional species number by centroid distance to next closest region. Because of young region ages or likely recent extinctions in Australia, India, and Java (unfilled points), which may have resulted in less time for speciation, we conducted analyses [(C) to (E)] with and without these three regions (*33*). Results of the statistical tests in (C) to (E) are given for analyses with these three regions excluded. When including all regions, results remained significant for (C) (R^2 = 0.54, P = 0.006, n = 12) and (D) (R^2 = 0.48, P = 0.012, n = 12), but not for (E) (R^2 = 0.03, P = 0.558, n = 12). Analyses are based on present-day region sizes and distances because the best biogeographic model was a time-uniform model based on current geography (table S9). For region abbreviations, see Fig. 1.

broadly generalizable to other plant and animal lineages.

Further research will be required to explore how these geography-driven patterns of speciation and dispersal are modulated by biotic factors such as ecological characteristics (e.g., soil preferences), species interactions (e.g., competition, predation, and mutualisms such as seed dispersal), and extinction (*10*, *12*, *47*). For a comprehensive understanding of the origins of tropical Asian biodiversity, densely sampled phylogenetic analyses of other species-rich lineages are needed to account for all dimensions of biodiversity, including life forms (e.g., plants and animals), habitats (e.g., wet and dry, montane and lowland, terrestrial and marine), dispersal modes (e.g., animal or wind dispersal), and lineage ages (old and young). Our regional classification can be applied to the study of biotic assembly in different lineages, geographic contexts, and spatial scales, so that a more general understanding of the evolution of biodiversity in the Asian tropics and other species-rich areas may be allowed to emerge.

REFERENCES AND NOTES

1. A. R. Wallace, *Tropical Nature, and Other Essays* (Macmillan, 1878).
2. W. L. Eiserhardt, T. L. P. Couvreur, W. J. Baker, *New Phytol.* **214**, 1408–1422 (2017).
3. R. M. Kooyman *et al.*, *Annu. Rev. Ecol. Evol. Syst.* **50**, 119–143 (2019).
4. M. de Bruyn *et al.*, *Syst. Biol.* **63**, 879–901 (2014).
5. A. Antonelli *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 6034–6039 (2018).
6. P. H. Raven *et al.*, *Sci. Adv.* **6**, eabc6228 (2020).
7. P. Weigelt, W. Jetz, H. Kreft, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15307–15312 (2013).
8. R. Hall, *Annu. Rev. Earth Planet. Sci.* **45**, 331–358 (2017).
9. W. P. Schellart, Z. Chen, V. Strak, J. C. Duarte, F. M. Rosas, *Nat. Commun.* **10**, 4480 (2019).
10. E. O. Wilson, R. H. MacArthur, *The Theory of Island Biogeography* (Princeton Univ. Press, 1967).
11. J. S. Clark, M. Silman, R. Kern, E. Macklin, J. HilleRisLambers, *Ecology* **80**, 1475–1494 (1999).
12. Y. Kisel, L. McInnes, N. H. Toomey, C. D. L. Orme, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **366**, 2514–2525 (2011).
13. L. Valente *et al.*, *Nature* **579**, 92–96 (2020).
14. A. R. Wallace, *J. Proc. Linn. Soc. Lond. Zool.* **4**, 172–184 (1860).
15. R. Schuster, *Bot. Rev.* **38**, 3–86 (1972).
16. S. T. Marsh, N. A. Brummitt, R. P. J. de Kok, T. M. A. Utteridge, *Blumea* **54**, 103–108 (2009).
17. A. Skeels *et al.*, *Science* **381**, 86–92 (2023).
18. A. V. Rudbeck *et al.*, *Ecography* **2022**, e06142 (2022).
19. D. J. Lohman *et al.*, *Annu. Rev. Ecol. Evol. Syst.* **42**, 205–226 (2011).
20. R. T. Pennington, Q. C. B. Cronk, J. A. Richardson, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **359**, 1455–1464 (2004).
21. J. S. Strijk *et al.*, *PLOS ONE* **15**, e0232936 (2020).
22. C. J. Raxworthy, B. T. Smith, *Trends Ecol. Evol.* **36**, 1049–1060 (2021).
23. J. Dransfield *et al.*, *Genera Palmarum – the Evolution and Classification of Palms* (Royal Botanic Gardens Kew, 2008).
24. M. M. Harley, R. J. Morley, *Rev. Palaeobot. Palynol.* **85**, 153–182 (1995).
25. J. Y. Lim *et al.*, *Glob. Ecol. Biogeogr.* **31**, 425–439 (2021).
26. A. Henderson, *Phytotaxa* **445**, 1–656 (2020).
27. B. G. Kuhnhäuser *et al.*, *Mol. Phylogenet. Evol.* **157**, 107067 (2021).
28. Royal Botanic Gardens Kew, "Plants of the world online" (2024); http://www.plantsoftheworldonline.org.
29. S. Zona, A. Henderson, *Selbyana* **11**, 6–21 (1989).

30. S. Zona, "Additions to 'A review of animal-mediated seed dispersal of palms'" (2006); http://www.virtualherbarium.org/palms/psdispersal.html.
31. W. D. Kissling *et al*., *Sci. Data* **6**, 178 (2019).
32. T. L. P. Couvreur *et al*., *Front. Genet.* **5**, 452 (2015).
33. Materials and methods are available as supplementary materials.
34. D. Burger, *Rev. Palaeobot. Palynol.* **65**, 153–163 (1990).
35. M. K. Macphail, G. J. Jordan, *Pap. Proc. R. Soc. Tasman.* **149**, 23–28 (2015).
36. W. J. D. Iles, S. Y. Smith, M. A. Gandolfo, S. W. Graham, *Bot. J. Linn. Soc.* **178**, 346–374 (2015).
37. R. J. Morley, *J. Trop. Ecol.* **34**, 209–234 (2018).
38. S. Parmar *et al*., *Rev. Palaeobot. Palynol.* **313**, 104890 (2023).
39. E. F. A. Toussaint *et al*., *Nat. Commun.* **5**, 4001 (2014).
40. V. S. F. T. Merckx *et al*., *Nature* **524**, 347–350 (2015).
41. R. Cámara-Leret *et al*., *Nature* **584**, 579–583 (2020).
42. J. W. F. Slik *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **108**, 12343–12347 (2011).
43. D. M. Crayn, C. Costion, M. G. Harrington, *J. Biogeogr.* **42**, 11–24 (2015).
44. G. G. Simpson, *Proc. Am. Philos. Soc.* **121**, 107–120 (1977).
45. J. R. Ali, L. R. Heaney, *J. Biogeogr.* **50**, 32–40 (2023).
46. J. A. Coyne, H. A. Orr, *Speciation* (Sinauer, 2004).
47. R. J. Whittaker, J. M. Fernández-Palacios, T. J. Matthews, M. K. Borregaard, K. A. Triantis, *Science* **357**, eaam8326 (2017).
48. B. G. Kuhnhäuser *et al*., Raw sequence data for: Island geography drives evolution of rattan palms in tropical Asian rainforests, European Nucleotide Archive (2024); https://www.ebi.ac.uk/ena/browser/view/PRJEB40689 and https://www.ebi.ac.uk/ena/browser/view/PRJEB60876.
49. B. G. Kuhnhäuser *et al*., Data and scripts for: Island geography drives evolution of rattan palms in tropical Asian rainforests, Zenodo (2025); https://doi.org/10.5281/zenodo.14391726.

ACKNOWLEDGMENTS

Plant material was provided by Royal Botanic Gardens Kew (UK), New York Botanical Garden (US), Naturalis (Netherlands), J. T. Adorador (University of the Philippines at Los Baños), Herbarium Universitatis Florentinae (Italy), Herbarium Bogoriense (Indonesia), L.H. Bailey Hortorium Herbarium (US), Forest Research Institute Malaysia (Malaysia), Missouri Botanical Garden (US), Singapore Herbarium, Australian National Herbarium, Papua New Guinea National Herbarium, Herbier National (France), Sarawak Herbarium (Malaysia), United States National Herbarium (US), Herbarium of the Arnold Arboretum (US), Herbarium of Aarhus University (Denmark), Bangkok Herbarium (Thailand), Natural History Museum (UK), Fairchild Tropical Botanic Garden (US), Conservatoire et Jardin botaniques de la Ville de Genève (Switzerland), Swedish Museum of Natural History and Á. Cano (University of Cambridge). The Sarawak Forestry Corporation facilitated field work in Borneo [permits (207)JHS/NCCD/600-7/2/107 and WL107/2018]. L. Csiba, R. Cowan, P. Malakasi, and D. Devey provided technical support. A. Antonelli, T. Utteridge, B. Lopez, and five anonymous reviewers gave feedback on the manuscript. Santos Ltd. and the previous operator, Oil Search Ltd., provided permission to publish the descriptions and locations of the fossils described here. Computational analyses were conducted using facilities of the Royal Botanic Gardens Kew and The James Hutton Institute and NIAB ("UK's Crop Diversity Bioinformatics HPC," BBSRC grant BB/S019669/1). **Funding:** This work was supported by Natural Environment Research Council grant NE/L002612/1 (B.G.K.); a Clarendon Scholarship by the Clarendon Fund and Brasenose College (B.G.K.); an Emily Holmes Memorial Scholarship by the Amar-Franses & Foster-Jenkins Trust (B.G.K.); the Garfield Weston Foundation under the Global Tree Seed Bank (S.B.); NERC Independent Research Fellowship grant NE/S014470/1 (G.C.); UKRI Frontier Research grant EP/X026868/1 (G.C.); VILLUM FONDEN research grant 00025354 (W.L.E.); Bentham-Moxon Trust grant BMT1-2017 (B.G.K., W.J.B.); Bentham-Moxon Trust grant BMT36-2018 (B.G.K., W.J.B.); IKEA of Sweden AB (W.J.B.); a National University of Singapore startup grant (J.Y.L.); and the Calleva Foundation (W.J.B.). **Author contributions:** Conceptualization: B.G.K., G.C., J.Y.L., S.B., S.J.H., W.J.B., W.L.E.; Funding acquisition: B.G.K., S.J.H., W.J.B.; Investigation: A.H., B.G.K., C.D.B., J.D., R.J.M., R.J.S., W.J.B.; Methodology: B.G.K., C.D.B., R.J.M., R.J.S., S.B., W.J.B., W.L.E.; Resources: A.H., B.G.K., C.G., H.R., S.J., W.J.B.; Project administration: B.G.K., S.J.H., W.J.B.; Supervision: G.C., S.B., S.J.H., W.J.B., W.L.E.; Visualization: B.G.K., R.J.M.; Writing – original draft: B.G.K., R.J.M.; Writing – review & editing: A.H., B.G.K., G.C., H.R., J.Y.L., R.J.M., R.J.S., S.B., S.J.H., W.J.B., W.L.E. All authors approved the final manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Raw sequence data have been deposited in the European Nucleotide Archive under project numbers PRJEB40689 and PRJEB60876 (*48*). Other data and scripts, as well as supplementary figures and tables, have been deposited on Zenodo (*49*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adp3437
Materials and Methods
Supplementary Text
Figs. S1 to S12
Tables S1 to S13
References (*50*–*134*)
MDAR Reproducibility Checklist

Submitted 24 April 2024; accepted 8 January 2025
10.1126/science.adp3437

COSMIC DUST

Three-dimensional maps of the interstellar dust extinction curve within the Milky Way galaxy

Xiangyu Zhang*† and Gregory M. Green†

Interstellar dust grains cause extinction (absorption and scattering) of light from background astronomical sources. The spectral shape of the extinction curve depends on the dust composition. We used low-resolution optical spectra to measure the extinction curve of 130 million stars. By inverting these data, we mapped the extinction curve parameter *R*(*V*) within the Milky Way in three dimensions and within the Magellanic Clouds in two dimensions. These maps provide improved extinction corrections for astronomical observations. We find that *R*(*V*) varies with extinction, consistent with dust grains growing by accretion in low-extinction regions and by coagulation in higher-extinction regions. Star-forming regions have high *R*(*V*) values, indicating either preferential destruction of small dust grains or additional supply of large dust grains in those regions.

Interstellar dust consists of microscopic solid grains and is ubiquitous in the interstellar medium (ISM). At ultraviolet (UV), optical, and near-infrared (NIR) wavelengths, dust scatters and absorbs light, thereby dimming, reddening, and polarizing the light from background sources. The combination of absorption and scattering by dust causes extinction, quantified as

$$A(\lambda) \equiv 2.5\log_{10}\left(\frac{f_0(\lambda)}{f_{\rm obs}(\lambda)}\right) \quad (1)$$

where λ is the wavelength, f_0 is the intrinsic flux (before extinction), and $f_{\rm obs}$ is the observed flux (after extinction). The spectral shape of A is referred to as the extinction curve.

The shape of the extinction curve depends on the optical properties (which depend on composition) and size distribution of the dust grains (*1*). The chemical composition of dust is constrained by the depletion of gas-phase elements (*1*, *2*). The grain-size distribution is set by the balance between growth (accretion, coagulation) and destruction (sputtering, shattering) processes (*3*, *4*). These processes can be investigated by comparing extinction curves with other observables of the ISM (*5*–*7*). During the evolution of a galaxy, the grain-size distribution changes, causing the extinction curve to vary both spatially throughout the galaxy and in time (*8*–*10*).

Galaxies and Cosmology Department, Max Planck Institute for Astronomy, Heidelberg, Germany.
***Corresponding author. Email: xzhang@mpia.de**
†These authors contributed equally to this work.

Within the Milky Way galaxy, the optical-NIR extinction curves of different sightlines can be described by a family of model curves, parameterized by $R(V) \equiv A(V)/[A(B) - A(V)]$, where B and V are standard blue and green optical filters, respectively (*11*–*13*). $R(V)$ therefore quantifies the slope of the extinction curve at optical wavelengths. The dust extinction curve is typically studied using spectroscopy of hot young stars (OB stars), which have smooth and well-understood intrinsic spectra (*11*, *14*). Large spectroscopic and photometric surveys provide statistical measurements of the extinction curve for a larger number of sightlines. $R(V)$ has been measured for samples of ~37,000 stars (*15*) and ~3 million stars (*16*). Such large samples can be used to map how $R(V)$ varies within the Milky Way, which has shown variations on kiloparsec scales (*17*).

Adopted datasets

We investigated how $R(V)$ varies using archival low-resolution, flux-calibrated blue photometer and red photometer (BP and RP, collectively referred to as XP) spectra collected by the Gaia space telescope (*18*–*21*). The XP spectra provide flux measurements at 61 wavelengths, many more than broadband photometric surveys, which is sufficient spectral resolution to determine the shape of the extinction curve (*22*–*24*). Most stars with XP spectra also have Gaia parallax measurements, which can be combined with photometry to determine the distance to the star. To supplement the XP spectra, we adopted additional photometry from NIR (*25*)

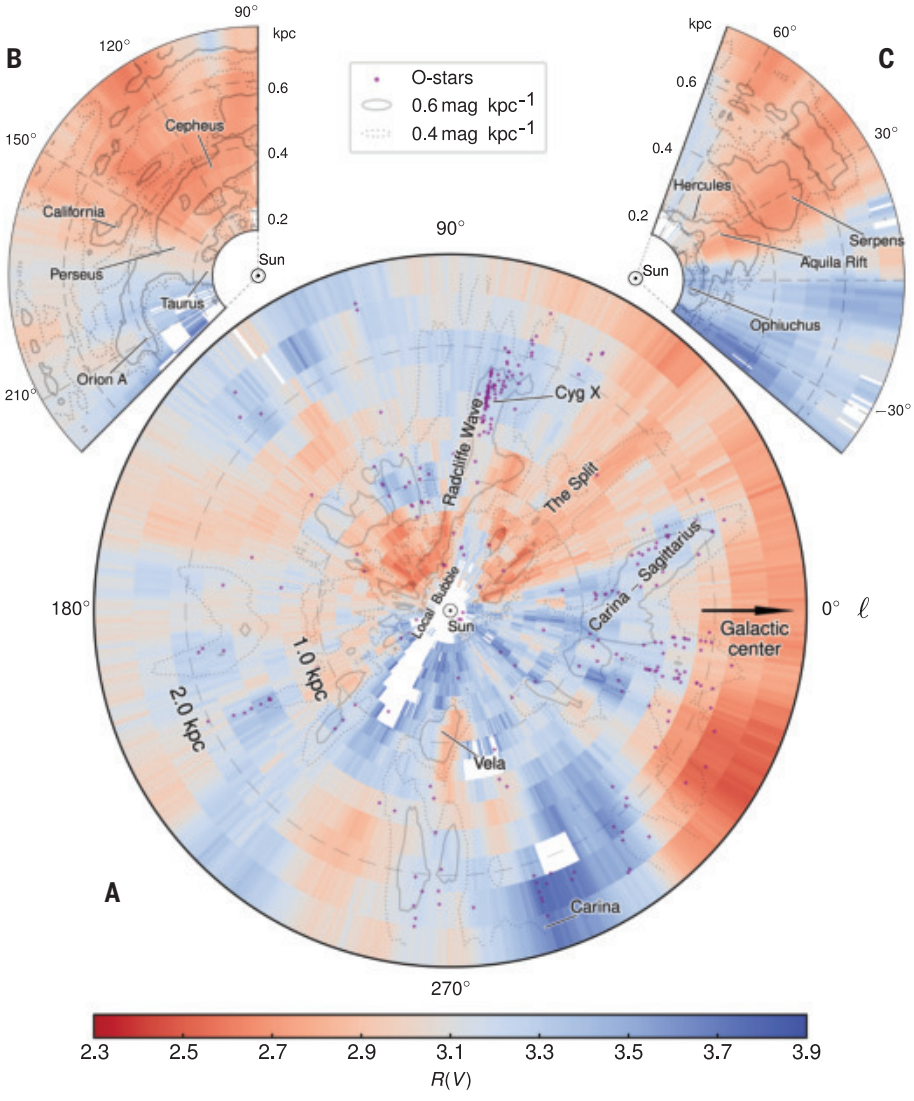
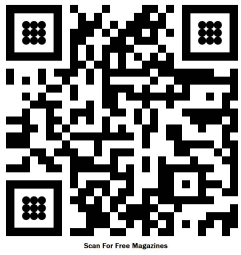


Fig. 1. Map of *R*(*V*) in the Galactic plane. (**A**) Polar map for targets within 2.6 kpc of the Sun (marked with the ⊙ symbol) and within 400 pc above or below the plane. ℓ is the Galactic longitude. $R(V)$ (color bar) was calculated by averaging our maps of differential extinction at 440 and 550 nm. Contours enclose regions of high differential extinction at 550 nm (see legend). White regions were excluded owing to insufficient extinction [$\Delta A(550\text{ nm})/\Delta D < 0.1$ mag kpc^{-1}] to measure $R(V)$, including within the Local Bubble. Purple dots indicate the locations of O-type stars (*47*). Specific regions are labeled. (**B** and **C**) Zoomed views of the same map in two ranges of Galactic longitude, out to a distance of 0.8 kpc.



and mid-infrared (MIR) (*26*) surveys, extending our wavelength coverage out to 4.6 μm.

We used machine learning to determine both a stellar model and a dust extinction curve along each sightline from the Gaia data (*27*). The machine learning model was trained on a small subset (~1%) of the Gaia XP catalog that has matching higher-resolution spectroscopy from the Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) surveys (*28*). The LAMOST data allowed the stellar parameters of the training set to be determined independently of the dust extinction parameters. The training set provides sufficient coverage for stars with effective temperatures of ~4000 to 8000 K (*27*).

Stellar model and quality cuts

We constructed a forward model that predicts the observed flux as a function of stellar and dust parameters. The intrinsic stellar spectrum is described by a neural network, which maps atmospheric parameters (effective temperature, metallicity, and surface gravity, which we collectively denote as θ) to the intrinsic flux ($\mathbf{F}_\lambda$) at each wavelength. The stellar flux is dimmed according to its distance, which we represent using stellar parallax (ϖ). In contrast to our previous work (*22*), we allowed the shape of the extinction curve, $\mathbf{R}(\xi)$, to vary within a family of curves (Eq. 2 and fig. S1A) controlled by a single fitting parameter ξ, which is allowed to vary between stars. ξ functions similarly to $R(V)$ in previous dust models (*11*–*13*). The total amount of extinction is quantified by a fitting parameter E. We implemented the model in the machine learning programming package TensorFlow 2 (*29*). Our forward model maps the stellar and dust parameters to the observed stellar spectra using the equation

$$\mathbf{f}_\lambda(\theta, \varpi, E, \xi) = \mathbf{F}_\lambda(\theta)\varpi^2 e^{-E\mathbf{R}(\xi)} \quad (2)$$

where the mapping from ξ to the extinction curve, $\mathbf{R}(\xi)$, is given by

$$\ln \mathbf{R}(\xi) = \ln(\mathbf{R}_0) + \tanh(\xi)\Delta\ln(\mathbf{R}) \quad (3)$$

where $\ln(\mathbf{R}_0)$ and $\Delta\ln(\mathbf{R})$ are model parameters shared by all stars that were learned while training the model. Applying the tanh function to ξ limits the size of its effect on the extinction curve.

We trained this model using 2.4 million stellar-type parameter estimates from LAMOST data release 8 (*28*, *30*). We used a catalog of hot stars (*31*) as priors on θ, reddening estimates from a three-dimensional (3D) dust map (*32*) as priors on E, parallax measurements from Gaia data release 3 (*19*) as priors on ϖ, and a unit normal distribution prior on ξ. For stars that are not in the prior reddening map (*32*), we adopted values from a 2D dust map (*33*). We sampled the XP spectral flux at 10-nm increments from 392 to 992 nm, similar to previous work (*22*).

We applied our trained model to all 220 million Gaia XP spectra [augmented with the

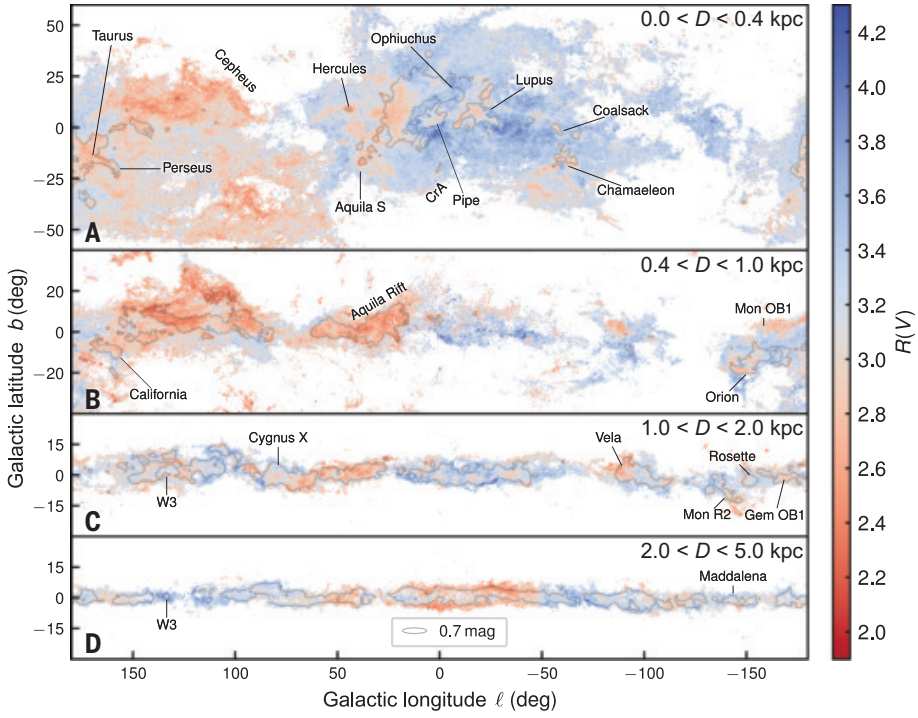


Fig. 2. Map of *R*(*V*) across the sky, in discrete distance bins. $R(V)$ (color bar) was calculated equivalently to Fig. 1 and projected onto the plane of the sky in four distance bins. $R(V)$ maps are shown in the distance slices (**A**) 0.0 to 0.4 kpc, (**B**) 0.4 to 1.0 kpc, (**C**) 1.0 to 2.0 kpc, and (**D**) 2.0 to 5.0 kpc. White areas are excluded owing to having $\Delta E/\Delta D < 0.1$ mag kpc^{-1} or $\Delta E < 0.1$ mag, where ΔE is the difference of extinction (E) in adjacent distance bins. Contours enclose high-extinction regions with $\Delta A(550\text{ nm}) > 0.7$ mag. Labels indicate the positions of dense molecular cloud complexes (*48*).

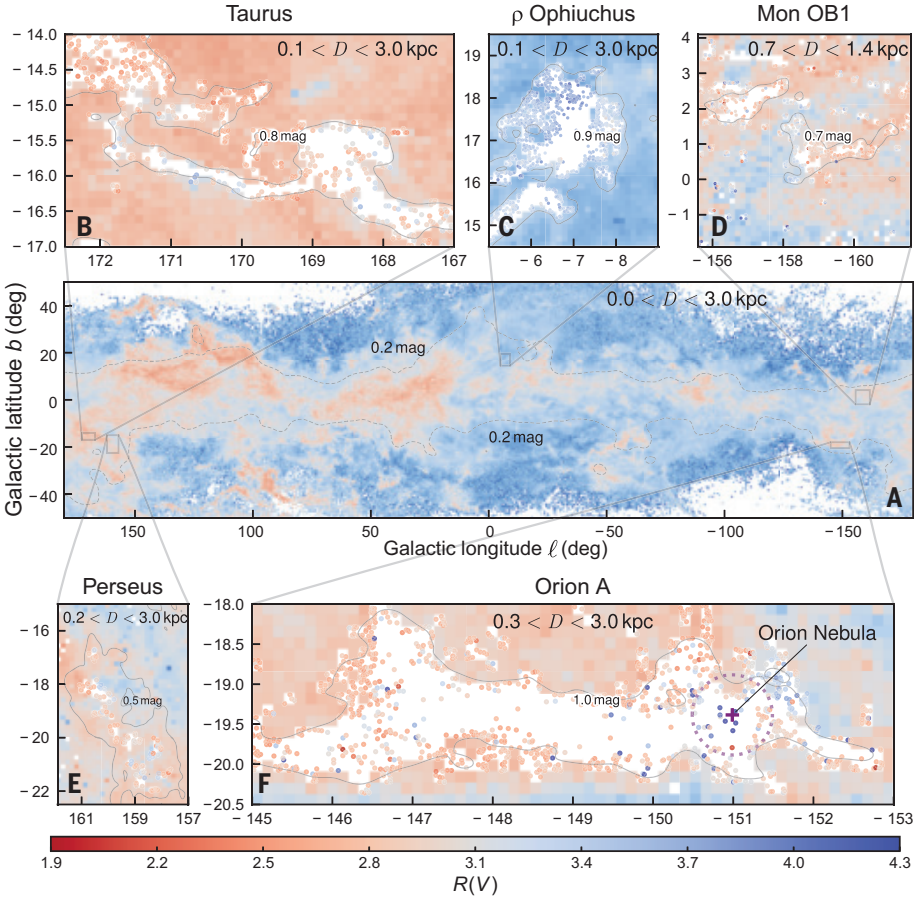


Fig. 3. Sky maps of *R*(*V*) around five dense clouds. (**A**) $R(V)$ map using stars with $E > 0.1$ mag at distances between 0 and 3 kpc. (**B** to **F**) Zoomed views of five dense clouds, with $R(V)$ calculated using stars in distance ranges (labeled in each panel) that contain the selected cloud, determined from a previous 3D dust map (*32*). Foreground extinction has been removed from each panel by subtracting maps of the average extinction of stars in front of each cloud. Gray contours enclose regions of high extinction (labels indicate the values of E), taken from the previous map (*32*). In highly extinguished regions with few observed stars, $R(V)$ values of individual stars are plotted as dots. In (F), the center and approximate radius of the Orion Nebula are denoted by the purple cross and circle, respectively.

NIR and MIR photometry (*25*, *26*) and Gaia parallaxes] to infer the parameters θ, ϖ, E, and ξ for each star. Because 99% of the sources with XP spectra do not have LAMOST observations, we imposed weak priors on the stellar parameters, on the basis of their distribution in our training set; this restricted the model to regions of stellar parameter space that were covered by the training set. We used the Fisher

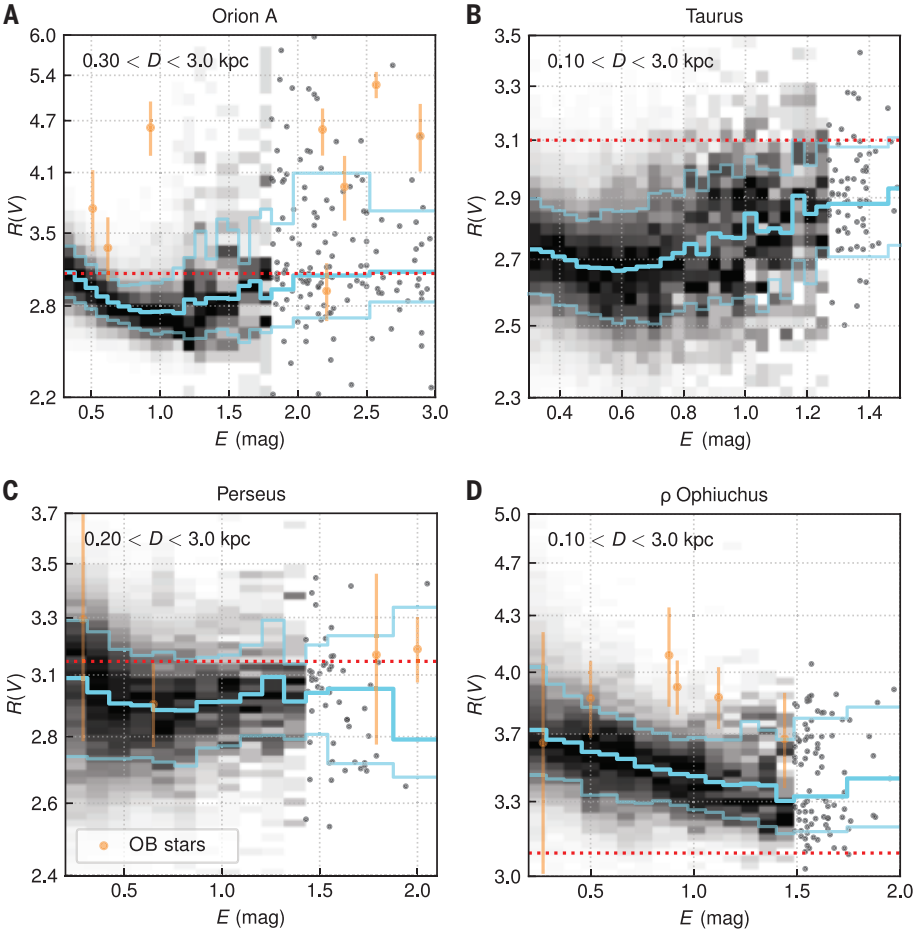


Fig. 4. The dependence of *R(V)* on extinction in four clouds. Shading in each panel shows the conditional distribution of $R(V)$ for each value of dust extinction E at a distance of 3 kpc, as predicted independently using a 3D dust extinction map (*32*). White indicates zero probability density and black corresponds to the maximum probability density in each extinction bin. Data are shown for four clouds: (**A**) 15,147 stars in Orion A, (**B**) 19,996 stars in Taurus, (**C**) 50,120 stars in Perseus, and (**D**) 44,830 stars in ρ Ophiuchus. For each cloud, we selected stars within a distance range (labeled in each panel) chosen to select stars that are either embedded in or behind the cloud. At high extinctions, where there are fewer than 20 stars in each bin, gray dots indicate individual stars. Cyan lines mark the 16th, 50th, and 84th percentiles of the $R(V)$ distribution in each extinction bin. Red dotted lines show the average Milky Way $R(V)$ of ~3.1. Yellow circles indicate $R(V)$ measurements (error bars show 1σ uncertainties) of individual hot stars from previous studies (*11*, *14*).

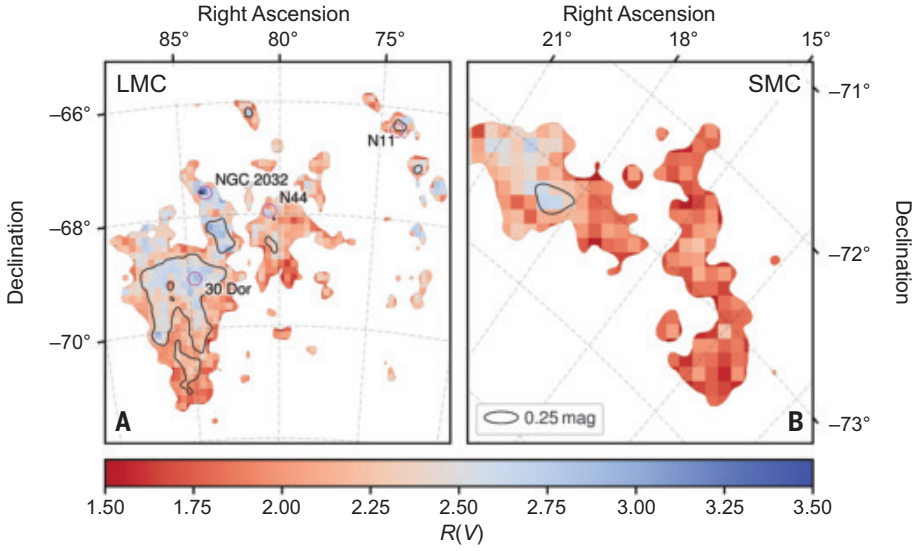


Fig. 5. Maps of *R(V)* in the Magellanic Clouds. $R(V)$ (color bar) mapped in (**A**) the LMC and (**B**) the SMC. Regions with $A(550\ \text{nm}) < 0.15$ mag have been masked out. Foreground extinction in the Milky Way has been subtracted (see text). Magenta circles in (A) indicate star-forming regions in the LMC, most of which have higher $R(V)$ than the LMC average. Dashed gray lines are a coordinate grid.

information matrix to propagate observational uncertainties into the stellar parameter estimates. We then converted the estimates of ξ into $R(V)$ for each star (*27*).

We then imposed quality cuts on the model output, to ensure that the extinction curve was sufficiently well constrained (*27*). We built three classifiers to set thresholds on the reliability of our estimates on effective temperature (T_{eff}), metallicity ([Fe/H]), and surface gravity ($\log g$) (*27*). For the 130 million stars that passed our quality cuts, we obtained median 1σ uncertainties of ±130 K in T_{eff}, ±0.03 mag in E, and ±0.18 in $R(V)$.

The learned family of extinction curves is consistent with those derived using higher-resolution spectra (see supplementary text in the supplementary materials). Our $R(V)$

measurements have a monotonic correspondence with previous work (see supplementary text).

Mapping the distribution of $R(V)$

We generated 3D maps of $R(V)$ from our extinction estimates of 130 million stars. We first calculated sky maps of the average extinction at 440 and 550 nm (the approximate central wavelengths of B and V bands) at different distances. We then calculated the value of $R(V)$ in the 3D space (voxel) between adjacent distance bins and angular positions, using $R(V) = \Delta A(V)/[\Delta A(B) - \Delta A(V)]$, where $\Delta A(B)$ and $\Delta A(V)$ are the difference of extinction of adjacent distance bins in B and V bands.

Figure 1 shows the projection of our $R(V)$ map onto the Galactic plane, for locations within 2.6 kpc of the Sun and <400 pc above or below the plane. We find spatial correspondence between the value of $R(V)$ and several known local structures (*34*, *35*), which are labeled in the figure. The Carina-Sagittarius spiral arm has higher $R(V)$ than the average, while the Split feature has lower $R(V)$; both regions contain numerous molecular clouds. We ascribe this difference to the high density of OB stars and star-forming regions in Carina-Sagittarius, which are absent in the Split (*34*). Averaging over the entire 3D map within 4 kpc of the Sun, we find a median $R(V)$ in each spatial voxel of 3.11 (1σ range from 2.60 to 3.88). Integrating along each line of sight, the median $R(V)$ is 3.08 (1σ range from 2.86 to 3.28) (see supplementary text). These values are consistent with previous studies that found a mean $R(V)$ within the Milky Way of about 3.1 (*12*, *15*).

Figure 2 shows maps of $R(V)$ across the sky in four distance bins. We find that regions containing molecular clouds have low $R(V)$ values. In the most distant bin, 2 to 5 kpc, we find low $R(V)$ values toward the Galactic Center ($\ell = 0°$), which is consistent with previous photometric results (*36*).

Figure 3 shows sky maps of $R(V)$ zoomed into the positions of five dense clouds. We find low $R(V)$ in the outer, intermediate-density regions of three of these clouds (Orion A, Taurus, and part of Mon OB1). However, in the lines of sight toward the densest regions of the clouds, such as the Orion Nebula, we find higher $R(V)$ values than in the outskirts. For reasons that are unclear, the Perseus cloud shows a more complex pattern. In ρ Ophiuchus, both the central and outer regions have higher $R(V)$ than the average value in the Milky Way (~3.1), which could be due to the recent and ongoing star formation in this region (*37*–*39*). Figure S9 shows the $R(V)$ maps of ρ Ophiuchus and the adjacent region of Upper Scorpius compared with the distribution of OB stars, which are tracers of star formation (see supplementary text).

We investigated how $R(V)$ depends on extinction in four clouds (Fig. 4). For Orion A, Perseus, and Taurus, we find a U-shaped relation: $R(V)$ decreases with increasing E at low extinctions and then shows the opposite trend at high extinction ($E \gtrsim 0.75$ mag). No such minimum is seen in the ρ Ophiuchus data.

We also mapped $R(V)$ in the Large Magellanic Cloud (LMC) and Small Magellanic Cloud (SMC), two satellite galaxies of the Milky Way. The LMC and SMC both have lower metallicity than does the Milky Way and so could potentially have different dust properties. After removing foreground extinction of the Milky Way (see supplementary text), we find that the SMC has lower $R(V)$ than the Milky Way average, which is consistent with previous studies (*40*). The LMC has a wider range of $R(V)$, with higher values in some star-forming regions (Fig. 5A).

Implications for dust properties

We have identified a general trend wherein $R(V)$ is usually higher than average in regions of the ISM with low column density of dust, as traced by E (Fig. 3A), and $R(V)$ typically decreases as E increases. In the highest-extinction regions of some clouds, such as Orion A, Taurus, and Perseus (Fig. 4, A to C), we find that this trend reverses, with $R(V)$ increasing with greater extinction. Previous work has shown that $R(V)$ tends to increase in the cold, dense cores of molecular clouds (*41*, *42*). We generally observe decreasing $R(V)$ with increasing extinction in the translucent (moderate-density) ISM, before $R(V)$ increases again in the densest inner regions of clouds. Next, we considered possible physical explanations for this trend.

The two dominant mechanisms of dust grain growth in the ISM are (i) accretion of elements from the gas phase as they condense onto the surface of grains and (ii) coagulation of grains that collide and stick together. Although both processes increase the average grain size, their effects on the grain-size distribution, and thus $R(V)$, are different (*4*). Accretion tends to deposit a layer of approximately equal thickness onto each grain, so smaller grains experience a larger fractional increase in surface area than do the larger grains. The resulting cross section–weighted average grain size decreases, so the extinction curve steepens and $R(V)$ decreases. In contrast, coagulation converts pairs of small grains into single larger grains, preferentially depleting the population of small grains. This drives the grain-size distribution toward larger grains, so the extinction curve becomes flatter and $R(V)$ increases.

Because accretion involves interactions between dust grains and the surrounding gas, it occurs at lower densities than does coagulation, which requires direct collisions between dust grains. As more-refractory (low condensation temperature) elements accrete onto grains, they become depleted from the gas phase, so the accretion rate decreases (*2*). We interpret our results as indicating that accretion is the dominant mechanism of grain growth in the moderately dense translucent regions of interstellar clouds, which is consistent with the observed decrease in $R(V)$ with dust density. We suggest that, at higher densities, coagulation becomes the dominant grain growth mechanism, so $R(V)$ increases.

In the maps, we identify regions of high $R(V)$ close to locations with recent (or ongoing) star formation (see supplementary text). We observe high $R(V)$ across the entire Upper Scorpius and ρ Ophiuchus region (Fig. 3C and fig. S9), which hosts an association of OB stars and contains both ongoing and recent star formation (*37*–*39*). Two-dimensional sky projections show that this region has high $R(V)$ in both low- and high-extinction regions (Fig. 4D), which might be typical of star-forming regions. In Fig. 1, concentrations of OB stars are often surrounded by regions of high $R(V)$. For example, although much of the Radcliffe Wave region has low $R(V)$ and contains few OB stars, Cygnus X (located at one end of the wave) contains a large concentration of OB stars and has higher $R(V)$. The Split region, which contains few OB stars, has low $R(V)$. The Carina-Sagittarius spiral arm has a higher density of OB stars and also has high $R(V)$. This behavior is also apparent in the LMC, where a few star-forming regions have higher $R(V)$ than the rest of that galaxy (Fig. 5A). There are several physical mechanisms that could produce this association between star formation and high $R(V)$, including preferential destruction of small grains by supernova shocks (*8*), or the formation of large grains in dense molecular cores (*3*, *43*), which are dispersed into the translucent ISM by stellar feedback (*44*).

Previous work (*15*) found an anticorrelation between $R(V)$ and the dust emissivity spectral index β (*45*), which quantifies the spectral shape of dust thermal emission in the far infrared. We find a similar trend for low-Galactic-latitude sightlines (see supplementary text and fig. S6A). Those sightlines are complex, with multiple interstellar clouds along each one, and so cannot be used to determine whether β and $R(V)$ are anticorrelated in individual clouds. We therefore investigated these parameters in clouds at higher Galactic latitudes (*27*), finding a weaker anticorrelation between β and $R(V)$ than in previous work (fig. S6B).

Utility for extinction corrections

Many astronomical observations require correction for foreground dust extinction. Previous dust extinction maps have assumed a constant value of $R(V)$, which we estimate introduces a systematic uncertainty of ~10% in extinction [assuming the typical range of $R(V)$ we observe in the Milky Way]. Our 3D map (*46*) provides the $R(V)$ value for foreground dust as a function of distance along any given sightline

in the Milky Way, which can be incorporated into dust correction calculations to avoid this systematic uncertainty.

REFERENCES AND NOTES

1. B. T. Draine, *Annu. Rev. Astron. Astrophys.* **41**, 241–289 (2003).
2. E. B. Jenkins, *Astrophys. J.* **700**, 1299–1348 (2009).
3. C. W. Ormel, D. Paszun, C. Dominik, A. G. G. M. Tielens, *Astron. Astrophys.* **502**, 845–869 (2009).
4. H. Hirashita, *Mon. Not. R. Astron. Soc.* **422**, 1263–1271 (2012).
5. B. L. Rachford *et al*., *Astrophys. J.* **577**, 221–244 (2002).
6. B. L. Rachford *et al*., *Astrophys. J. Suppl. Ser.* **180**, 125–137 (2009).
7. D. Van De Putte, S. I. B. Cartledge, K. D. Gordon, G. C. Clayton, J. Roman-Duval, *Astrophys. J.* **944**, 33 (2023).
8. R. S. Asano, T. T. Takeuchi, H. Hirashita, T. Nozawa, *Mon. Not. R. Astron. Soc.* **432**, 637–652 (2013).
9. K.-C. Hou, H. Hirashita, K. Nagamine, S. Aoyama, I. Shimizu, *Mon. Not. R. Astron. Soc.* **469**, 870–885 (2017).
10. H. Hirashita, S. Aoyama, *Mon. Not. R. Astron. Soc.* **482**, 2555–2572 (2019).
11. K. D. Gordon *et al*., *Astrophys. J.* **950**, 86 (2023).
12. J. A. Cardelli, G. C. Clayton, J. S. Mathis, *Astrophys. J.* **345**, 245 (1989).
13. E. L. Fitzpatrick, *Publ. Astron. Soc. Pac.* **111**, 63–75 (1999).
14. L. A. Valencic, G. C. Clayton, K. D. Gordon, *Astrophys. J.* **616**, 912–924 (2004).
15. E. F. Schlafly *et al*., *Astrophys. J.* **821**, 78 (2016).
16. R. Zhang, H. Yuan, B. Chen, *Astrophys. J. Suppl. Ser.* **269**, 6 (2023).
17. E. F. Schlafly, J. E. G. Peek, D. P. Finkbeiner, G. M. Green, *Astrophys. J.* **838**, 36 (2017).
18. T. Prusti *et al*., *Astron. Astrophys.* **595**, A1 (2016).
19. A. Vallenari *et al*., *Astron. Astrophys.* **674**, A1 (2023).
20. F. De Angeli *et al*., *Astron. Astrophys.* **674**, A2 (2023).
21. P. Montegriffo *et al*., *Astron. Astrophys.* **674**, A3 (2023).
22. X. Zhang, G. M. Green, H.-W. Rix, *Mon. Not. R. Astron. Soc.* **524**, 1855–1884 (2023).
23. R. Andrae, H.-W. Rix, V. Chandra, *Astrophys. J. Suppl. Ser.* **267**, 8 (2023).
24. J. Li, K. W. K. Wong, D. W. Hogg, H.-W. Rix, V. Chandra, *Astrophys. J. Suppl. Ser.* **272**, 2 (2024).
25. M. F. Skrutskie *et al*., *Astron. J.* **131**, 1163–1183 (2006).
26. E. F. Schlafly, A. M. Meisner, G. M. Green, *Astrophys. J. Suppl. Ser.* **240**, 30 (2019).
27. Materials and methods are available as supplementary materials.
28. X.-Q. Cui *et al*., *Res. Astron. Astrophys.* **12**, 1197–1242 (2012).
29. M. Abadi *et al*., in *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation* (The USENIX Association, 2016), pp. 265–283.
30. G. Zhao, Y.-H. Zhao, Y.-Q. Chu, Y.-P. Jing, L.-C. Deng, *Res. Astron. Astrophys.* **12**, 723–734 (2012).
31. M. Xiang *et al*., *Astron. Astrophys.* **662**, A66 (2022).
32. G. M. Green, E. Schlafly, C. Zucker, J. S. Speagle, D. Finkbeiner, *Astrophys. J.* **887**, 93 (2019).
33. D. J. Schlegel, D. P. Finkbeiner, M. Davis, *Astrophys. J.* **500**, 525–553 (1998).
34. J. Alves *et al*., *Nature* **578**, 237–239 (2020).
35. R. Lallement *et al*., *Astron. Astrophys.* **625**, A135 (2019).
36. A. Udalski, *Astrophys. J.* **590**, 284–290 (2003).
37. S. Ratzenböck *et al*., *Astron. Astrophys.* **678**, A71 (2023).
38. G. Briceño-Morales, J. Chanamé, *Mon. Not. R. Astron. Soc.* **522**, 1288–1309 (2023).
39. N. Miret-Roig *et al*., *Astron. Astrophys.* **667**, A163 (2022).
40. K. D. Gordon, G. C. Clayton, K. A. Misselt, A. U. Landolt, M. J. Wolff, *Astrophys. J.* **594**, 279–293 (2003).
41. N. L. Chapman, L. G. Mundy, S.-P. Lai, N. J. Evans, *Astrophys. J.* **690**, 496–511 (2009).
42. J. Steinacker, L. Pagani, A. Bacmann, S. Guieu, *Astron. Astrophys.* **511**, A9 (2010).
43. H. Hirashita, H. Yan, *Mon. Not. R. Astron. Soc.* **394**, 1061–1074 (2009).
44. N. Murray, E. Quataert, T. A. Thompson, *Astrophys. J.* **709**, 191–209 (2010).
45. A. Abergel *et al*., *Astron. Astrophys.* **571**, A11 (2014).
46. X. Zhang, G. M. Green, Unveiling the Milky Way dust extinction curve in 3D [Data set], Zenodo (2024); https://doi.org/10.5281/zenodo.11394477.
47. Y. Xu *et al*., *Astron. Astrophys.* **616**, L15 (2018).
48. C. Zucker *et al*., *Astron. Astrophys.* **633**, A51 (2020).
49. X. Zhang, G. M. Green, Astrozxy/gaia_XP_forward_model: XP_forward_model V1.1, Zenodo (2024); https://doi.org/10.5281/zenodo.11242348.

ACKNOWLEDGMENTS

We acknowledge discussions with A. Li, B. Hensley, and T. Henning on dust physics, and with H.-W. Rix on the work and manuscript. We thank G. Clayton, T. Zwitter, and an anonymous reviewer for constructive comments that improved the manuscript. **Funding:** G.M.G. and X.Z. are supported by a Sofja Kowalewskaja Award, granted to G.M.G. by the Alexander von Humboldt Foundation. **Author contributions:** X.Z. and G.M.G. jointly developed the forward model, wrote the software to implement it, performed the model training, and derived the stellar parameters. X.Z. and G.M.G. made equal contributions to the analysis of the results, generation of the figures, and writing of the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Our trained stellar model, catalog of parameter inferences for all 220 million stars, derived 3D map of dust $R(V)$, and example scripts are archived at Zenodo (*46*). Our software to train the model, derive the stellar and extinction parameters, and produce the maps is also archived at Zenodo (*49*). The Gaia XP spectra (*21*) were obtained from https://sdsc-users.flatironinstitute.org/~gaia/dr3/hdf5/XpContinuousMeanSpectrum/. The LAMOST spectral parameters (*30*) were obtained from https://www.lamost.org/dr8/v2.0/catalogue.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ado9787
Materials and Methods
Figs. S1 to S10
Table S1
References (*50*–*68*)

Submitted 1 March 2024; accepted 17 January 2025
10.1126/science.ado9787

SUSTAINABILITY

Geological resource production constrained by regional water availability

Kamrul Islam[1], Keitaro Maeno[1], Ryosuke Yokoi[1], Damien Giurco[2], Shigemi Kagawa[3], Shinsuke Murakami[4], Masaharu Motoshita[1]*

Although the global economy requires geological resource mining, production has substantial environmental impacts, including the use of regional available water. In this study, we shed light on the global production capacity of 32 mined geological resources, considering regional water availability as a constraint. We found that current resource mining greatly exceeds regional water constraints for several, notably copper (37% of current production exceeds available water capacity) in 2010. Changing the location of production to regions of lower water stress would alleviate current exceedances of water constraints; however, considering economic factors shows that this is not always feasible. Future demand for geological resources is expected to require a considerable increase in water consumption. Considering the constraints of water resources in geological resource production is crucial for sustainability.

Mined geological resources, for example, minerals, metals, and rocks, are essential for developing and sustaining the global economy. Since the early 2000s, geological resource extraction has risen by more than 50% (*1*), driven by increasing demand for raw materials, and this upward trend is expected to continue because of the build-up of global material stocks (*2*) and the expansion of low-carbon infrastructures, such as wind and solar energy and battery storage capacities (*3*, *4*). The extraction and processing of geological resources can lead to several adverse environmental effects, including land use changes (*5*–*10*), biodiversity loss (*11*–*13*), increased CO_2 emissions (*14*), acid mine drainage (*15*), periodic tailings dam disasters (*16*), and water pollution (*17*). The significant increase in the production of mined geological resources is a part of the "great acceleration" (*18*) arguably pushing the global socioeconomic metabolism beyond planetary boundaries (*19*, *20*), which defines a safe operating space for the current society to develop and thrive while maintaining the resilience and functioning of the earth system.

Mining and processing operations of geological resources require great amounts of water, often entering the operations from surface and underground water sources (*21*–*23*). Moreover, water use and consumption in geological resource mining and processing is a critical challenge, as it competes with water use in other production systems, such as agriculture. Mining, although constituting a small fraction of global water use (2 to 4.5% in mining-intensive countries), substantially strains regional water supplies, impacting quantity and quality (*24*). Our previous study determines sustainable water use by regional carrying capacities (RCCs), which are defined as the remaining water for humanity after

[1]Research Institute of Science for Safety and Sustainability, National Institute of Advanced Industrial Science and Technology, AIST Tsukuba West, 16-1 Onogawa, Tsukuba, Ibaraki, Japan. [2]Institute for Sustainable Futures, University of Technology Sydney, Ultimo, New South Wales, Australia. [3]Faculty of Economics, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka, Japan. [4]Department of Technology Management for Innovation, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan.
***Corresponding author. Email: m-motoshita@aist.go.jp**

securing water for ecosystems (*25*). According to the estimate based on this approach, freshwater use currently exceeds the limits of water resources at the regional level, depriving aquatic ecosystems of the water they need to endure. Therefore, geological resource production at the location where water is overexploited beyond the carrying capacity will need to reduce production (to the limit for aquatic ecosystem conservation).

In this study, we aimed to determine a sustainable capacity for geological resource production under the constraint of regional available water and identify the potential gaps between sustainable production and projected future demand. Firstly, we established datasets on water consumption intensity for producing 32 geological resources in 2010 (table S1), representing all geological resources available in the SNL database, through an extensive literature review. The SNL database provides the operational data of global mines with the largest coverage, including the mined volume of each geological resource (*26*). Linking water consumption intensity with geological resource production data from the SNL database enabled estimates of the water consumption volume for geological resource production on a global scale. Secondly, we defined the overproduction of geological resources based on the water volume consumed for geological resource production beyond the RCCs of water resources of global watersheds, which were estimated in a previous study (*25*). Thirdly, we explored the theoretical potential of alleviating the overproduction of geological resources based on three defined scenarios. Lastly, we demonstrated how water constraints may cause gaps between sustainable production and projected future demand for geological resources following socioeconomic pathways (SSPs) as future scenarios (see details in materials and methods) (*27*–*29*). The definitions of key terms in this work are available in table S2.

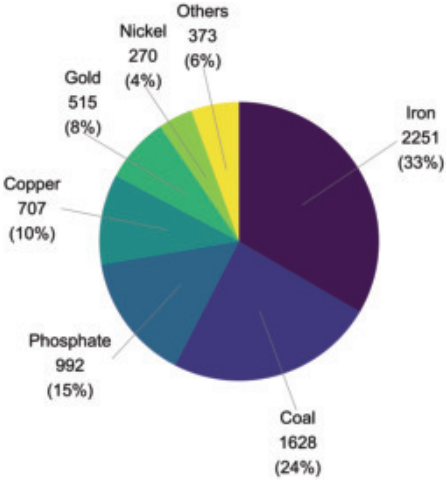


Fig. 1. Breakdown of the total water consumption (million cubic meters) for mineral production by geological resource.

Results

Water consumption for geological resource production and its spatial distribution

The total water consumption for geological resource production [including extraction, crushing, processing, and refining (see details in materials and methods)] in 2010 from the 3319 mines studied was estimated to be 6739 (±1564) million m^3. The estimated volume of water consumption for geological resource production was equivalent to 7 (±2)% of total industrial water consumption in 2010 (96,146 million m^3). Six major geological resources accounted for 94% of the total water consumption for geological resource production in the world: iron (33%), coal (24%), phosphate (15%), copper (10%), gold (8%), and nickel (4%) (Fig. 1; see table S8 for details). Iron and coal required a relatively smaller volume of water consumption per ton produced (table S3), whereas the relatively larger production volumes contribute to the dominant water consumption for these geological resources (table S4). By contrast, phosphate production was less than one-tenth that of iron and coal but resulted in a comparable amount of water consumption for its production because it is so water intensive.

Sustainability of water consumption for geological resource production

Constraints on water use depend not only on how much water is needed for geological resource production but also on how much water is available in the region. We estimated the overproduction of geological resources, namely the proportion of production inducing water consumption beyond RCC, at global mining sites aggregated into watersheds for 2010 (Fig. 2). Mines in Chile and Peru had the highest overproduction rates (25 to 100%). The scarcity of surface and groundwater in these regions has led to the use of desalinated ocean water for geological resource production (*30*). A total of 215 watersheds where 132 mines were located, accounting for 6% of the total watersheds with mines, experienced geological resource overproduction; although overproduction rates vary, close to half of these watersheds (105 out of 215 watersheds) had more than 75% overproduction rates across 64 mining sites. Most of these mines with higher overproduction rates (>75% overproduction) mined copper (22 sites), gold (14 sites), iron (11 sites), and coal (8 sites).

Top 10 geological resources in water overconsumption for their production accounts for around 98% of the total water overconsumption: iron, coal, phosphate, copper, gold, nickel, zinc, bauxite, chromite, and manganese (Fig. 3; see fig. S1 for details of all geological resources). Copper shows the largest proportion of overconsumption (37% of current production exceeding water resource capacity in 2010) to the current production, whereas iron shows the largest volume of water consumption but a lower proportion of overconsumption (9% of current production exceeding water resource capacity in 2010). Considering the volume of water overconsumption, coal shows the largest volume (382 million m^3), followed by copper (260 million m^3). This implies that the current momentum of decarbonization through an energy shift from coal may have synergistic effects on the reduction in greenhouse gas emissions and sustainable water use; a massive reduction of coal production will be needed to make a difference in the context of water stress alleviation. Meanwhile, decarbonization technologies are expected to require copper, and their large-scale implementation will increase the future demand for copper (*31*, *32*), possibly increasing its overproduction beyond the sustainable capacity of regional water resources.

Potential for alleviating overproduction

Six major geological resources showed major overproduction relative to their current production owing to the consumption of large water volumes. We found that countries with large geological resource production faced overproduction, whereas some countries with no overproduction had the capacity for additional production with respect to water resource availability (fig. S2 and table S5), indicating a large substitution potential for geological resource production to alleviate overproduction in other producing countries. Therefore, we analyzed the alleviation potential for substituting overproduction by countries with no overproduction in three defined scenarios: alleviation by substitution is limited by only production capacity (scenario 1), production capacity and market competitiveness (scenario 2), and production capacity and overproduction rate (scenario 3). Scenario 1 represents the maximum alleviation potential; others additionally consider more possible constraints in the context of economic competitiveness by prioritizing alleviation in countries with large volumes of overproduction (scenario 2) or high water stress (scenario 3) (see the detail of scenarios in table S6).

By adhering solely to the annual production capacity constraints, which are based on the water resources' carrying capacity, we succeeded in avoiding the overproduction of all six major geological resources in scenario 1; however, the production increase in substituting countries may not be feasible due to the limitations of facility capacity, labor, and other socioeconomic constraints for some geological resources, such as coal (166%),

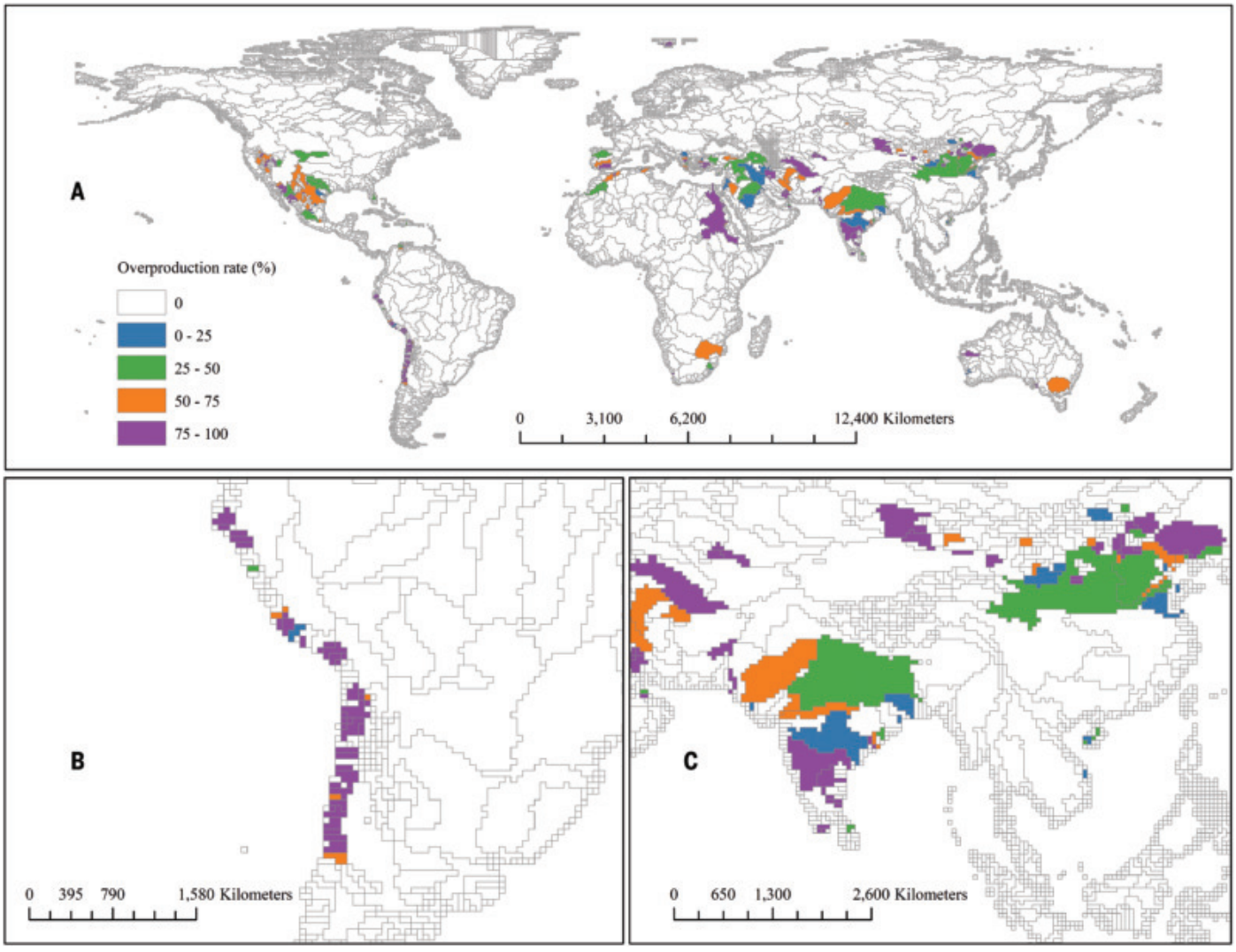


Fig. 2. Rate of overproduction of geological resources at watershed level during 2010. (**A**) The overproduction rate (%) in each watershed for global watersheds. (**B**) The zoomed-in view of mines within watersheds located in Chile and Peru. (**C**) The zoomed-in view of mines in watersheds in India and northern China.

copper (217%), gold (101%), and phosphate (83%) (fig. S3). Conversely, the substitution potential decreased when we considered the economic competitiveness of goods produced in a country in scenarios 2 and 3. For example, some proportion of the overproduction in the current scenario (Business as Usual) will still remain in scenarios 2 and 3: 33 to 42% for coal, 30 to 32% for copper, 6 to 17% for iron, 64 to 77% for gold, 93% for phosphate, and 25 to 28% for nickel (table S7). These results indicate that the theoretical potential of substituting production with limitation by only production capacity (scenario 1) was sufficiently large to alleviate overproduction, but the feasible potentials when we considered the economic competitiveness of substituting countries (scenarios 2 and 3) were insufficient to cover the overproduction of all geological resources under the current situation. Detailed information on the major substituting countries in all scenarios is provided in table S8.

Future water consumption for geological resource production

We demonstrated that the current production of some geological resources already exceeds the production capacity under the constraint of sustainable water use (Fig. 4). This situation may worsen, given the projected increase in geological resource demand in the future. Coal will gradually increase water consumption until the midcentury. After the midcentury, water consumption for coal will increase in the three scenarios (middle of the road, SSP2; regional rivalry, SSP3; fossil-fueled development, SSP5), whereas a slight decrease of water consumption occurs in the sustainability (SSP1) and inequality (SSP4) scenarios (see the details of each scenario in fig. S4). In particular, SSP5 will result in a rapid increase after the midcentury owing to the economic growth and dependency on fossil fuels. Coal shows distinct temporal trends compared with metals (copper, iron, and nickel) because metals can be stocked in society and recycled after the lifetime of the embodied products. The three major metals' production is projected to cause a rapid increase in water consumption until the midcentury, then decline with the increased recycling of these metals in all scenarios, except for SSP3, which will reduce the demand for primary metals and associated water consumption. In the regional rivalry scenario (SSP3), the population in middle-income countries will continue increasing, although other scenarios will decrease the population in middle-income countries after the midcentury. The temporal change will differ between the SSPs, although, across scenarios, the water consumption

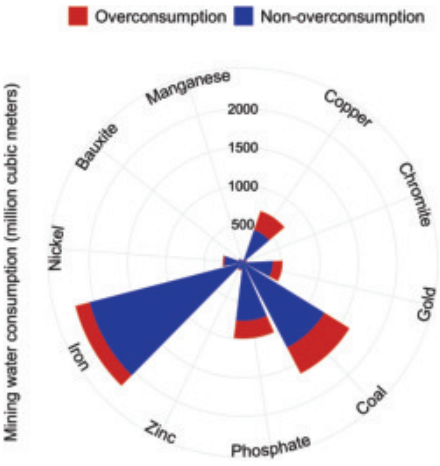


Fig. 3. Water consumption of the top 10 geological resources with the largest volume of water overconsumption. Overconsumption (red) is defined as the water volume consumed for geological resources production beyond the regional water availability (blue). The version covering all geological resources is available in fig. S1. The detailed data are available in table S11.

for producing these metals is expected to reach a higher level in the future than the consumption in 2010. Considering the sustainability scenario (SSP1), water consumption associated with the production of these metals will increase by up to 241% for copper, 119% for iron, and 239% for nickel during the midcentury compared with those for their 2010 production (Fig. 4). This highlights that water constraint for geological resource production may become more severe in the future. In this case, the crucial point is that, owing to the changing climate, future changes in water availability and associated RCC will differ worldwide (*33*). Because we could not consider this aspect when estimating the future water consumption for geological resource production, our projections could be considered to contain uncertainty in the estimates in addition to the incompleteness of the coverage of global mines.

Discussion

Our results highlight that the environmental constraints, in particular, water scarcity, must be considered when discussing the sustainable production of geological resources, in addition to factors such as geological resources' abundance in the earth's crust, energy requirements in extraction and processing, and purely economic considerations, such as the availability of labor and other inputs. Both the magnitude and water consumption intensity for geological resource production determine the overproduction of geological resources. Overproduction varies by location, allowing for the potential to fully avoid overproduction of geological resources under the constraint of sustainable water use by shifting production from water-limited to water-rich regions. However, the feasibility of substitution will be restricted by several factors, including economic realities. Apart from the factors considered in this study, such substitution of geological resource production could be limited by the operational capacity of mining and production sites, including labor, quantity, and quality of deposits in substituting mining sites as well as production cost and market price of commodities. The detailed estimation of potential substitution with the consideration of these additional factors is a complex challenge but necessary in future analyses for governments and industries to plan sustainable geological resource production. An adequate regulatory framework and regular long-run planning of geological resource production are crucial for achieving sustainable water management in geological resource production. Appropriate allocation of the rights of water users (*34*) and consideration of relevant factors, including external cost of production activities, will be an effective option to avoid water constraints on future geological resource production.

Our findings indicate that increasing demand for major geological resources in the future, owing to economic and population growth, will increase water consumption associated with their production by around triple that of the present and may worsen the overproduction of geological resources. Although some scenarios demonstrate the decrease in water consumption for metal production owing to recycling metal stocks in society, the effects can be observed only after the midcentury, as it will take some time until stocks become available for recycling (*14*). This highlights that the improvement of water intensity for geological resource production and changing the location of geological resource production are crucial solutions to reduce the pressure of the overproduction of geological resources from a short-term perspective. At the same time, the advancement of recycling technologies and systems will also contribute to sustainable geological resource use from a long-term perspective. In addition, the demand for some metals, for example, copper and nickel, is expected to rapidly increase in response to the deployment of renewable energy technologies for decarbonization (*32*). Some previous studies have already suggested that such an increasing demand for metals for renewable energy technologies will exceed the capacity of production facilities (*35*) and the reserve of metals (*31*). Our results of increasing water consumption of metal production in the future confirm and complement these concerns on the future availability of metals.

As a first attempt to explore the relationship between geological resource production and water use, we highlighted the potential influence of the RCC of water use on geological resource production. Although the primary constraints on geological resource supply are often related to extraction capabilities and funding, our findings suggest the possibility that water use can pose a limiting factor in regional water availability and aquatic ecosystem conservation. However, the results and conclusions of this study have some methodological limitations. The limitations listed in the following paragraph should be overcome in future studies to obtain more robust results.

One of the primary objectives of this study was to illustrate the global picture of the current pressure of geological resource production on regional water resources. The mining datasets were derived from the SNL database, a comprehensive database covering global mines. However, the coverage of global mines is imperfect, as shown in tables S4, S9, and S10. There is potential for updating this analysis with comprehensive and up-to-date information on global mining sites (*36*), and site-specific production and commodity-level information need to be disclosed and obtained in parallel. Moreover, the intensity of water consumption for geological resource production has been derived from extensive literature reviews; however, we could not fully consider the effects of site-specific conditions of water use, including the effects on the water flows by geological changes for mining. Water for geological resource production is recycled or reused, and desalinated water is used in some production sites located in water-stressed areas. In addition, different water flows (e.g., evapotranspiration, product incorporation, and water transfers) are relevant to geological resource production (*37*), whereas the discrimination of water flows will be needed when considering countermeasures. In fact, some of the geological resources industry have developed their own framework for water accounting that enables tracing and management of relevant water flows for mining (*38*). However, such a good practice still needs time to be widely adopted by global production sites. For a more accurate estimation, the location-, technology- and water flow–specific water intensity for geological resource production should be pursued, although collecting location- and technology-specific data is currently challenging.

Regarding the estimation of the geological resource production capacity in the watershed, we assigned the RCCs for all water users in proportion to the amount of the total human water consumption for each use. This was a simplified assumption; however, it may have affected the estimation of the overproduction of geological resources. A completely acceptable

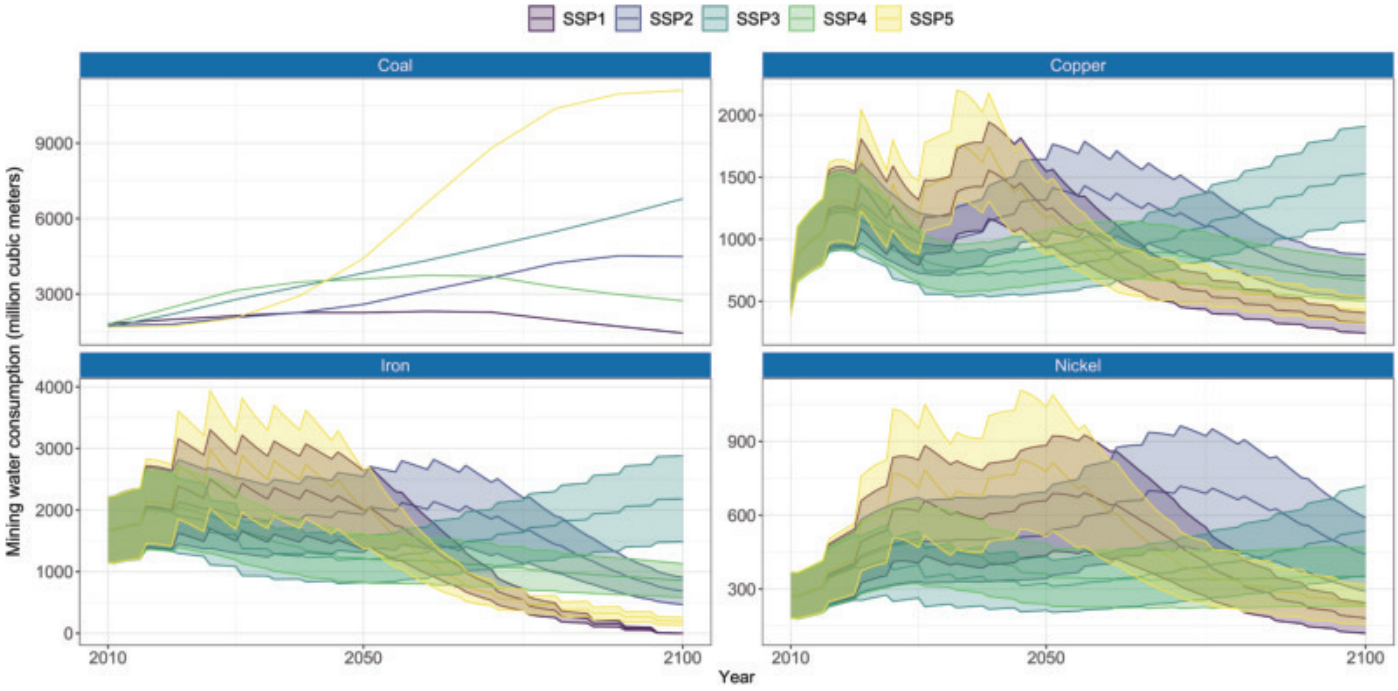


Fig. 4. Future water consumption associated with coal, copper, iron, and nickel production during 2010 to 2100 under different SSPs. Note that the shaded area in the plot denotes the range of future water consumption as the uncertainty of the estimation based on the standard deviation of water consumption intensity. Owing to the unavailability of data, the uncertainty of water consumption for coal production could not be estimated. SSP1, sustainability; SSP2, middle of the road; SSP3, regional rivalry; SSP4, inequality; SSP5, fossil-fueled development.

allocation method remains undevised; therefore, several viable options for allocating RCCs to each water user should be tested in the future. For future projections of water consumption for geological resource production, we estimated only the global amount of water consumed to produce major geological resources without assessing geological resource overproduction owing to the lack of reliable data on the geographical distribution of geological resource production and water availability and consumption.

REFERENCES AND NOTES

1. F. P. Carvalho, *Food Energy Secur.* **6**, 61–77 (2017).
2. F. Krausmann, D. Wiedenhofer, H. Haberl, *Glob. Environ. Change* **61**, 102034 (2020).
3. A. Elshkaki, L. Shen, *Energy* **180**, 903–917 (2019).
4. T. Watari, K. Nansai, K. Nakajima, *Resour. Conserv. Recycling* **164**, 105107 (2021).
5. K. Islam, R. Yokoi, M. Motoshita, S. Murakami, *Resour. Conserv. Recycling* **183**, 106384 (2022).
6. K. Islam, X. Vilaysouk, S. Murakami, *Resour. Conserv. Recycling* **154**, 104630 (2020).
7. V. Maus *et al.*, *Sci. Data* **9**, 433 (2022).
8. L. Tang, T. T. Werner, *Commun. Earth Environ.* **4**, 134 (2023).
9. V. Maus *et al.*, *Sci. Data* **7**, 289 (2020).
10. T. T. Werner *et al.*, *Glob. Environ. Change* **60**, 102007 (2020).
11. L. Cabernard, S. Pfister, *Environ. Sci. Technol.* **56**, 16357–16368 (2022).
12. S. Luckeneder, S. Giljum, A. Schaffartzik, V. Maus, M. Tost, *Glob. Environ. Change* **69**, 102303 (2021).
13. L. J. Sonter, M. C. Dade, J. E. M. Watson, R. K. Valenta, *Nat. Commun.* **11**, 4174 (2020).
14. R. Yokoi, T. Watari, M. Motoshita, *Energy Environ. Sci.* **15**, 146–157 (2022).
15. R. Koide, C. Tokoro, S. Murakami, T. Adachi, A. Takahashi, *Mine Water Environ.* **31**, 287–296 (2012).
16. K. Islam, S. Murakami, *Glob. Environ. Change* **70**, 102361 (2021).
17. K. Islam, S. Murakami, *Sustainability (Basel)* **12**, 9660 (2020).
18. W. Steffen *et al.*, *Science* **347**, 1259855 (2015).
19. J. Rockström *et al.*, *Nature* **461**, 472–475 (2009).
20. K. Richardson *et al.*, *Sci. Adv.* **9**, eadh2458 (2023).
21. S. A. Northey, G. M. Mudd, T. T. Werner, N. Haque, M. Yellishetty, *Water Resour. Ind.* **21**, 100104 (2019).
22. S. A. Northey, C. Madrid López, N. Haque, G. M. Mudd, M. Yellishetty, *J. Clean. Prod.* **184**, 788–797 (2018).
23. S. A. Northey, G. M. Mudd, E. Saarivuori, H. Wessman-Jääskeläinen, N. Haque, *J. Clean. Prod.* **135**, 1098–1116 (2016).
24. S. Meißner, *Resources* **10**, 120 (2021).
25. M. Motoshita, S. Pfister, M. Finkbeiner, *Environ. Sci. Technol.* **54**, 9083–9094 (2020).
26. SNL, S&P Global Mining Intelligence (2023); https://www.spglobal.com/marketintelligence/en/campaigns/metals-mining.
27. B. C. O'Neill *et al.*, *Clim. Change* **122**, 387–400 (2014).
28. B. C. O'Neill *et al.*, *Glob. Environ. Change* **42**, 169–180 (2017).
29. K. Riahi *et al.*, *Glob. Environ. Change* **42**, 153–168 (2017).
30. R. E. Dixon, *Desalination Water Treat.* **51**, 5–10 (2013).
31. A. Månberger, B. Stenqvist, *Energy Policy* **119**, 226–241 (2018).
32. B. K. Sovacool *et al.*, *Science* **367**, 30–33 (2020).
33. S. A. Northey *et al.*, *Glob. Environ. Change* **44**, 109–124 (2017).
34. J. Doolan, B. T. Hart, "Water Resource Policy, Planning and Management in Australia: An Overview" in *Decision Making in Water Resources Policy and Management* (Elsevier, 2017), pp. 3–19.
35. S. Davidsson, M. Höök, *Energy Policy* **108**, 574–582 (2017).
36. M. G. Macklin *et al.*, *Science* **381**, 1345–1350 (2023).
37. A. C. Schomberg, S. Bringezu, M. Flörke, *Commun. Earth Environ.* **2**, 11 (2021).
38. MCA, *Minerals Industry Water Accounting Framework* (Minerals Council of Australia, 2022).
39. K. Islam, K. Maeno, R. Yokoi, D. Giurco, S. Kagawa, S. Murakami, M. Motoshita, Constraints on global mined geological resource production from limited regional water availability, Zenodo (2024); http://doi.org/10.5281/zenodo.8088743.

ACKNOWLEDGMENTS

The authors thank S. Northey from the University of Technology, Sydney, for his helpful discussions and suggestions. **Funding:** Grant-in-Aid for Fostering Joint International Research (B) (JSPS KAKENHI JP18KK0303); Grant-in-Aid for Scientific Research (B) (JSPS KAKENHI JP23H03612); Grant-in-Aid for Scientific Research (A) (JSPS KAKENHI JP21H04944); Grant-in-Aid for Early-Career Scientists (JSPS KAKENHI JP24K20966). **Author contributions:** Conceptualization: K.I., R.Y., M.M.; Methodology: K.I., K.M., R.Y., M.M.; Investigation: K.I., R.Y., S.M., M.M.; Visualization: K.I., R.Y., M.M.; Funding acquisition: K.I., R.Y., S.M., M.M.; Project administration: K.I., R.Y., M.M.; Supervision: D.G., S.M., M.M.; Writing – original draft: K.I., K.M., R.Y., D.G., S.K., S.M., M.M.; Writing – review & editing: K.I., K.M., R.Y., D.G., S.K., S.M., M.M. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data and codes are available at Zenodo (*39*). **License information:** Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adk5318
Materials and Methods
Supplementary Text
Figs. S1 to S11
Tables S1 to S19
References (*40*–*79*)
MDAR Reproducibility Checklist

Submitted 28 August 2023; resubmitted 15 April 2024
Accepted 3 February 2025
10.1126/science.adk5318

ICE SHELVES

Deep learning the flow law of Antarctic ice shelves

Yongji Wang[1,2,3]*, Ching-Yao Lai[1,3]*, David J. Prior[4], Charlie Cowen-Breen[5]†

Antarctic ice shelves buttress the grounded ice sheet, mitigating global sea level rise. However, fundamental mechanical properties, such as the ice flow law and viscosity structure, remain under debate. In this work, by leveraging remote-sensing data and physics-informed deep learning, we provide evidence over several ice shelves that the flow law follows a grain size–sensitive composite rheology in the compression zone. In the extension zone, we found that ice exhibits anisotropic properties. We constructed ice shelf–wide anisotropic viscosity maps that capture the suture zones, which inhibit rift propagation. The inferred stress exponent near the grounding zone dictates the grounding-line ice flux and grounding line stability, whereas the inferred viscosity maps inform the prediction of rifts. Both are essential for predicting the future mass loss of the Antarctic Ice Sheet.

Sea level rise is one of the biggest environmental threats brought upon society by a warming climate. One of the major contributors to the projected rise in sea levels is the mass loss from the Antarctic Ice Sheet. Ice shelves, the floating extensions of ice sheets, buttress the grounded ice sheets and play a critical role in slowing their discharge (*1*–*3*). Substantial reduction of ice shelf buttressing is expected to increase ice discharge across the grounding line (*3*, *4*), thereby accelerating global sea level rise (*5*).

The motion of ice shelves is driven by gravity, modeled by the two-dimensional (2D) shallow-shelf approximation (SSA) equations (*6*) (upper right of Fig. 1). The flow law (rheology) of glacial ice governs its deformation in response to stress, thereby determining its discharge from grounded ice and the resulting mass loss of ice sheet into the ocean. However, our understanding of the flow law of glacier ice has substantial uncertainties, primarily due to the challenges associated with directly measuring the rheology of glacial ice on ice sheet–wide scales.

The flow law of ice, derived from laboratory experiments on polycrystalline synthetic ice (*7*, *8*), is commonly modeled by a power-law relation between the effective stress τ_e and strain rate $\dot{\epsilon}_e$, the second invariant of their tensors,

$$\dot{\epsilon}_e = A\tau_e^n \quad (1)$$

where n is the stress exponent. The prefactor A is a function of various physical properties, including temperature, grain size, fabric orientation, and chemical impurities (*9*). The first ice flow law was calibrated by Glen in 1955 (*7*). The power-law flow law with $n = 3$ is often referred to as Glen's law (*10*–*12*) and is the most common flow law used in ice sheet models. Validating the applicability of the laboratory-derived flow law is important because misrepresenting the flow law of glacial ice can lead to imprecise forecasts of ice mass loss into the oceans and its consequent impact on global sea level rise.

Antarctic ice shelf rheology

Ice shelves consist of a mixture of firn, meteoric glacial ice, and marine ice (*13*) with nonhomogeneous local structure and considerable anisotropy (*14*, *15*). Ice shelf rheology, therefore, may differ substantially from that of polycrystalline synthetic ice (*8*). In addition, large-scale features such as crevasses (*16*, *17*) and suture zones (*13*, *18*), regions of ice merged from different upstream sources, can introduce further complexity to the glacial rheology at ice shelf scales.

The rheology of ice shelves can be discerned through analysis of the relationship between stress and strain rate (Eq. 1) (*19*–*21*). The effective strain rate of ice shelves can be computed from observed ice velocities in two horizontal directions u, v (Fig. 2B). However, estimating effective stresses within an ice shelf is comparatively challenging. Direct measurements in boreholes (*22*) or indirect measurements from ice core samples (*23*) are rare. Stress calculations from ice velocity u, v and thickness h derived under assumptions about flow regimes (*19*–*21*) enable rheology assessments using observations, but their applicability is confined to narrow regions of ice shelves [supplementary materials (SM) section II.B; fig. S9].

Continent-wide calculations of the effective stress τ_e of ice shelves can be achieved by inferring the effective viscosity $\mu = \tau_e/2\dot{\epsilon}_e$, a fundamental mechanical property of glacial ice. Given that continent-wide data of ice velocity u, v and thickness h are available through remote-sensing observations (*24*, *25*), the effective viscosity μ is the sole unknown in the 2D SSA equations, which can be solved as an inverse problem (*26*). Prior studies, by using the control methods (*27*–*29*), inferred the ice hardness $B = A^{-1/n}$ in Eq. 1 assuming $n = 3$. By contrast, our approach infers the effective viscosity μ directly (*30*), avoiding power-law assumptions (Eq. 1) for the flow law and enabling a data-driven discovery of the underlying viscosity model.

We leverage physics-informed neural networks (PINNs) (*31*) to infer the effective viscosity μ. PINNs have been used for inferring ice hardness (*32*) and are useful for assimilating sporadic and noisy data. Figure 1 shows the workflow of PINNs. Detailed settings, methodological advantages, and robustness tests of PINNs are provided in the SM (SM section I; figs. S1 to S8). For example, Fig. 2A shows the viscosity inferred by PINNs on the Amery Ice Shelf. The effective strain rate (Fig. 2B) is computed according to the derivative of the neural network–approximated velocity fields u, v through automatic differentiation, and the effective stress τ_e (Fig. 2C) is then, by definition, obtained as the product of μ and $\dot{\epsilon}_e$. These two quantities, effective strain rate and stress, are used for determining the ice shelf rheology.

Extension versus compression zones

Figure 2, D to E, shows the relationship of stress to strain rate of ice in two square windows (50 km by 50 km) located at upstream and downstream regions of the Amery Ice Shelf (Fig. 2B). Neither shows a clear pattern with low coefficient of determination $r^2 < 0.5$. Note that the variation of surface temperature and the corresponding ice hardness B within each window is too small (*33*) to cause the spread of the data. This indicates that the rheology of the ice shelf is more complicated than what a simple power law can describe.

To investigate how ice rheology evolves through an ice shelf, we examined the relationship of stress to strain rate along a single flow line (black curves in Fig. 2, F to H), connecting the upstream region near the grounding line to the downstream region near the calving front. Figure 2I shows that this relationship, along a 60-km-long section of the flow line in the upstream region, closely follows a power law with n around 1. However, the relationship gradually transitions to a vertical line in the downstream region, where it resembles the perfect plastic rheology (*34*). To elucidate this transition, we computed the strain rate tensor $\dot{\epsilon}$ in the flow-line coordinates $(\hat{x}, \hat{y})$, with $\hat{x}$ and $\hat{y}$ being tangential and normal to local flow directions, respectively (SM section II.A). Figure 2, G and H, displays the shear strain rate $\dot{\epsilon}_{\hat{x}\hat{y}}$ and transverse strain rate $\dot{\epsilon}_{\hat{y}\hat{y}}$, which are negligible compared with the longitudinal strain rate $\dot{\epsilon}_{\hat{x}\hat{x}}$ (Fig. 2F) along the selected flow line. Based on $\dot{\epsilon}_{\hat{x}\hat{x}}$, the flow line can be segmented into (i) the compression zone with $\dot{\epsilon}_{\hat{x}\hat{x}} < 0$, indicating that a flow parcel at the front moves

[1]Department of Geophysics, Stanford University, Stanford, CA, USA. [2]Department of Mathematics, New York University, New York, NY, USA. [3]Department of Geosciences, Princeton University, Princeton, NJ, USA. [4]Department of Geology, University of Otago, Dunedin, New Zealand. [5]Department of Mathematics, Princeton University, Princeton, NJ, USA.
*Corresponding author. Email: yongjiw@stanford.edu (Y.W.); cyaolai@stanford.edu (C.-Y.L.)
†Present address: Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA.

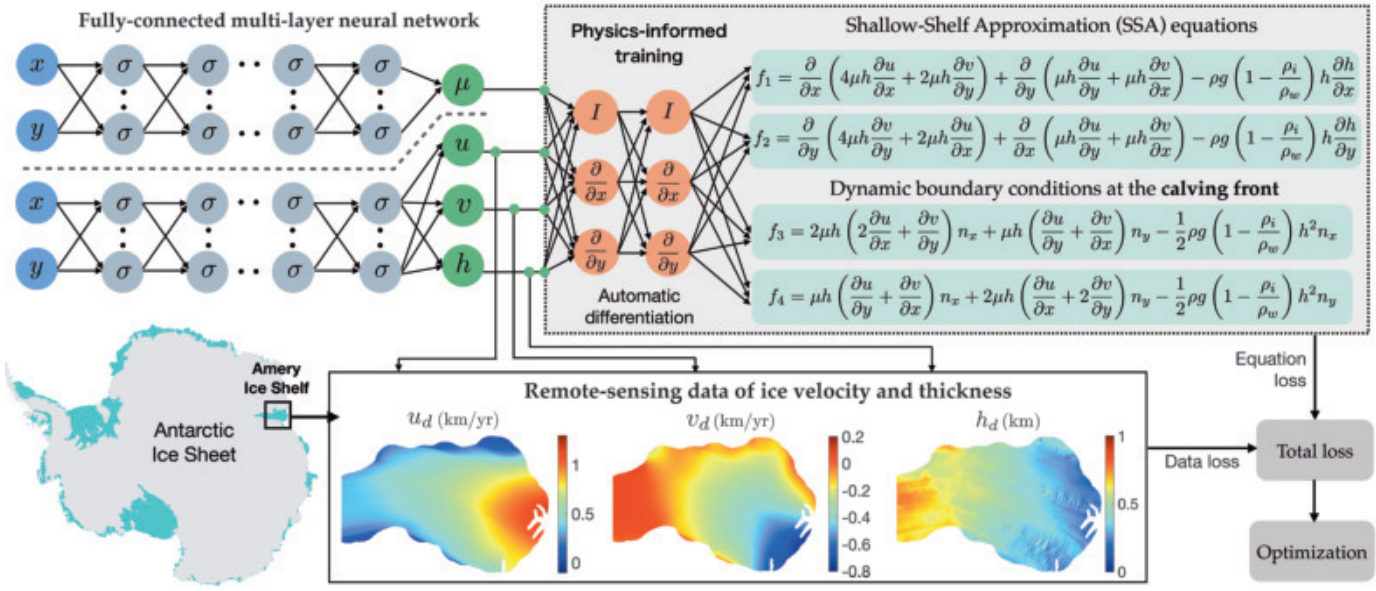


Fig. 1. PINN setup. The structure and workflow of PINNs for inferring effective viscosity μ of ice shelves from remote-sensing data [ice shelf velocity *u*, *v* (*23*) and thickness *h* (*24*)], with the SSA equations (f_1, f_2 = 0) (*6*), and the dynamic boundary conditions (f_3, f_4 = 0) at the calving front of ice shelves (*61*). The cost function *L* contains two terms: the data loss L_d and the equation loss L_e.

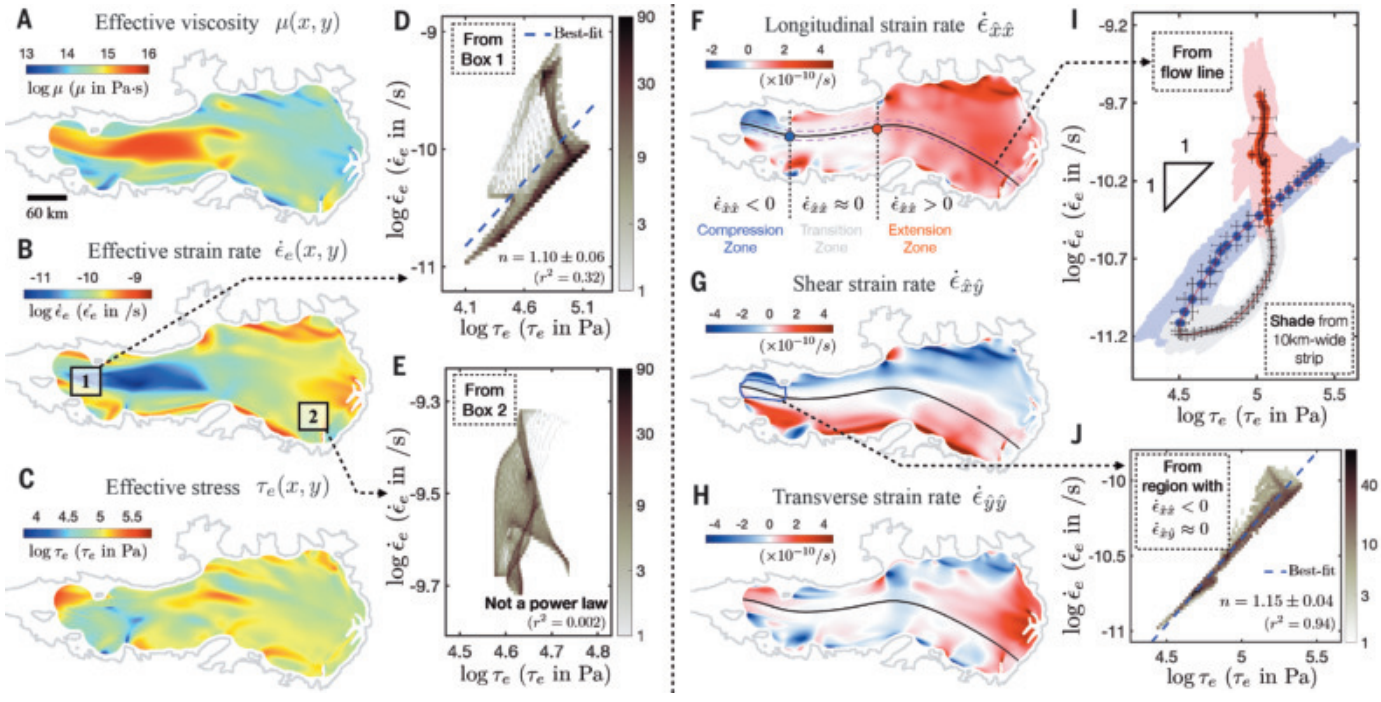


Fig. 2. The complex flow law of the Amery Ice Shelf. (**A**) Effective viscosity μ, (**B**) effective strain rate $\dot{\epsilon}_e$, and (**C**) effective stress τ_e of the Amery Ice Shelf inferred from PINNs. (**D** and **E**) The stress–strain rate relations in two square windows in (B), with best-fit stress exponent *n* and associated r^2 value. The color scale represents the number of data points. (**F** to **H**) Longitudinal $\dot{\epsilon}_{\hat{x}\hat{x}}$, shear $\dot{\epsilon}_{\hat{x}\hat{y}}$, and transverse $\dot{\epsilon}_{\hat{y}\hat{y}}$ strain rates in the flow-line coordinates, respectively. The black curve indicates a single flow line, with two nearby flow lines (purple dashed lines) that are 10 km away. (**I**) Stress–strain rate relation along the black flow line in (F). The data points in the compression, transition, and extension zones are labeled in blue, gray, and red colors, respectively. Error bars indicate the standard deviation of 15 repetitions of PINN inversions with different random initialization. The shaded regions show the stress and strain-rate data between the pink dashed lines in (F). (**J**) The stress–strain rate relation in the entire shear-free zone ($\dot{\epsilon}_{\hat{x}\hat{y}} \approx 0$) and compression zone ($\dot{\epsilon}_{\hat{x}\hat{x}} < 0$) obeys a power law with *n* around 1. The color scale indicates the number of data points.

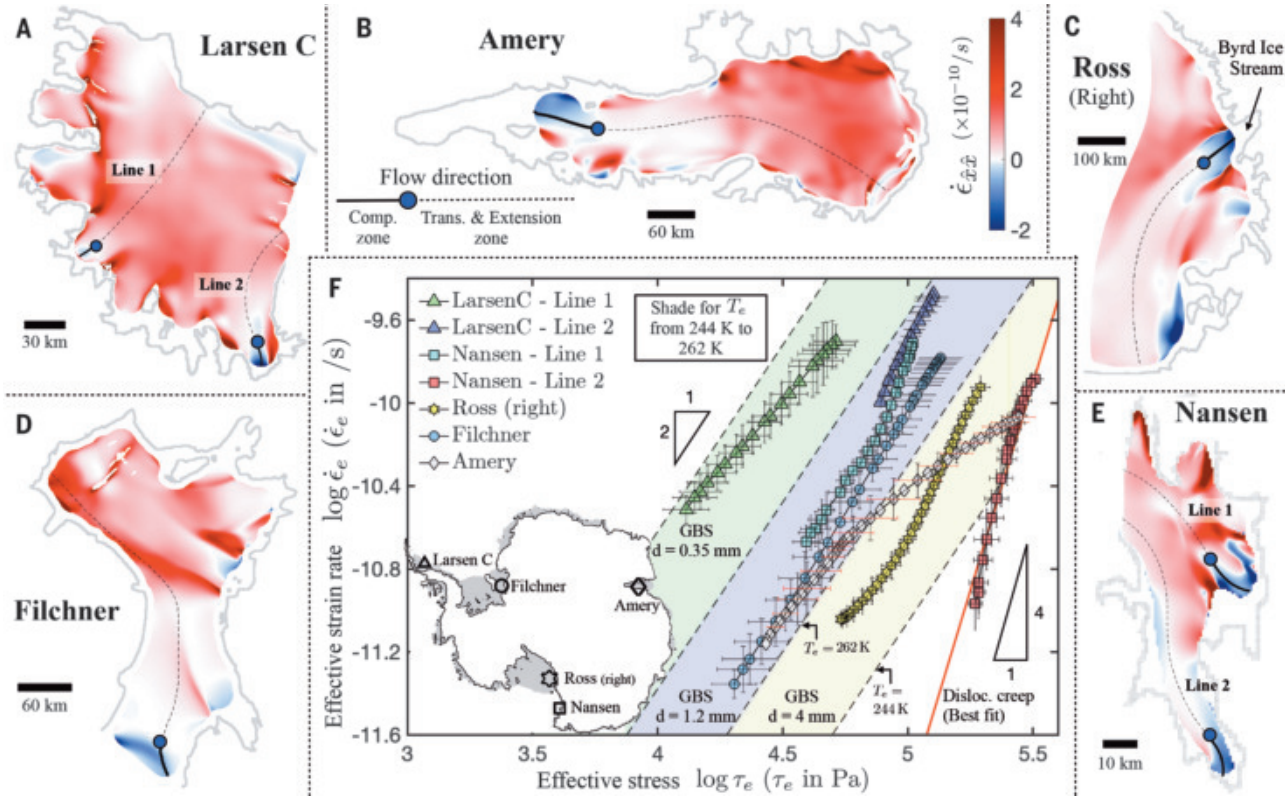


Fig. 3. Composite rheology in the compression zone. (**A** to **E**) Longitudinal strain rate $\dot{\epsilon}_{\hat{x}\hat{x}}$ in the flow-line coordinates of the five Antarctic ice shelves. The black curves indicate the flow lines that pass through the compression zones. The thick solid line represents the section in the compression zone, whereas the thin dashed line lies in the transition and extension zones. (**F**) The stress–strain rate relations in the compression zones on the five Antarctic ice shelves. Error bars indicate the standard deviation of 10 repetitions of PINN inversions with different random initialization. The different color shades represent the prediction of the stress–strain rate relations using GBS with different grain sizes (dashed lines) (Eq. 2). The red solid line represents the DC term in Eq. 2 with A_d fitted to the data. The width of the shaded regions corresponds to the range of temperatures T_e (244 to 262 K) used to plot Eq. 2. The inset shows the location of flow lines on each ice shelf.

slower than one at the rear and thus implying compression of flow; (ii) the extension zone with $\dot{\epsilon}_{\hat{x}\hat{x}} > 0$; and (iii) the transition zone, a region with vanishing $\dot{\epsilon}_{\hat{x}\hat{x}} \approx 0$ between the compression and extension zones. Its quantitative definition is provided in SM section III.A.

We found that the boundary between the compression zone and transition zone corresponds to the inflection point at which the relationship of stress to strain rate deviates from the power law (Fig. 2I). The predominance of power-law rheology in the compression zone was not assumed a priori but revealed by the data themselves. Expanding the analysis from a single flow line to a 10-km-wide band (indicated by the purple dashed lines in Fig. 2F), we observed consistent power-law rheology, shown by the blue-shaded region in Fig. 2I. Moreover, examining the relationship of stress to strain rate over the entire zone dominated by compression (about 50 km by 25 km, enclosed by the blue curve in Fig. 2G) reaffirms this power-law behavior (Fig. 2J). By contrast, in the extension zone, the stress and strain-rate data within the 10-km band exhibits a wide spread with no clear power-law relationship (red-shaded region in Fig. 2I).

Composite rheology in the compression zone

Clear evidence of power-law rheology in the compression zone is consistently observed across different ice shelves, including the Larsen C, Ronne-Filchner, Ross, Amery and Nansen Ice Shelves (SM section III.B; figs. S11 and S12), with the best-fit stress exponent n generally increasing with the stresses, as shown in Fig. 3F (also see the table in fig. S12F). These findings resemble the composite rheology of ice found by laboratory experiments (*8*, *35*) (SM section III.C). A composite rheology is the mixture of two (or more) distinct physical deformation mechanisms. For ice (*8*, *36*), the composite rheology comprises two mechanisms: the sliding between ice grains, referred to as grain-boundary sliding (GBS), and the glide-and-climb of dislocations within the grain, known as dislocation creep (DC). The composite rheology model can be expressed in terms of the depth-averaged effective strain rate $\dot{\epsilon}_e$ and stress τ_e (SM section III.C)

$$\dot{\epsilon}_e = \underbrace{A_g d^{-p_g} \tau_e^{n_g} \exp\left(-\frac{Q_g}{RT_e}\right)}_{\text{GBS}} + \underbrace{A_d \tau_e^{n_d} \exp\left(-\frac{Q_d}{RT_e}\right)}_{\text{DC}} \quad (2)$$

where stress exponents n_g and n_d, activation energies Q_g and Q_d, and scaling factors A_g and A_d are empirical constants (*8*, *37*). T_e is the depth-averaged temperature of ice shelves, R is the universal gas constant, and d is the grain size with the grain size exponent p_g = 1.4.

Despite the uncertainty in depth-averaged temperature T_e (the range of realistic T_e is shown by the shaded regions in Fig. 3F), for most compression zones with maximum stress less than 100 kPa, the relationship of stress to strain rate has exponent n around 2 (SM section III.D; fig. S12), close to the empirical stress exponent n_g = 1.8 of the GBS term in Eq. 2. This suggests that the GBS mechanism dominates ice deformation there.

By contrast, the compression zone found on the Nansen Ice Shelf (along line 2) is under high stress (τ_e > 200 kPa). The exponent n is around 4, close to the empirical stress exponent n_d = 4 of the DC term in Eq. 2, which suggests predominance of the DC mechanism.

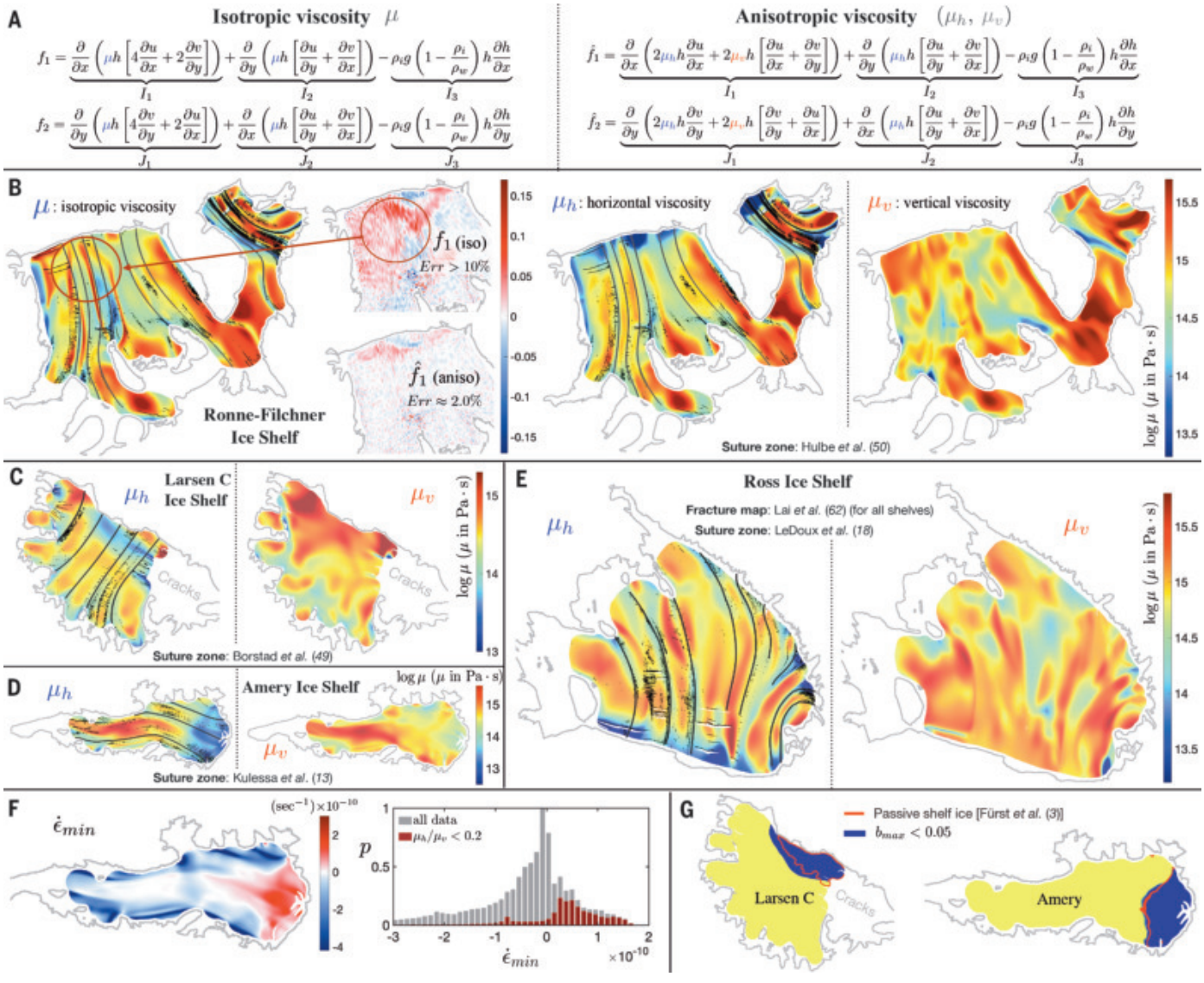


Fig. 4. Anisotropic viscosity of ice shelves. (**A**) Expressions of isotropic and anisotropic SSA equations with horizontal and vertical viscosities (μ_h, μ_v). (**B**) Comparison of the inferred isotropic effective viscosity μ of the Ronne-Filchner Ice Shelf to our inferred anisotropic viscosities (μ_h, μ_v). The equation residue $\hat{f}_1$ for anisotropic viscosity is much smaller than that for isotropic viscosity f_1. The black lines indicate the previously mapped suture zones of the ice shelves, which align well with the locations with small μ_h. Such alignments were not observed for the isotropic viscosity. (**C** to **E**) Horizontal μ_h and vertical viscosity μ_v of three other Antarctic ice shelves, the (C) Ross, (D) Larsen C, and (E) Amery Ice Shelves. The thick black lines indicate the suture zones, and the thin black lines indicate fracture locations (*62*). (**F**) The minimal principal strain rate $\dot{\epsilon}_{min}$ of the Amery Ice Shelf. The region $\dot{\epsilon}_{min} < 0$ is known as the compressive arch. The histogram shows that regions with strong anisotropy ($\mu_h \ll \mu_v$) mainly lie outside the compressive arch. (**G**) The overlap between the areas with a low maximum buttressing number b_{max} and the passive shelf ice (*3*).

Using the empirical values of n_d and Q_d from experiments (*8*), we inferred the best-fit value of $A_d = 0.9 \times 10^5$ MPa^{-4} s^{-1}, which is an order of magnitude lower than the value from experiments (*12*). The experimental value was obtained from ice with fabric orientations that facilitate deformation (*12*). By contrast, the fabric formed in grounded ice may be oriented unfavorably for deformation kinematics in the ice shelf, making DC harder.

The compression zone around the Byrd Glacial Ice Stream on the Ross Ice Shelf experiences stress between 100 and 200 kPa, and the corresponding stress exponent n is approximately 3, which suggests that both GBS and DC mechanisms contribute to deformation in this region. This is consistent with the fact that the stress and strain-rate data lie at the interface between the yellow-shaded regions (GBS dominated) and the red line (DC dominated) in Fig. 3F.

Although the relationships of stress to strain rate in various low-stress compression zones are well described by the GBS mechanism with a consistent stress exponent of n around 2, the offset among those flow laws are notable. For example, at a stress of around 4.5 kPa, the compression zone of Larsen C (line 1) deforms at a strain rate that is an order of magnitude faster than that of Filchner. This difference cannot be solely attributed to the variation in depth-averaged temperature T_e. The strain rate difference predicted by the GBS term in Eq. 2 within a realistic range from 244 to 262 K (shown by the shaded regions in Fig. 3F) is about half an order of magnitude or less, much smaller than the spread of strain rates at a fixed stress

in Fig. 3F. Thus, different grain sizes d in these compression zones are needed to explain the offset among those flow laws, with the inferred grain size using Eq. 2 shown in Fig. 3F.

We note that the grain sizes inferred from the GBS mechanism (Eq. 2) appear to be smaller than the steady-state grain size predicted for these low-stress compression zones (τ_e < 100 kPa in Fig. 3F and figs. S12F and S13) (*38*, *39*). We hypothesized that this discrepancy results from the rapid decrease in stress as glacier ice moves across the grounding line through the high-stress flexure zones (*40*) to the low-stress zones of ice shelves. The small ice grain size, related to higher stresses when grounded, may be preserved because the timescale required for ice grains to grow to the steady-state size corresponding to lower-stress conditions (*41*) is substantially longer (~1000 years) than the time it takes for ice to travel to the low-stress (τ_e < 100 kPa) compression zones on the ice shelf (~10 years). This leads to the predominance of the GBS mechanism (*42*) in governing the rheological behavior of low-stress compression zones on the ice shelves.

The evidence presented thus far reinforces the conclusion that the ice shelf rheology in the compression zones follows the composite rheology (Eq. 2). However, exceptions exist, such as the compression zone on the Amery Ice Shelf, where the stress exponent $n = 1.15 \pm 0.04$ (Fig. 2J) is observed, which deviates markedly from the composite rheology (Fig. 3F). Linear viscous behavior ($n = 1$) has been observed in creep experiments at the melt temperature (*43*, *44*), and it is plausible that Amery, which has substantial surface melting (*45*), may exhibit similar behavior. Although most of our data are consistent with a composite rheology in the compression zones, other physical processes could additionally influence local rheological behavior, necessitating further careful examination.

Anisotropic viscosity in the extension zone

The stress and strain-rate data in the extension zones do not exhibit clear power-law relationships. This is seen in Amery Ice Shelf (Fig. 2I) and all others considered (SM section III.B; figs. S10 and S11). We observed that the residue of the governing equations f_i is consistently larger in the extension zones than anywhere else (Fig. 4B and fig. S14). Systematic tests show that this larger error arises from a contradiction between the data and the ice shelf equations used in the inversion (SM section IV.A; fig. S16).

Given the ice velocity u, v and thickness h from observations, the effective viscosity μ, as the only unknown, is overconstrained by the two SSA equations (f_1, f_2 in Fig. 1) (SM section IV.B). We elucidate in SM section IV.C that the overconstrained issue is due to the isotropic viscosity assumption underlying the classical SSA equations and vanishes if one allows for anisotropic viscosity. Potential factors contributing to the anisotropic property of ice shelves include the existence of aligned crevasses (*46*) and the orientation of ice fabrics (*14*, *15*).

We derived a set of new anisotropic ice shelf equations (Fig. 4A; SM section V.E) with local transverse isotropic anisotropy (*47*, *48*), characterized by distinct horizontal and vertical viscosity (μ_h, μ_v) (SM section IV.C), that further minimizes data misfit and equation residues (figs. S18 and S19). Figure 4B shows that the residue of our proposed anisotropic SSA equations in the extension zones is 10 times smaller than that of the isotropic equations. This substantial improvement in inversion accuracy is consistent across other ice shelves (SM sections IV.E and IV.F; fig. S18).

Figure 4, C to E, shows the inferred horizontal and vertical viscosities (μ_h, μ_v) for various ice shelves. Two lines of physical evidence further validate the inferred anisotropic viscosity. First, the horizontal viscosity pattern accurately captures the suture zones (*13*, *18*, *49*, *50*) (solid black lines in Fig. 4, B to E), which are bands of the ice shelf extending downstream from an obstruction (e.g., an ice rise) and forming marine ice where the low viscosity is expected. Such clear alignment with suture zones is not present in the isotropic viscosity fields in existing ice sheet models, likely because of the large inference error caused by the overconstrained inversion (Fig. 4B). Accurately representing the low-viscosity pattern in the suture zone is critical for predicting rift propagation, given that rifts often terminate at the suture zone (*13*, *51*).

Second, the low–horizontal viscosity regions near the calving fronts of different ice shelves correspond closely to regions outside the compressive arch (*52*), where both of the principal strain rates in the horizontal direction are positive, promoting crevasse and fracture growth (Fig. 4F). Damage mechanics implies that the presence of vertical crevasses should cause a decrease in horizontal viscosity (*53*), which provides a rationale for the low horizontal viscosity inferred outside the compressive arch. Fabric patterns across an ice shelf are likely to be complex, comprising areas with subhorizontal *c*-axis maxima (*15*) related to horizontal shear on vertical planes, vertical girdles related to horizontal extension (*14*), and areas with subvertical maxima advected from outlet glaciers and ice streams (*14*, *54*) and formed on the ice shelf. These mechanisms could explain the local transverse isotropic anisotropy, but there are insufficient data to extend this argument further. Empirically, we also observed that the low–horizontal viscosity area exhibits a low maximum buttressing number b_{max} (*2*) (SM section IV.G; fig. S20), which aligns well with the passive shelf ice zone outlined in Fürst *et al.* (*3*) (Fig. 4G). More in situ measurements are required to further refine the anisotropic inversion. However, this study could serve as a useful framework for future research on anisotropic ice shelf rheology.

Implications for future studies

We assessed the flow law and viscosity structure across different ice shelves by leveraging continent-scale remote-sensing data. Our analysis shows that ice shelf rheology differs substantially between the compression and extension zones. We found clear evidence that the rheology in several compression zones is consistent with a composite rheology. The ice shelf compression zones typically occur where grounded ice enters ice shelves through the grounding line. Theoretical studies (*1*, *55*) have shown that the grounding line flux and its position are highly dependent on the stress exponent n. Recent work has also shown that the stability of the grounding line is sensitive to the value of stress exponent n (*56*, *57*), making the precise quantification of n near the grounding line critical. Therefore, the impact of our data-driven flow law on the prediction of grounding line flux, stability, the future mass loss of ice, and the consequential global sea level rise is an important subject of future work. At a fundamental level, there are gaps in our understanding of ice rheology when under extension: Ice rheology experiments are predominantly conducted under compression (*7*, *8*). Future experiments quantifying ice rheology in response to extension stresses, such as fractures that could alter ice deformation, would be essential for understanding the ice shelf rheology in the extension zone.

The data observed in the extension zone indicate that we should consider more complex viscosity models than the simple isotropic power-law rheology. In this study, we introduce anisotropic ice viscosity inversion. Our inferred anisotropic viscosity fields accurately capture previously mapped suture zones, which exhibit low viscosity that can arrest rift propagation on the ice shelves. Therefore, anisotropic viscosity inversion is likely to be a powerful tool for predicting rheological heterogeneities that will underpin processes such as shear zone localization, rift propagation, and ice shelf calving. Moreover, we demonstrate that the regions outside the compressive arch, where crevasses can form more easily as a result of tensile stress, correspond closely to regions that exhibit anisotropy. Future work addressing the coupling of ice damages (including crevasses and rifts) (*46*, *58*–*60*) with ice shelf rheology is crucial for elucidating the anisotropic viscosity maps inferred in this study and assessing their impact on the future evolution of the Antarctic Ice Sheet. Lastly, we demonstrate how extensive observational datasets combined with deep-learning techniques can reveal the behavior of Earth and planetary materials in natural settings. The machine learning–based method

developed here has the potential to assist scientific discoveries for a wide spectrum of data-driven research and may hold considerable interest to the broader earth science community.

REFERENCES AND NOTES

1. C. Schoof, *J. Fluid Mech.* **573**, 27–55 (2007).
2. G. Gudmundsson, *Cryosphere* **7**, 647–655 (2013).
3. J. J. Fürst *et al*., *Nat. Clim. Chang.* **6**, 479–482 (2016).
4. R. Reese, G. H. Gudmundsson, A. Levermann, R. Winkelmann, *Nat. Clim. Chang.* **8**, 53–57 (2018).
5. S. Sun *et al*., *J. Glaciol.* **66**, 891–904 (2020).
6. D. R. MacAyeal, *J. Geophys. Res.* **94**, 4071–4087 (1989).
7. J. W. Glen, *Proc. R. Soc. London Ser. A* **228**, 519–538 (1955).
8. D. Goldsby, D. L. Kohlstedt, *J. Geophys. Res.* **106**, 11017–11030 (2001).
9. K. M. Cuffey, W. S. B. Paterson, *The Physics of Glaciers* (Academic Press, 2010).
10. R. LeB. Hooke, *Rev. Geophys.* **19**, 664–672 (1981).
11. W. F. Budd, T. Jacka, *Cold Reg. Sci. Technol.* **16**, 107–144 (1989).
12. S. Fan *et al*., Flow laws for ice constrained by seventy years of laboratory experiments. ResearchSquare 10.21203/rs.3.rs-4809184/v1 [Preprint] (2025); https://doi.org/10.21203/rs.3.rs-4809184/v1.
13. B. Kulessa *et al*., *Nat. Commun.* **10**, 5491 (2019).
14. A. Treverrow, R. C. Warner, W. F. Budd, M. Craven, *J. Glaciol.* **56**, 877–890 (2010).
15. F. Lutz *et al*., *J. Geophys. Res. Earth Surf.* **125**, e2020JF005707 (2020).
16. C. van der Veen, *Cold Reg. Sci. Technol.* **27**, 31–47 (1998).
17. W. R. Buck, C.-Y. Lai, *Geophys. Res. Lett.* **48**, e2021GL093110 (2021).
18. C. M. LeDoux, C. L. Hulbe, M. P. Forbes, T. A. Scambos, K. Alley, *Ann. Glaciol.* **58**, 88–98 (2017).
19. R. H. Thomas, *J. Glaciol.* **12**, 55–70 (1973).
20. K. C. Jezek, R. B. Alley, R. H. Thomas, *Science* **227**, 1335–1337 (1985).
21. J. D. Millstein, B. M. Minchew, S. S. Pegler, *Commun. Earth Environ.* **3**, 57 (2022).
22. W. T. Pfeffer, N. Humphrey, B. Amadei, J. Harper, J. Wegmann, *Ann. Glaciol.* **31**, 229–235 (2000).
23. M. Jackson, B. Kamb, *J. Glaciol.* **43**, 415–426 (1997).
24. J. Mouginot, E. Rignot, B. Scheuchl, MEaSUREs Phase-based Antarctica Ice Velocity Map, Version 1, National Snow and Ice Data Center (2019); https://doi.org/10.5067/PZ3NJ5RXRH10.
25. M. Morlighem, MEaSUREs Bedmachine Antarctica, Version 2, National Snow and Ice Data Center (2020); https://doi.org/10.5067/E1QL9HFQ7A8M.
26. A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation* (SIAM, 2005).
27. D. R. MacAyeal, *J. Glaciol.* **39**, 91–98 (1993).
28. J. J. Fürst *et al*., *Cryosphere* **9**, 1427–1443 (2015).
29. S. Wang *et al*., *J. Geophys. Res. Earth Surf.* **127**, e2021JF006346 (2022).
30. A. Vieli, A. J. Payne, Z. Du, A. Shepherd, *Philos. Trans. A Math. Phys. Eng. Sci.* **364**, 1815–1839 (2006).
31. M. Raissi, P. Perdikaris, G. Karniadakis, *J. Comput. Phys.* **378**, 686–707 (2019).
32. B. Riel, B. Minchew, *J. Glaciol.* **69**, 1167–1186 (2023).
33. Y. Wang, C. Zhao, R. Gladstone, B. Galton-Fenzi, R. Warner, *Cryosphere* **16**, 1221–1245 (2022).
34. J. Weertman, *J. Glaciol.* **13**, 3–11 (1974).
35. W. B. Durham, O. Prieto-Ballesteros, D. Goldsby, J. Kargel, *Space Sci. Rev.* **153**, 273–298 (2010).
36. D. L. Goldsby, in *Glacier Science and Environmental Change*, P. G. Knight, Ed. (Wiley, 2006), pp. 308–314.
37. E.-J. Kuiper *et al*., *Cryosphere* **14**, 2429–2448 (2020).
38. T. Jacka, L. Jun, *Ann. Glaciol.* **20**, 13–18 (1994).
39. J. Platt, *J. Struct. Geol.* **177**, 104987 (2023).
40. H. Still *et al*., *Front. Earth Sci. (Lausanne)* **10**, 828313 (2022).
41. S. Fan *et al*., *Cryosphere* **17**, 3443–3459 (2023).
42. R. H. Wightman, D. J. Prior, T. A. Little, *J. Struct. Geol.* **28**, 902–918 (2006).
43. C. J. Adams, N. R. Iverson, C. Helanow, L. K. Zoet, C. E. Bate, *Front. Earth Sci. (Lausanne)* **9**, 702761 (2021).
44. C. M. Schohn, N. R. Iverson, L. K. Zoet, J. R. Fowler, N. Morgan-Witts, *Science* **387**, 182–185 (2025).
45. J. F. Arthur *et al*., *Nat. Commun.* **13**, 1711 (2022).
46. A. Pralog, M. Funk, *J. Geophys. Res.* **110**, B01309 (2005).
47. C. Martin, G. H. Gudmundsson, H. D. Pritchard, O. Gagliardini, *J. Geophys. Res.* **114**, F04001 (2009).
48. W. Sharples, L. N. Moresi, M. Velic, M. A. Jadamec, D. A. May, *Phys. Earth Planet. Inter.* **252**, 77–90 (2016).
49. C. Borstad, D. McGrath, A. Pope, *Geophys. Res. Lett.* **44**, 4186–4194 (2017).
50. C. L. Hulbe, C. LeDoux, K. Cruikshank, *J. Glaciol.* **56**, 459–472 (2010).
51. D. McGrath *et al*., *J. Geophys. Res. Earth Surf.* **119**, 588–602 (2014).
52. C. Doake, H. Corr, H. Rott, P. Skvarca, N. Young, *Nature* **391**, 778–780 (1998).
53. A. Huth, R. Duddu, B. Smith, *J. Adv. Model. Earth Syst.* **13**, MS002292 (2021).
54. S. Picotti, A. Vuan, J. M. Carcione, H. J. Horgan, S. Anandakrishnan, *J. Geophys. Res. Solid Earth* **120**, 4237–4262 (2015).
55. S. S. Pegler, *J. Fluid Mech.* **857**, 605–647 (2018).
56. M. Haseloff, O. V. Sergienko, *J. Glaciol.* **64**, 417–431 (2018).
57. M. Ranganathan, B. Minchew, *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2309788121 (2024).
58. J. Krug, J. Weiss, O. Gagliardini, G. Durand, *Cryosphere* **8**, 2101–2117 (2014).
59. S. Jiménez, R. Duddu, J. Bassis, *Comput. Methods Appl. Mech. Eng.* **313**, 406–432 (2017).
60. R. Duddu, S. Jime'nez, J. Bassis, *J. Glaciol.* **66**, 415–429 (2020).
61. D. N. Goldberg, P. Heimbach, I. Joughin, B. Smith, *Cryosphere* **9**, 2429–2446 (2015).
62. C.-Y. Lai *et al*., *Nature* **584**, 574–578 (2020).

ACKNOWLEDGMENTS

We thank J. Millstein and B. Minchew for helpful discussions regarding rheology in different flow regimes and O. Sergienko for valuable conversations about the control method. We thank the Stanford Doerr School for providing computational resources and partial support through the Discovery Grant. We acknowledge the Office of the Dean for Research at Princeton University for partial funding support through the Dean for Research Fund for New Ideas in the Natural Sciences. C.-Y.L. acknowledges the National Science Foundation for funding through grant no. DMS-2245228 and NASA for partial support through Cryosphere Award 80NSSC21K1003. We also acknowledge financial support from the Schmidt Data X Fund at Princeton University, which is made possible through a major gift from the Schmidt Futures Foundation. D.J.P. acknowledges support from Marsden Fund UOO1716 from the Royal Society of New Zealand. **Author contributions:** Y.W. and C.-Y.L. conceived the project. Y.W. conducted the experiments and analyzed the results. Y.W. designed the training scheme of PINN using the SGD technique proposed by C.C.-B. C.-Y.L. and D.J.P. helped interpret the results. Y.W. prepared the figures and wrote the first draft of the manuscript. Y.W., C.-Y.L., D.J.P., and C.C.-B. all contributed to the editing of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Preprocessed data of different ice shelves used for the PINN training are available at https://github.com/YaoGroup/DIFFICE_jax/tree/main/examples/real_data. The raw velocity data across Antarctica are from MEaSUREs Phase-Based Antarctica Ice Velocity Map, Version 1 (NSIDC-0754) (*24*), available at https://nsidc.org/data/nsidc-0754/versions/1. Thickness data across Antarctica are from MEaSUREs BedMachine Antarctica, Version 2 (NSIDC-0756) (*25*), available at https://nsidc.org/data/nsidc-0756/versions/2. To accelerate the applications of PINNs in glaciology and various problems with real observational data, we made our PINN code written in JAX open source available on Github at https://github.com/YaoGroup/DIFFICE_jax/; documentations and tutorial examples are provided at https://diffice-jax.readthedocs.io. The PINN methodology used in this study implemented several advanced features (such as extended network architecture, data and equation normalization, collocation points resampling, extended-PINNs for large ice shelves, and anisotropic viscosity inversion) (refer to SM section I) that are beyond the original PINN setup (*31*). **License information:**

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adp3300
Material and Methods
Supplementary Text
Figs. S1 to S20
Table S1
References (*63–85*)

Submitted 19 March 2024; accepted 16 January 2025
10.1126/science.adp3300

By **Muhammed Rashid**

After the flood

Four years into my undergraduate degree in the southern Indian state of Kerala, I awoke to a friend urgently shaking me, telling me the heavy rain the previous night was flooding our area. We rushed outside to a chaotic and distressing scene. Families were stranded on rooftops. Children were crying for help. Hospital patients were being evacuated. My friends and I immediately joined the rescue teams, volunteering to distribute food to the affected people. I had intended to devote days or weeks to the effort—however long it took to help my community. But the event had a more lasting effect: propelling me toward a new career path.

I had gone to college to study civil engineering. Growing up in an impoverished rural area, I was fascinated by architecture in cities, and I hoped to design buildings myself someday. However, as my degree progressed, I became enthralled with research and wanted to continue my academic journey. I wasn't sure what to study. Then the flood came, and it all became clear.

After spending much of my summer volunteering in affected communities, I decided to focus my career on natural disasters, exploring the various factors behind them, as well as their impacts on society. As a first step, I applied for a master's program in development studies—a multidisciplinary program that included coursework in sustainable development, geography, and economics, as well as hands-on community projects. This decision shocked my friends and family. Relatives kept asking my parents why I was leaving engineering to study social sciences. They warned my parents, who only have a primary education, that this path might make it hard for me to find a job. However, I was determined and went ahead with what I thought was right for me.

The program, which exposed me to projects that helped communities plan for climate change and other disruptions, convinced me I was on the right path. I decided to apply for a Ph.D. It was a bold move given my uncertain finances. I also found that many universities preferred candidates who had consistent academic backgrounds and hadn't moved around between engineering and the social sciences. During interviews, faculty members often questioned why I chose to risk my career by changing fields. I received multiple rejections, both in India and abroad. But one program agreed to accept me, and I set out to study how better water resource management and policy can help vulnerable populations adapt to the uncertainties posed by climate change.

"The opportunity to delve into an issue I care so passionately about ... makes it all worthwhile."

I was excited to start the program and be surrounded by some of the brightest minds in the country. It wasn't easy, though. Most of my peers had followed a traditional path, without jumping around between fields, and had family members with degrees in higher education. I sometimes felt unsure about my depth of knowledge. I also had trouble connecting with my peers at times, leaving me feeling lonely.

My feelings of isolation began to change with the arrival of a new Ph.D. student who also had a multidisciplinary background and a similar family situation. Meeting him was a huge relief. We connected easily because we had both dealt with rejections from Ph.D. programs and faced mockery for our career decisions. Opening up with one another didn't make our problems go away. But from that point on, I felt less alone.

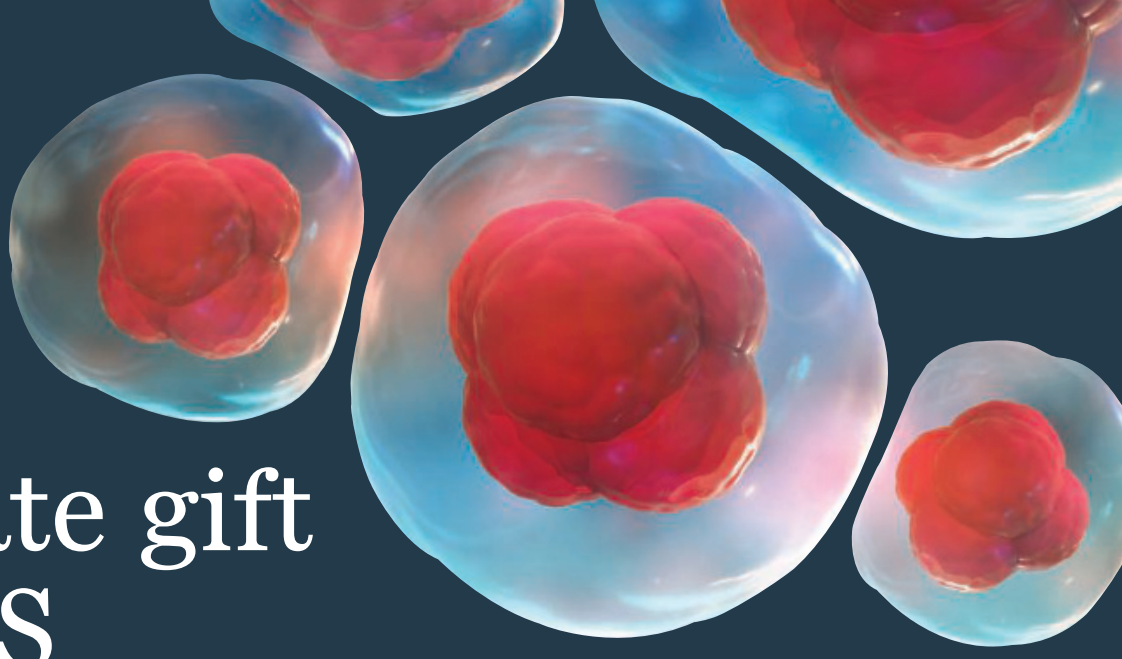
Sharing what I've learned with people in my community has also given me satisfaction. A barber from my hometown whom I have known since childhood now regularly talks with me about the increasing frequency of extreme weather events. Despite only having a primary education, he's filled with curiosity, excitement, and deep respect for the research I discuss with him.

Now, 3 years into my program, I am confident in my decision to pursue a Ph.D. There have been challenges along the way, but the opportunity to delve into an issue I care so passionately about and try to help people makes it all worthwhile. Above all, my journey has taught me that it's OK to follow your heart, even when the path seems unconventional. ■

Muhammed Rashid is a Ph.D. student at the Indian Institute of Technology Bombay. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

1848 SOCIETY
AAAS



An estate gift to AAAS

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

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

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

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

—Prof. Elisabeth Ervin-Blankenheim, 1848 Society member

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

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

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

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

cut here

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

1848 SOCIETY
AAAS

PLEASE CONTACT ME AT:

Name: _____

Address: _____

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

Email: _____ Phone: _____

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