AAAS

Science

SPECIAL ISSUE

plants & heat

12 JUNE 2025

How warming will worsen conflict over Indus waters p. 1118

Jaundice limits malaria severity pp. 1132 & 1181

Sensing magma's moves minutes before an eruption p. 1189

ASU Science

PRIZE FOR TRANSFORMATIONAL IMPACT

Apply now to the Arizona State University and *Science* Prize for Transformational Impact!

Arizona State University and *Science* have partnered to create the **ASU–*Science* Prize for Transformational Impact**. This prize recognizes transformational research that uses innovative methods and approaches to identify problems and develop solutions with impacts on policy and decision-making.

The grand prize winner will receive a prize of **US$30,000** and their prize-winning essay will be published in *Science* online and print. A runner-up will receive **US$10,000** and have their essay published in *Science* online.

APPLY BY **AUGUST 15, 2025**

www. Science.org/ASU

CONTENTS

12 JUNE 2025 | VOLUME 388 | ISSUE 6752

Thermal imagery captures variation in temperature across a forest canopy in Colorado, US.

CREDIT: ANDREW D. RICHARDSON ADAPTED BY M. HERSHER/*SCIENCE*

ON THE COVER

Infrared imaging shows heat levels in sun-exposed leaves of *Alstonia scholaris* from the Australian Wet Tropics. As climate change increases temperatures and the severity of heat waves, both natural ecosystems and agricultural plants are increasingly affected by heat. Examining heat responses at cellular, genetic, physiological, and ecosystem scales, this special issue explores how plants sense and respond to high temperatures.
See the section beginning on page 1146.
Image: Kali Middleby

SPECIAL ISSUE

Plants & heat

FEATURES

1125 Hot zone
In the world's hottest forests, scientists are probing how plants cope with rising temperatures —E. Stokstad
PODCAST

BOOKS ET AL.

1142 North America's last prairies
The region's grasslands—vital ecosystems and carbon sinks—have been farmed and ranched beyond recognition —S. Boon

INTRODUCTION

1146 Plants facing the heat
—Madeleine Seale, Bianca Lopez, Sarah Lempriere

PERSPECTIVES

1148 Can wild plant adaptations help crops tolerate heat?
Wild plant species harbor a vast but largely unknown diversity of temperature stress solutions —S. Yeaman

1150 Plant microbiomes feel the heat
Rising temperatures change the structure and function of plant microbial communities —A.-L. Laine and S. Leino

REVIEWS

1153 Safeguarding crop photosynthesis in a rapidly warming world
—C. J. Bernacchi *et al.*

1161 Dispersed components drive temperature sensing and response in plants
—A. Singh Yadav *et al.*

1167 Scaling plant responses to heat: From molecules to the biosphere
—M. E. K. Evans *et al.*

ON THE PODCAST
Hot forests as testbeds for climate change

EDITORIAL

1113 An American research crisis with real human consequences
—M. B. Parlange

1115 Strangling intellectual independence
—M. S. Harris

NEWS

1116 Did science reformers play into Trump's hands?
Reform movement should have seen call for "gold standard science" coming, critics say
—C. O'Grady

1118 Warming threatens fragile pact over Indus River waters
Scientists say India-Pakistan treaty needs to be rethought for a changing world —S. Pathak

1120 Trump budget proposes killing nursing research institute
A tiny sliver of NIH, the institute has provided evidence base for bedside care —M. Wadman

1121 Many scientific societies are losing publishing revenue
Driven largely by open access, the trend puts society programming at risk —J. Brainard

1122 Proposed LIGO cut could snuff out brand new form of astronomy
"You're killing a newborn baby," says one astrophysicist —A. Cho

1124 DNA in the air could track biodiversity—and humans
Technology could be a boon for science, but raises ethical concerns
—W. Cornwall

COMMENTARY

PERSPECTIVES

1130 Global nitrogen budget revisited
Bottom sediments are important for nitrogen production in inland and coastal waters —P. L. Brezonik
RESEARCH ARTICLE p. 1205

1132 Rethinking jaundice
A buildup of unconjugated bilirubin may be a protective response to malaria —J. Kloehn and D. Soldati-Favre
RESEARCH SUMMARY p. 1181

1133 Synthesizing vascular trees at speed
A computational algorithm can render a complex artificial vascular structure in minutes —Y. Y. S. Huang and L. A. Ju
RESEARCH ARTICLE p. 1198

1134 Bacteria poison and eat their neighbors
Bacteria leverage a secretion system to kill and scavenge nutrients from nearby competitors
—C. D. Nadell and C. J. Marx
RESEARCH ARTICLE p. 1214

POLICY FORUM

1136 A liability framework for high-risk neural devices
A no-fault compensation scheme may help balance innovation and patient protection —A. Rotenberg *et al.*

1138 New case law and liability risks for manufacturers of medical AI
Recent case law can shape how innovation unfolds —S. Gerke and D. A. Simon

BOOKS ET AL.

1141 Dark matters
A new film celebrates the subterranean —N. J. Dominy

LETTERS

1143 The hidden costs of gold mining in Ecuador
—D. Cogălniceanu and D. Szekely

1143 Address Arctic shipping crisis at COP30 —Z. Du *et al.*

1144 US tariffs jeopardize medicine supply chains
—G. Casiraghi *et al.*

RESEARCH

HIGHLIGHTS

1174 From *Science* and other journals

RESEARCH SUMMARIES

1177 Bacterial immunity
Cat1 forms filament networks to degrade NAD^+ during the type III CRISPR-Cas antiviral response
—C. F. Baca *et al.*

1178 Speciation
Introgression dynamics of sex-linked chromosomal inversions shape the Malawi cichlid radiation
—L. M. Blumer *et al.*

1179 Cancer genomics
Aberrant basal cell clonal dynamics shape early lung carcinogenesis
—S. Gómez-López *et al.*

1180 Enzymology
Evolutionary-scale enzymology enables exploration of a rugged catalytic landscape —D. F. Muir *et al.*

1181 Malaria
A metabolite-based resistance mechanism against malaria
—A. Figueiredo *et al.*
PERSPECTIVE p. 1132

1182 Electrochemistry
Acid-humidified CO_2 gas input for stable electrochemical CO_2 reduction reaction —S. Hao *et al.*

RESEARCH ARTICLES

1183 Device physics
Gate-driven band modulation hyperdoping for high-performance p-type 2D semiconductor transistors
—B. Zhao *et al.*

1189 Volcanology
Minute-scale dynamics of recurrent dike intrusions in Iceland with fiber-optic geodesy —J. Li *et al.*

1194 Photophysics
Differential absorption of circularly polarized light by a centrosymmetric crystal
—K. A. Parrish *et al.*

1198 Tissue engineering
Rapid model-guided design of organ-scale synthetic vasculature for biomanufacturing
—Z. A. Sexton *et al.*
PERSPECTIVE p. 1133

1205 Nitrogen fixation
Global importance of nitrogen fixation across inland and coastal waters —R. W. Fulweiler *et al.*
PERSPECTIVE p. 1130

1210 Climate
Observed trend in Earth energy imbalance may provide a constraint for low climate sensitivity models
—G. Myhre *et al.*

1214 Microbiology
Antagonism as a foraging strategy in microbial communities
—A. K. M. Stubbusch *et al.*
PERSPECTIVE p. 1134

1218 Virology
Pan-viral ORFs discovery using massively parallel ribosome profiling
—S. Weingarten-Gabbay *et al.*

1225 Molecular biology
RNA transcripts regulate G-quadruplex landscapes through G-loop formation —K. Sato *et al.*

WORKING LIFE

1234 How burnout snuck up on me —M. D. Xames

1114 *Science* Staff
1233 *Science* Careers

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, commentary, 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.** *Science* (ISSN 0036-8075) is published weekly on Thursday, 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. The title *Science* is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $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.

An American research crisis with real human consequences

Marc B. Parlange

Ivy League universities have dominated recent news headlines, having become popular targets for critics of higher education. But the threats they face—cuts to federal research funding, assaults on academic freedom, and bans on admitting international students—extend far beyond their campuses. Research universities across the country—large and small, public and private—are grappling with these same pressures. These institutions are behind the breakthroughs that shape daily lives. Undermining them doesn't just jeopardize higher education, it threatens national and global strength. This means that economic, technological, and intellectual collapse is inevitable if US research institutions fall to federal and state disinvestment.

Universities with vibrant research communities are feeling the impact of growing hostility toward academia. In recent years, higher education has been painted as out of touch with "real America"—elite and disconnected. However, public research universities, like the University of Rhode Island (URI, where I am president), are by definition "real America." URI was founded under the Morrill Land-Grant Act of 1862, which gave federal land to states to create colleges for the working class. These weren't conceived as ivory towers—they were intended to be institutions that democratized knowledge, expanded opportunity, and equipped a growing young nation with practical skills in agriculture, science, and mechanical arts.

Today, there are 105 public and 7 private land-grant institutions, serving students in every state, the District of Columbia, and 5 inhabited US territories. This network includes 19 Historically Black and 35 Tribal Colleges and Universities. These institutions continue to embody the land-grant research mission, educating nurses who staff community hospitals, teachers working in public schools, and engineers translating research into companies. Students in land-grant schools are often the first in their families to attend college, and faculty research is tied to both local industries and national priorities. Defunding these universities doesn't punish a few—it hurts working families, veterans, immigrants, and young people striving for something better. In attacking these institutions, critics irreparably harm the people and the nation that these universities were built to serve.

Public research institutions are as vital to the national research enterprise as their billion-dollar–endowed private counterparts. These schools are the lifeblood of local communities and are especially vulnerable to budget cuts and federal policy shifts. There are 187 universities designated "R1"—a distinction given to schools that conduct the most intensive and high-impact research in the nation. They include universities like URI, whose impact far outpaces limited research budgets, and schools that invest billions each year in research. But keeping national research efforts strong means supporting all R1 universities. These universities are where research shapes teaching and where solutions are developed by directly addressing local needs. When they falter, there is a ripple effect on everyday lives.

Public research institutions... are the lifeblood of local communities and are especially vulnerable to budget cuts...

Unfortunately, we are seeing this now. For example, URI—Rhode Island's flagship research university—has received termination notices on more than 25 awards totaling $49 million. Critical initiatives have been halted, including coastal resiliency efforts and Department of Defense projects. Public health research programs have been eliminated, including those focused on community health and infectious disease prevention. Awards from the National Institutes of Health (NIH) that support students in biomedical research have been canceled. Grant-funded jobs have been cut, affecting the employment of local residents. Loss of federal support does not just mean fewer publications but also the end of clinical trials, increased threats to national security, the disappearance of advances in natural disaster forecasting and mitigation, and fewer nurses, teachers, and pharmacists. This is a crisis with real human consequences.

We have seen what happens when research investment dwindles. After the 2008 financial crisis, US federal budgets were slashed, resulting in fewer grants and a wave of researchers leaving academia. Promising engineering and agricultural projects were delayed or abandoned. NIH-backed studies were paused. It took more than a decade to rebuild labs, rehire talent, and regain momentum in critical fields.

Over the past 17 years, the US has been recovering from this recession, but the resurgence didn't happen by chance—it took sustained investment, bipartisan support, and public trust in science. The real beneficiaries of this productivity are not the privileged few—they are farmers raising climate-resilient crops, patients receiving better treatments, and small businesses spun out of university campuses. And just as research institutions are regaining ground from the pandemic, a different kind of threat has emerged. The erosion of federal support is already underway, and if it continues, the consequences will be swift, far-reaching, and perhaps even irreparable. □

Marc B. Parlange is president of the University of Rhode Island, Kingston, RI, USA. marc.parlange@uri.edu

10.1126/science.adz6477

Science AAAS

EDITOR-IN-CHIEF **Holden Thorp**, hthorp@aaas.org

EXECUTIVE EDITOR Valda Vinson

EDITORS, RESEARCH Sacha Vignieri, Jake S. Yeston **EDITOR, COMMENTARY** Lisa D. Chong

DEPUTY EXECUTIVE EDITOR Lauren Kmec

DEPUTY EDITORS Stella M. Hurtley (UK), Phillip D. Szuromi **SENIOR 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 **SENIOR PROGRAM ASSOCIATE** Maryrose Madrid **EDITORIAL MANAGER** Joi S. Granger **EDITORIAL ASSOCIATES** Aneera Dobbins, Lisa Johnson, Jerry Richardson, Anita Wynn **SENIOR EDITORIAL COORDINATORS** Alexander Kief, Ronmel Navas, 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** Carly Hayward (UK) **ASI SR. OFFICE ADMINISTRATORS** Simon Brignell (UK), Jessica Waldock (UK) **COMMUNICATIONS DIRECTOR** Meagan Phelan **DEPUTY DIRECTOR** Matthew Wright **SENIOR WRITERS** Walter Beckwith, Joseph Cariz, Abigail Eisenstadt **WRITER** 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) **SENIOR 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, Sara Reardon, Richard Stone (Senior Asia Correspondent), Gretchen Vogel (Berlin), Lizzie Wade (Mexico City) **INTERN** 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 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 **PUBLISHING PLATFORM MANAGER** Jessica Loayza **PUBLISHING SYSTEMS SPECIALIST, PROJECT COORDINATOR** 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 **SENIOR TRAFFIC ASSOCIATE** Christine Hall

DIRECTOR, PRODUCT MANAGEMENT Kris Bishop **PRODUCT DEVELOPMENT MANAGER** Scott Chernoff **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

Strangling intellectual independence

Michael S. Harris

The phrase "Sputnik moment" is often used to describe a moment when a country—usually the United States—needs to respond to some technological leap made by another nation. The wake-up call is meant to provoke more investment in research, development, and education. Today, the United States faces another Sputnik moment, but this time, the threat isn't coming from abroad—it's coming from within.

When the Soviet Union launched its Sputnik satellite in 1957, American leaders realized that science and education were national imperatives essential for the long-term economic and national security of the United States. The end result? The National Defense Education Act, along with massive federal investment in research and development, ushered in an era of US higher education that positioned American universities at the forefront of global innovation.

American higher education is now at risk—not because of any single rival, but because of two self-inflicted wounds. It's hard to imagine a more destructive one-two punch than what is happening today. One blow has been the federal government's ideologically driven cuts to research funding at universities. At center stage have been the nation's premier private institutions, with Harvard University bearing the brunt of attack at the moment. The other jab has been the ideologically driven attacks on academic freedom at state universities, as typified by recent legislation in Texas. Last week, Texas lawmakers pushed Senate Bill 37 across the finish line, giving the state's political appointees control of faculty governance as well as the design of curricula across its public university system, one of the largest in the country. Noncompliance could result in the loss of already appropriated funding from the state, among other penalties. This extreme form of control—described by state political leaders as necessary to ensure academic accountability and eliminate political advocacy—amounts to censorship.

Texas is home to some of the nation's top institutions of higher education. Its 106 public universities, community colleges, and medical schools are inhabited by over 50,000 faculty and 1.2 million students. Over the past 3 years, the state has undertaken an aggressive campaign to reshape higher education in its political image. State bills have already passed to eliminate diversity, equity, and inclusion initiatives and limit tenure protections. The new legislation attacks academic freedom further by curtailing the authority of faculty senates—for example, by imposing term limits on faculty senators, limiting a senate's scope of authority, and requiring that half of senators be appointed bythe university president. Appointed senators can serve 6 years, whereas elected senators may only serve 2 years. The legislation also requires that certificates and minor degree programs have workforce outcomes, a move intended to close off certain areas of study. It also establishes a process whereby any individual can report institutions for investigation on issues such as which courses are included in general education. These are not simple administrative changes or modifications to ensure accountability, as proponents claim. They are fundamental incursions into the intellectual independence of universities.

Most alarming is how these two trends—politicized attacks on academic freedom and declining federal research support—reinforce each other.

When politicians dictate what topics are acceptable in classrooms or the kinds of research that can be carried out in universities, they don't just chill inquiry, they suffocate it. Scientific breakthroughs do not emerge from politically convenient consensus. More often, they come from scholarly challenge to the status quo, asking uncomfortable questions, and crossing disciplinary and methodological barriers. Undermining academic freedom doesn't just slow progress—it shuts down the very engines of discovery.

Meanwhile, at the federal level, the Trump administration slashed funding for major areas of research—climate science, reproductive health, racial and educational disparities—not out of fiscal necessity, but political hostility. Proposed reductions to research budgets and operational support reflect a dangerous departure: Science is no longer seen as a common public good, but as a political battleground. Yet, other countries are sprinting ahead. China, Japan, South Korea, and the European Union, for example, are investing heavily in artificial intelligence, biotechnology, renewable energy, and other major science priorities.

Most alarming is how these two trends—politicized attacks on academic freedom and declining federal research support—reinforce each other. Why would world-class researchers stay in US universities under these conditions? Why would the best international students come to institutions under siege when other countries offer stability, funding, and respect for academic expertise?

With the greatest threat to American science no longer orbiting overhead but sitting in state legislative chambers and the White House, will the United States have the courage to respond as strongly as before? The world will not wait while American researchers are defunded and muzzled. No nation can keep pace in the race for global leadership in innovation by politicizing its universities. □

Michael S. Harris is a professor of higher education at Southern Methodist University, Dallas, TX, USA. harrism@mail.smu.edu

Published online 5 June 2025; 10.1126/science.adz5024

POLICY

Did science reformers play into Trump's hands?

Reform movement should have seen call for "gold standard science" coming, critics say

CATHLEEN O'GRADY

The science reform movement—which aims to improve the rigor of research—has unwittingly handed the administration of President Donald Trump a way to attack science, critics say.

Trump's executive order calling for "gold standard science"—widely interpreted as an attempt to target findings that are inconsistent with the administration's political agenda—was a foreseeable outcome of a movement that overstated the problems in science, according to some researchers. But science reformers say integrity concerns are too crucial to tamp down over fears of misuse.

The 23 May order points to a "reproducibility crisis" and high-profile cases of data falsification. It calls for science to be transparent and reproducible, and to acknowledge uncertainty. Those goals are familiar to Brian Nosek, executive director of the Center for Open Science (COS), a nonprofit that has pushed to improve science through measures such as sharing data openly and describing a study plan in full before embarking on it. But 6 days after Trump signed the order, COS released a statement describing the dangers it poses and how its goals depart from those of the reform movement.

Nosek says the order treats ideas for reform as requirements rather than targets, when even the best studies are unlikely to meet every criterion: "The implication is that policymakers won't be able to use the best available evidence. … And that is in nobody's interest." The order also empowers political appointees to enforce its vision of scientific integrity, meaning researchers whose work is deemed to fall short face disciplinary action, Nosek says. And it could also make it easier for courts to throw out scientific evidence that's not deemed up to standard, warns Joe Bak-Coleman, a collective behavior scientist at the University of Washington.

Bak-Coleman and others say the reformers should have seen that their claims and concerns could be weaponized by antiscience movements, and done more to prevent that. "The writing was on the wall in huge letters," says Stephan Lewandowsky, a cognitive scientist at the University of Bristol who points to attempts by the tobacco industry to introduce "sound science" legislation and sponsor "research integrity" symposia as early as the 1990s.

> **"The writing was on the wall in huge letters.**
>
> **Stephan Lewandowsky**
> University of Bristol

The reform movement gathered momentum in psychology in the early 2010s, when high-profile failed replications raised concerns that much of the psychological literature was not robust. Critics pointed to problems such as publication bias—the tendency of journals to only publish positive findings and reject studies that found no effect, skewing the overall picture of the literature. They also noted some scientists' tendency to dig around in their data until they came up with favorable results.

These and other "questionable research practices" were driving a literature full of false findings, the reformers said. Various studies subsequently found that many psychology findings could not be replicated. Similar concerns were raised in economics, neuroscience, and other fields, giving rise to the idea of a "replication crisis."

But the "crisis" narrative was overblown, Bak-Coleman and others say. "Suddenly the whole conversation changed from 'some of these fields are having certain issues' to 'science is in crisis,'" says University of Idaho metascientist Berna Devezer. That was unwarranted, she says. For a start, some of the problems driving psychology's replication issues—such as fuzzy measurements of ill-defined concepts—do not plague many scientific fields. And she asks whether questionable research practices are widespread enough even within psychology to support claims of a crisis.

Nosek agrees that the term has been unhelpful: "We don't use the crisis narrative." Still, media coverage amplified the idea of science in crisis, and some people began to use the idea of a "replication crisis" to justify antiscience moves, Devezer says. In 2017, reformers were alarmed when the Republican-controlled U.S. House of Representatives passed the HONEST Act, which aimed to restrict the Environmental Protection Agency to using only evidence with publicly available data and materials. The HONEST Act later died in the Senate, but conservative think tanks like the Heritage Foundation began to emphasize the idea that science is in crisis and needs reform, Lewandowsky says. "That narrative has been carefully constructed and curated over the last 10, 20 years."

Bak-Coleman thinks metascientists and science reformers should have been more careful about how they communicated their ideas, avoiding the claim that the problems they identified affect all of science. Reformers could also have communicated with more humility and less

certainty about the prevalence of questionable research practices and other issues, Devezer says. Although it's difficult to predict how claims could be misused, "you don't want to make it easy," she says.

And reform advocates could also have done more to contextualize the failings of science within the bigger picture of its successes, Lewandowsky says. "I'm old enough to remember when HIV was a death sentence," he says. "And now you can get life insurance." There are significant problems in science, but its achievements are overwhelming, and a narrative that all of science is in crisis and needs total reform "throws the baby out with the bathwater, twice over."

Devezer says reformers could learn from geneticists, who have long grappled with racists who weaponize and distort their findings. In both fields, "mostly good-faith people are making very strong claims that can easily be picked up by people in bad faith and used to achieve their ideological goals," says Kevin Bird, an evolutionary biologist at the University of California, Davis. But there wasn't as much warning for science reformers that their work could also be misused, he says.

Many geneticists have learned to write defensively, anticipating how their work could be distorted. They frame their papers carefully, understanding that many readers may never see more than the abstract, which makes it unhelpful to bury caveats deeper in the paper or in separate blog posts, Bird says. He thinks the executive order should be a wake-up call for science reformers to communicate more carefully.

Lewandowsky agrees. The work of reform is necessary, he says, but "let's make sure that it isn't done in a way that inadvertently makes us useful idiots of an antiscience movement." □

IN FOCUS

PHOTO: KONRAD WOTHE/PICTURE PRESS/REDUX

U.S. military trims access to sea ice measurements

A long-standing record that tracks shrinking Arctic sea ice is at risk, after the U.S. Department of Defense (DOD) told climate scientists last month it would "deprioritize" access to data collected by military weather satellites. "We're not getting all the data anymore, at least regularly," says Walt Meier, a remote sensing scientist at the U.S. National Snow and Ice Data Center. NSIDC, which produces the record for a wide variety of users, said last week it would transition to using data from a Japanese satellite. Cloud-penetrating microwave sensors on three satellites in the U.S. Defense Meteorological Satellite Program are used not only to track wind speeds over the oceans, but also to detect reflections off sea ice. In May, however, DOD officials told NSIDC scientists that because of staffing cuts at the facility that receives and processes the data, the Fleet Numerical Meteorology and Oceanography Center, they could not continue to reliably provide full coverage of sea ice. By October 2026, DOD plans to turn off all three of the satellites, which are aging. A new one, the Weather System Follow-on–Microwave satellite, was launched last year but is not supplying data to NSIDC. —*Paul Voosen*

The Indus River, shown here near the border of India and Pakistan, is a major source of water for irrigation and hydropower.

PHOTO: UMAR ALTAF/MIDDLE EAST IMAGES/AFP VIA GETTY IMAGES

SCIENCE & POLICY

Warming threatens fragile pact over Indus River waters

Scientists say India-Pakistan treaty needs to be rethought for a changing world **SUSHMITA PATHAK**

The world relaxed a bit last month after nuclear-armed neighbors India and Pakistan agreed to end days of air and artillery strikes ignited by a terrorist attack on Indian tourists in Kashmir, a disputed Himalayan territory split between the countries. That truce, however, did not end another high-stakes conflict between the two nations, this one over water.

In response to the 22 April attack that killed 26 people, India said it was putting "in abeyance" its participation in the Indus Waters Treaty (IWT), a 65-year-old pact that governs how Pakistan and India share water from one of Asia's major river systems. Pakistan, in turn, said it would consider any move by India to withhold Indus water, which irrigates most of its farmland, as an act of war. Researchers say the jousting highlighted not only the political sensitivity of the treaty, but also the increasingly urgent need to update the IWT to reflect the ongoing impacts of climate change and population growth.

"The treaty was crafted for a river system that no longer exists," says Fazlul Haq, a geographer at Ohio State University. The current tensions mean any renegotiation won't happen soon. But in the long run, Haq says, "We need to rethink the IWT not as a static political agreement, but as a living, evolving environmental treaty."

The Indus originates in the rugged mountains of Tibet's Hindu Kush Himalayas (HKH), then flows some 3180 kilometers through India and Pakistan to the Arabian Sea. The IWT gives Pakistan rights to water in most of the main stem of the Indus as well as two of its western tributaries, the Jhelum and the Chenab; India controls three eastern tributaries, the Ravi, Sutlej, and Beas.

When negotiators finalized the pact in 1960, they did not foresee global warming, which now poses an existential threat to the glaciers and snowpack that feed the system, or the spiraling demand for water from the basin's human population, now nearly 300 million and still growing. As a result, the Indus is now "the most sensitive river basin from both climate as well as socio-economic point of view," says climate scientist Arun

Shrestha at the International Centre for Integrated Mountain Development (ICIMOD).

At current emissions rates, the HKH region will warm by up to 2.2°C by the end of the century, Shrestha says. Its glaciers are already feeling the heat: They disappeared 65% faster in the 2010s compared with the previous decade and could lose up to two-thirds of their volume by 2100, a 2023 ICIMOD study found.

The nature of regional precipitation is also changing, says glacio-logist Shakil Romshoo at the Islamic University of Science and Technology. There's less snowfall and more rain, which shrinks glaciers, and the snow that falls now melts more quickly. For the past 3 years, snow persistence, or the time snow remains on the ground, has been below normal in the Indus Basin.

Such changes are expected to alter when seasonal water flows reach their peak. Currently, rivers in the Upper Indus Basin see maximum flows between June and August, just when farmers need water for irrigation. But the precipitation and temperature changes could shift this peak to earlier in the year, Romshoo's research has found. Sometime around the middle of the century, overall water flows in the Indus Basin are forecast to reach a modern-era peak as glacier melt accelerates. "But after that, the water availability will start to diminish" because of a lack of snow and ice, Shrestha says.

Those forecasts, however, remain uncertain because of sparse data. For example, researchers currently monitor just 16 of the roughly 33,000 glaciers in the Indian Himalayas, Romshoo says. And because the monitoring stations tend to be at lower altitudes, they provide limited insight into how weather phenomena, such as storms known as the western disturbances, might dump snow at high altitudes and help compensate for the loss of ice, Shrestha says. There's also been little study of how other factors, such as the debris and black carbon particles that settle on glaciers and snow, might influence melting. "We need to continue improving our understanding," Shrestha says.

Better data could help negotiators develop a more flexible IWT, able to respond to changing river flows. "Rather than allocating fixed river quotas, the treaty should adopt dynamic flow-based management," Haq says. If satellite data reveal sparse snow cover, for example, managers could plan ahead to institute water conservation measures or reschedule reservoir releases. Dynamic management could also help managers plan to divert or store excessive flows, such as the floods that result when lakes filled with glacier meltwater burst through ice dams. Such "outburst floods" are expected to increase in the future.

Advances in remote sensing and other technologies make dynamic management more feasible than ever, researchers say. But developing robust monitoring networks will require funding that, so far, is in short supply. And simply collecting more data won't be enough, observers note: India and Pakistan would also need to share them. The IWT already mandates regular information exchange, but Pakistan has often accused India of sharing only limited hydrological information.

For the moment, India is withholding all data, saying it won't comply with the IWT until Pakistan cracks down on terrorists on its soil. That could "lead to uncertainties about water availability downstream in Pakistan," says Srinivas Chokkakula at the Centre for Policy Research in New Delhi. India is also reportedly planning to periodically flush sediment from reservoirs behind hydropower dams on the Chenab, sending bursts of turbid water downstream, without first warning Pakistan.

The IWT bars each nation from unilaterally exiting the treaty. But India's move could mean the suspension of a bilateral body called the Permanent Indus Commission, Chokkakula says, giving Pakistan no forum for airing or resolving complaints. Even before this latest standoff, deep-rooted distrust has stalled efforts to negotiate changes to the IWT. "Any attempt to revise the treaty is immediately perceived as a zero-sum move" that would benefit only one nation, Haq says.

Amid that stalemate, regional temperatures are continuing to climb. "Geopolitical conflict can suspend cooperation," Haq says. "But climate change can render the treaty itself obsolete." □

Sushmita Pathak is a journalist based in New Delhi.

TRUMP TRACKER

NIH KILLS EMERGING DISEASE NETWORK

President Donald Trump's administration is shuttering a network of National Institutes of Health (NIH)-funded centers that aims to prevent future pandemics. The 10 Centers for Research in Emerging Infectious Diseases (CREID) were launched 5 years ago with a planned $82 million in funding to collect, characterize, and respond to viruses and other pathogens that could jump from animals to people. NIH expected to renew the network this year. But a 5 June stop-work order states its work "has been deemed unsafe for Americans and not a good use of taxpayer funding." Virologist David Wang of Washington University in St. Louis, who led one center, calls that logic "totally unsubstantiated." The move, he says, "is incredibly short sighted and actually makes America and the whole world less safe." —*Jocelyn Kaiser*

PUSHBACK ON NEW DEI RULES

Williams College, an elite liberal arts school in Massachusetts, says it is not "comfortable" accepting federal research grants until Trump's administration clarifies a new requirement that grantee institutions certify they aren't doing anything to "promote or advance diversity, equity, and inclusion." Williams appears to be the first U.S. institution to have taken such a dramatic stance since NIH adopted the new rule in April and the National Science Foundation followed suit last month. "This new condition goes well beyond a standard certification that the college comply with all applicable nondiscrimination laws," two school officials wrote in a 30 May memo to faculty. "Some people believe the institution is taking a moral stand and living up to its values," says a faculty member who requested anonymity. "But others say it is signaling it doesn't care about research." —*Jeffrey Mervis*

IN OTHER NEWS

NIH STAFF PROTEST TRUMP POLICIES Nearly 300 current and laid off scientists and staff at the National Institutes of Health released a statement on 9 June charging NIH officials and their superiors with politicizing science. The signers, about 75% of them anonymous, implore NIH Director Jayanta "Jay" Bhattacharya to reverse cuts to external grants on topics disliked by conservatives and undo funding freezes to Harvard, Columbia, and other universities that President Donald Trump has targeted. The Bethesda Declaration, named for NIH's main campus in Maryland, also aims to test Bhattacharya's avowed support for scientific dissent, organizers say. More than 13,000 people signed an open letter supporting the statement, including at least 22 Nobel laureates. The declaration came the day before Bhattacharya responded to a Senate spending committee's questions about a White House plan to slash the agency's $47 billion budget by 40% in 2026 by saying determining a final budget would be "a collaborative effort" with Congress. *—Jocelyn Kaiser*

PHOTO: LEILA NAVIDI/*MINNEAPOLIS STAR TRIBUNE* VIA ZUMA

RESEARCH POLICY

Trump budget proposes killing nursing research institute

A tiny sliver of NIH, the institute has provided evidence base for bedside care

MEREDITH WADMAN

Hospital patients and their families often learn that nurses know far more about their condition and care than doctors. Last year, a clinical trial published in *Nature Medicine* underscored that with an eye-popping number: Hospitalized patients were 36% less likely to die if their caregivers used an artificial intelligence–based system that analyzes patterns of nurses' observations to provide an early warning that patients are deteriorating.

Now, President Donald Trump's proposed 2026 budget asks Congress to eliminate the tiny, low-profile institute within the U.S. National Institutes of Health (NIH) that made the study possible. The National Institute of Nursing Research (NINR), which focuses on building evidence to improve patient care, from hospitals to nursing homes to schools, supported the development of the early warning system and the recent clinical trial. NINR also provided training for several members of the Columbia University–led team behind the research.

When she saw the proposal to abolish NINR, "I was like: 'You've got to be kidding me,'" says Martha Curley, a clinical trialist in pediatric intensive care nursing at the Children's Hospital of Philadelphia. "What are the patients and families going to do if they don't have any science to support how they can live with, and die with, their illnesses?" (NINR has funded key research in palliative care.)

"If NINR goes away, nursing science will go away," adds Michele Balas, associate dean for research at the University of Nebraska Medical Center College of Nursing. "That is our primary institute. There will be no funding to teach people how to do research in nursing."

NINR, founded as the NIH Center for Nursing Research in 1986 and elevated to institute status in 1993, is NIH's third smallest. Its $198 million budget accounts for 0.4% of the $48.5 billion agency's spending. But its backers say it serves a unique and critical function with its focus on the whole patient and on practical measures for improving patient care.

"You are talking about a very small institute in terms of investment that has yielded a very massive impact on the lives and well-being of the people we serve," says Mary Naylor, a nurse-scientist at the University of Pennsylvania (UPenn) School of Nursing. Her seminal, NINR-funded trial, published in 1999, showed that sending elderly hospitalized patients home with transitional nursing care from

Two nurses care for a girl at Children's Hospital in St. Paul, Minnesota. The National Institutes of Health's nursing institute supports research on patient care.

highly qualified nurses significantly reduced hospital readmissions and per-patient health care costs.

Another landmark study documented a 7% increase in patient death rates with each patient added to a single hospital nurse's workload. Based in part on this finding, Nevada lawmakers last week passed a bill setting limits on the number of patients nurses can be assigned. (Hospitals oppose the bill, and it's not clear that Governor Joe Lombardo [R] will sign it.)

Proponents of NINR also stress its important role in training nurse-scientists. The institute spent $6.6 million on training grants in 2024—the second-largest proportion of an NIH institute budget spent on training, after the National Institute of General Medical Sciences. NINR's focus on advanced training is "what allows nursing schools to have expert faculty, which is a major constraint on growing the nursing workforce," says Matthew McHugh, a nurse-scientist who is director of the Center for Health Outcomes and Policy Research at the UPenn School of Nursing. A dearth of faculty is forcing nursing programs to turn away thousands of qualified applicants, compounding a shortage of nurses in the United States, particularly in some states and in rural areas.

The Trump administration's budget, released in final form on 30 May, also proposes eliminating three other small institutes and centers, among the 27 that make up NIH: the Fogarty International Center, the National Institute on Minority Health and Health Disparities, and the National Center for Complementary and Integrative Health. Andrew Nixon, a spokesperson for the Department of Health and Human Services, NIH's parent agency, said in a statement: "The Budget reflects needed reforms to put health care spending on a sustainable fiscal path and proposes policies to Make America Healthy Again."

Earlier, on 1 April, the Trump administration placed Shannon Zenk, NINR's director since 2020, on administrative leave and offered her a transfer to a position with the Indian Health Service in one of several western states. Zenk had been outspoken in promoting a focus on health equity at NINR, a research topic Trump has aggressively targeted. The institute's strategic plan for the years 2022–26 says advancing health equity is integral to its mission, and declares advancing diversity, equity, and inclusion a guiding principle. Zenk could not be reached for comment.

NIH Director Jayanta "Jay" Bhattacharya did not respond to written questions from *Science* about NINR's proposed elimination. But at a 10 June Senate committee hearing, he said he "absolutely" supports nursing research. Senator Jeff Merkley (D–OR), whose wife is a nurse, asked the NIH director why he would back NINR's elimination when "such important work is done through that institute." Bhattacharya pledged he would not eliminate NINR without congressional approval, but then added: "[Nursing] research needs to get done. … Where exactly it happens is less important."

However, nurse-scientists who have received NINR grants say they are skeptical that they would be able to pivot to find funding at the remaining NIH institutes, should Congress agree to eliminate NINR. "Those institutes are gonna have priority areas they need to fund … biomedical focused and basic science focused," Balas says. "The priorities will not be related to the questions that nursing science answers and that are important to patients and families."

> "If NINR goes away, nursing science will go away."
>
> **Michele Balas**
> University of Nebraska Medical Center College of Nursing

Nurse-researchers are already feeling the pain of Trump's broader cuts to NIH grants. In March, the team at Columbia responsible for the early warning system that slashed mortality in hospitalized adults learned that its NINR grant to continue the trial in pediatric and emergency room patients had been killed. □

RESEARCH COMMUNITY

Many scientific societies are losing publishing revenue

Driven largely by open access, the trend puts society programming at risk

JEFFREY BRAINARD

For many scholarly societies, selling subscriptions to the journals they publish has historically been a key source of revenue, helping subsidize other work such as advocacy and providing scholarships. But the trend toward open-access publishing is threatening that income stream even as costs are rising, a recent survey indicates. More than 90% of the 66 societies that participated said the revenue they collect from publishing journals is not keeping pace with inflation, and about half reported a decrease.

The respondents represent just a small fraction of all societies that publish journals, estimated to be in the thousands globally, and U.K. societies made up three-quarters. But Michael Clarke, managing director of Clarke & Esposito, a consulting company for scholarly publishers, says the findings are similar to what clients in several countries have told him: Many are seeing flat publishing revenues, which represent declines after accounting for inflation. (*Science* is produced by the scientific society AAAS, one of the larger scientific societies in the United States; *Science*'s News section is editorially independent.)

Of the respondents—almost all focused on science, engineering, or math—nearly half said they rely on publishing revenue for more than 50% of their total budget. Most respondents publish five or fewer journals, a relatively small volume that is typically less financially efficient than the stables of hundreds of journals at the largest publishers. Declining revenues have forced some to cut back on other work. "One society described having to considerably reduce the scale of charitable activities, while another ceased all new book commissioning," says Rob Johnson, managing director of Research Consulting, the firm that produced the report, published on 28 May. (It does not identify which societies participated in the survey, which the firm distributed largely with the help of publishing trade groups and scholarly society associations.)

"If we lose these societies, it's just going to consolidate more power and clout to a couple of large commercial publishers," says Kathryn Spiller, co-chair of the Society Publishers' Coalition, which has more than 120 members. Scholarly societies provide a range of services,

including advocacy on public policy, education, and career support for their members. Without them, Spiller says, "The world would be a very much poorer place."

Several industry trends have contributed to what Johnson calls "a slow-motion car crash" for societies. Historically, leading society journals have tended to pursue high selectivity and low publishing volume. That started to change in the 2000s with the advent of paid open-access publishing, in which authors pay a fee to make papers free to read when published—and which critics say incentivizes quantity over quality. Many societies hoped to gradually transition from subscriptions to open access. But, "We don't necessarily have the ability or the desire to suddenly grow" publishing volume, says Spiller, who is director of publishing at the society Applied Microbiology International. Even with open-access fees averaging about $2000 per paper among publishers of all types, societies tend to make less money than under the subscription model, Spiller says.

Rather than navigating the daunting transition to open access on their own, many societies have outsourced publishing operations to corporate giants such as Elsevier and Springer Nature in return for a negotiated percentage of revenues or profits. But with average revenue per paper dropping, commercial publishers' terms have become less generous over time, Johnson says. Some societies have settled for contract renewals providing 20% to 40% less money than they received in the past. "A highly selective [society] journal, where you're turning away 80% of your potentially revenue generating items, is not as attractive in that regime," he says.

Moreover, the big publishers typically make society journals part of increasingly common "transformative agreements" between publishers and institutions. Under such deals, a university or consortium pays a single price to enable its researchers to read the publisher's journal portfolio and publish papers open access without paying a per-article fee. How much revenue the publisher shares with a society whose journals are included in such agreements depends on the number of articles published under them—and if few of the university's scholars do, that can leave the society worse off than under the subscription model. "It could be that many of those [transformative] deals are not good deals for societies," Clarke says.

Like other publishers, societies also face increased costs, for example to check manuscripts for signs of research fraud. Artificial intelligence (AI) tools to spot red flags such as nonexistent or irrelevant citations may ultimately help improve efficiency and quality, but they require investments and come with steep learning curves. The survey respondents said AI represents both the biggest challenge and opportunity for them over the next 5 years.

Some societies are exploring new sources of revenue, such as online courses and magazines written for nonspecialist readers, Spiller says. "It does feel as if many societies are very, very overreliant on publishing income and are fully aware that they need to diversify." □

THEY SAID IT

Now the fear is that the ACIP will be filled up with people who know nothing about vaccines except suspicion.

Senator Bill Cassidy (R–LA), a physician, on the 9 June removal of all members of the U.S. government's Advisory Committee on Immunization Practices (ACIP), a key expert panel, by Health and Human Services Secretary Robert F. Kennedy Jr.

TRUMP BUDGET

Proposed LIGO cut could snuff out brand new form of astronomy

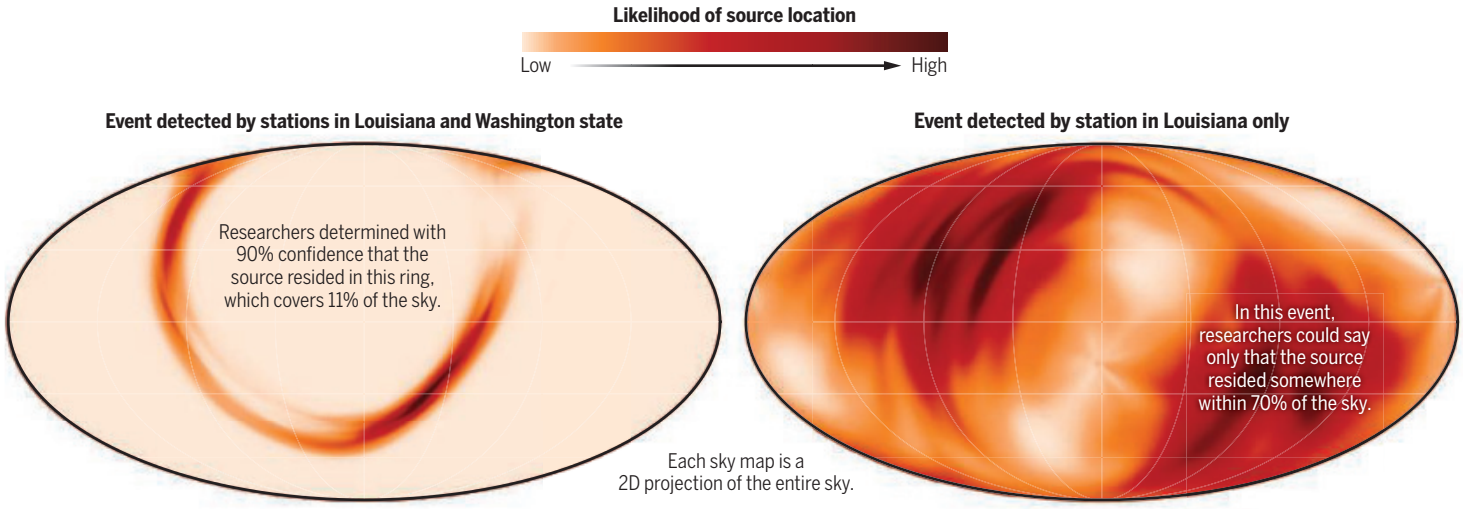
"You're killing a newborn baby," says one astrophysicist **ADRIAN CHO**

Just 10 years ago, 900 scientists working with two of the biggest and most sensitive scientific devices ever built made a discovery that opened new eyes on the universe. The twin detectors of the Laser Interferometer Gravitational-Wave Observatory (LIGO) spotted ripples in space itself set off when two massive black holes 1.3 billion light-years away spiraled into each other. For the first time, humans had detected an astrophysical object with something besides electromagnetic radiation. It was a dramatic demonstration of general relativity, Albert Einstein's theory of gravity, and it marked the birth of a new field, gravitational wave astronomy.

Now, however, a budget-cutting plan by the administration of President Donald Trump would close one of the two LIGO interferometers, enormous L-shaped optical instruments in which light bounces between mirrors in arms 4 kilometers long. The move would save roughly $20 million while drastically cutting LIGO's ability to identify and localize the cataclysmic celestial events that produce gravitational waves. Physicists say it would largely eliminate the field LIGO birthed. "You're killing a newborn baby," warns Gianluca Gemme, an astrophysicist with Italy's National Institute for Nuclear Physics and spokesperson for Virgo, a smaller gravitational wave detector near Pisa.

Three early leaders of the LIGO effort won the 2017 Nobel Prize in Physics for the first detection of a gravitational wave. The LIGO detectors in Louisiana and Washington state, working together with the less sensitive Virgo, have now captured signals from more than 300 sources, almost all black hole mergers. And all the detectors continue to gain sensitivity through upgrades, says David Reitze, a physicist at the California Institute of Technology and director of LIGO. "We're at the tip of the iceberg in terms of the amount of science we can do."

However, in its proposed budget for 2026, the Trump administration would slash the National Science Foundation's (NSF's) $9 billion budget by 57%, and, as part that cut, trim LIGO's annual operating budget from $48 million to $29 million. The proposed budget explicitly



It takes two

These sky maps show real data from two recent Laser Interferometer Gravitational-Wave Observatory events, one that registered a signal in both of the observatory's stations and one that produced a signal in only one of them. The comparison shows how triggering both detectors helps researchers locate the source of the gravitational waves.

calls for eliminating one of LIGO's interferometers, but doesn't specify which one. In a two-sentence email to *Science*, an NSF spokesperson said the plan reflects "a strategic alignment of resources in a constrained fiscal environment."

The loss would inflict disproportionate damage on the observatory's scientific reach, scientists say. Researchers would find it more difficult to distinguish a black hole collision in the distant cosmos from a nearby seismic tremor or anything else that shakes the remaining interferometer. A real gravitational wave likely produces signals in both U.S. detectors within 10 milliseconds, the time it takes the ripple to cover the cross-country distance. Without the ability to compare data from the two instruments, detections would be less sure and their rate would fall at least 25%, Reitze says, and possibly much more.

With a single detector LIGO would also lose the ability to roughly locate the source of a signal on the sky—the key to its working as an observatory. LIGO showed the power of source localization in 2017 when it, working with Virgo, detected two neutron stars spiraling into each other. Around the globe and in space, myriad telescopes across the electromagnetic spectrum wheeled to see the dramatic explosion that ensued, a so-called kilonova that produced a gamma ray burst and spewed gold and other newly formed elements. Having more than one LIGO detector made that possible, says Roger Blandford, a theoretical astrophysicist at Stanford University. "There's a reason they built two of them."

To locate a source, researchers use the arrival time of the signals at different detectors to triangulate. If Virgo and the two LIGO detectors all see a signal, they can now locate its source to as little as 5 square degrees. The two LIGO detectors can also locate a source to a reasonably small circle on the sky. But a single LIGO detector can do little better than half the sky, Reitze says. "Astronomers look at that and go, 'Why did you even wake us up?'"

Closing a LIGO interferometer would also likely mean cutting scientific staff, a prospect that worries some people more than anything else. "There's a whole culture there that has worked together to build these two magnificent facilities," Blandford says, "and you don't recreate this just by hiring some people in an ad." Gabriela Gonzalez, a physicist and LIGO member at Louisiana State University, notes that the majority of graduate students who work on LIGO go on to jobs in high-tech industries. "It's not just the science results, it's the people we train that will be lost."

France Córdova, an astrophysicist who served as NSF director from 2014 to 2020, says the proposed lopping of LIGO in half may be, oddly, a testament to the prestige of the project even in the White House's Office of Management and Budget (OMB). "Somebody in OMB understands that this was significant enough not to cancel it altogether, but didn't understand the science" well enough to know why it couldn't be cut in half, she speculates.

Physicists are still hopeful the proposed closure isn't set in stone. "We're having discussions with NSF to understand better where it came from and what it really means," Reitze says. "Is it binding language or is there some latitude?" Gonzalez notes that Congress sets the agency's final budget. "Legislators are smart," she says. "I know that they are in difficult positions, and they do a very difficult job, but they listen to people, and we are talking to them." □

IN OTHER NEWS

FLAWED STUDY SEARCH The nonprofit Center for Scientific Integrity received a $900,000 gift for a new effort to find flawed meta-analyses and other research articles that could distort medical treatment guidelines and harm patients. The Medical Evidence Project will be led by James Heathers, a data sleuth, for 2 years with funding from Open Philanthropy. It comes as the number of retracted papers has grown, along with fears that fraud and errors have polluted the medical literature. Heathers plans to promptly publish findings in the center's Retraction Watch blog because universities and publishers can take years to complete investigations of reportedly flawed papers and publicly share the results, which often lack details. —*Jeffrey Brainard*

CREDITS: (GRAPHIC) M. HERSHER/*SCIENCE*; (DATA) LIGO/NASA/LEO SINGER

ENVIRONMENT

DNA in the air could track biodiversity—and humans

Technology could be a boon for science, but raises ethical concerns **WARREN CORNWALL**

During a single week in April 2023, the area around Florida's Washington Oaks Gardens State Park was abuzz. A bobcat passed by, perhaps stalking the eastern gray squirrels. An eastern diamondback rattlesnake slithered through the undergrowth. The spaces among the grand oaks teemed with wildlife—a big brown bat, mosquitoes, and an osprey—and people.

Scientists didn't directly see any of these creatures. Instead, cutting-edge DNA technology captured evidence of them from tiny specks of organic material floating in the air. A similar analysis from the streets of Dublin also revealed a rich—though very different—tableau of life. Together the work, published last week in *Nature Ecology & Evolution*, shows the potential of airborne DNA-detection tools to surveil biodiversity, disease-causing microbes, and invasive species—and to imperil the privacy of people whose DNA is also captured. "It is boundary-pushing work," says Ryan Kelly, a University of Washington marine scientist who was not involved in the study.

> "It's a little bit of fear mongering, I fear."
> **Lara Urban**
> University of Zürich

Over the past decade, scientists have increasingly analyzed DNA in the water, air, and dirt, commonly known as environmental DNA or eDNA. Just last month a team reported it had used the approach to survey biodiversity across the United Kingdom.

But the most common method for analyzing eDNA samples, known as metabarcoding, can require days to weeks for processing. And because it relies on a limited set of DNA "primer" sequences to amplify and ultimately sort sequences, or barcodes, that are distinctive to groups of organisms, it provides information on a limited number of species.

By contrast, the new study made use of another approach known as shotgun sequencing, in which all the DNA in a sample is randomly chopped up and billions of the sequences are "read." Computer programs spot overlapping bits and reassemble the digital fragments into larger sequences. The results can be compared with libraries of distinct genetic sequences for more than half a million species.

In the past, shotgun sequencing has been too costly and time intensive for routine processing of environmental samples, says David Duffy, a biologist at the University of Florida's Whitney Laboratory for Marine Bioscience. But recently developed tools, including so called nanopore sequencers, have lowered costs and enabled researchers to get results as quickly as in 2 days. Nanopore sequencers also read longer strands of DNA than other technologies, providing more complete genetic information, and they can be smaller than a cigarette packet.

Duffy first used shotgun sequencing of DNA in water and sand to monitor for disease in sea turtles in Florida. In the latest work his team cast a broader net, gathering airborne DNA from the same coastal Florida areas using pumps that sucked air across filters and swabbing surfaces such as windows. They then checked the samples for a broad range of species. "What we're trying to show is this is already cost effective, it's feasible, and it lets us answer a whole lot of new questions," Duffy says.

The researchers could pick out enough genetic details of the bobcat from the Florida sample to match it to a specific population in the northern part of the state. It also assembled long enough reads of human DNA to trace the ancestry of people passing through the state park to Europe, Asia, and Africa.

For a big city sample, they repeated the process in Dublin, where Duffy grew up. DNA from the city center reflected a larger variety of people genetically connected to different parts of the world, as well as 221 different human pathogens such as viruses—seven times more than in the Florida forest sample.

The work marks a step toward the dream of having an affordable way to get a comprehensive picture of all the DNA present in a place, says Simon Creer, a molecular ecologist at Bangor University. "If you can step back and look at everything you can have a better picture of what's happening in every tree of life," he says.

Jan Gogarten, a disease ecologist and eDNA expert at the Helmholtz Institute for One Health, cautions that shotgun sequencing isn't infallible. He notes that Duffy's team identified cowpox, a virus that is not known to be circulating in North America, in Florida seawater. "It's a tiny piece of DNA that's in a sample. What does that mean, and how much can you make of it?" he asks.

Others are alarmed about the ethical implications of scavenging any and all genetic information from the air. The new study only inferred the broad genetic heritage of passersby, but the method could potentially be used to identify individuals and examine their DNA. Natalie Ram of the University of Maryland Francis King Carey School of Law, an expert on genetic privacy, fears governments might use airborne eDNA to track people's movements and health information at an unprecedented level. "I don't know about you, but I'm not feeling so sanguine about massive surveillance efforts by the government these days, or ever, really," she says.

Creer says he recently saw a presentation by a scientist who works at a private company that was exploring the possibility of building a small shotgun-sequencing device that could be placed discreetly in a room, scanning the air for DNA from things such as dangerous viruses. "He couldn't talk about many things because much of his funding had come from defense contracts," Creer says.

However, Lara Urban, a statistical geneticist at the University of Zürich who works with eDNA, warns against letting hype outrun reality. It could be devilishly hard to pinpoint which person was the source of a tiny speck of DNA in the air. "It's a little bit of fear mongering, I fear," Urban says.

Duffy sees his team's work as highlighting some of the tool's capabilities, both good and bad. They thought of titling the paper "Towards a tricorder?"—a reference to a fictional handheld device depicted in the original *Star Trek* TV series that could be used on alien planets to instantly scan for life. "We're not claiming to be there," Duffy says. "We're saying we are a lot closer … and you can foresee it being a reality in the future." □

FEATURES

HOT ZONE

In the world's hottest forests, scientists are probing how plants cope with rising temperatures

ERIK STOKSTAD

Infrared panels push plants on the forest floor in Puerto Rico beyond their comfort zone.

PHOTO: MAXWELL FARRINGTON

It was an experiment ahead of its time. Working in his laboratory in Breslau, Prussia, botanist Julius von Sachs placed a large glass bell jar over a flowerpot and heated it from below with alcohol burners. The pioneering 19th century physiologist measured the rising temperature and watched for signs that the plant was in distress, repeating the process with various species. It was the first effort to determine the hottest temperature plants could survive. His finding—50°C—must have seemed unimaginably hot.

Sachs, now best known for his work on photosynthesis and plant nutrition, was simply curious. But today, as global temperatures climb and ecosystems are hit by heat waves, the question of how plants respond to extreme heat has taken on practical importance. Johan Uddling, a plant ecophysiologist at the University of Gothenburg, says a global experiment is underway, with "potentially devastating" consequences. "You could lose a lot of species that are on the edge," adds Lina Mercado, a plant ecophysiologist at the University of Exeter.

Perhaps the biggest challenge confronting plants will be in the tropics, where temperatures are already high—and are projected to rise as much as 4°C by the end of the century if more isn't done to curb climate change. "They're the hottest forests," says ecologist Kenneth Feeley of the University of Miami. "So, the question is: What happens when we go into unprecedented heat?"

In a lodge in Peru, researchers analyze leaf samples to measure how well plants photosynthesize.

PHOTO: KEN FEELEY

Scientists already see signs that plants are responding to the roughly 1.5°C of global warming so far. Some tree species in Mexico, for example, have shifted to higher elevations, where temperatures are cooler, a recent paper in *Science* reported (7 March, p. 1057). But across the Americas, tropical species aren't moving fast enough to keep pace with warming, and researchers are skeptical that they will be able to cope by migrating upslope or away from the equator. Nor is evolution likely to be the answer. Trees can take decades to start reproducing, making it unlikely they can evolve new genetic adaptations for heat tolerance fast enough to withstand rising temperatures.

That means that over the coming decades, the fate of tropical forests may hinge in part on how well individual plants can acclimate to warmer temperatures by adjusting aspects of their structure or physiology. The question is especially urgent for the tallest species. "The trees that make up the canopy right now, during their lifetimes, they're going to see changes in temperature that are quite significant," says Martijn Slot, a plant ecologist at the Smithsonian Tropical Research Institute.

To find out what plants are capable of, researchers are going far beyond Sachs's simple experiment. They are manipulating temperatures in plots of forest to push tropical species beyond their comfort zone. They are probing the traits of plants in a naturally superheated forest. And they are transplanting trees into warmer habitats to see how they fare.

The work has already revealed hints of resilience. "We know that plants can handle a little bit of warming," says Slot, who has measured how temperature affects photosynthesis in dozens of species. "They have the machinery to make adjustments that allows them to acclimate to gradual changes." But whether tropical forests can withstand the surge of warming likely in the coming decades remains a focus of feverish research.

THE STAKES EXTEND well beyond the plants themselves. In the world's megadiverse tropical forests, many species depend on the shade and microhabitats created by gargantuan trees that can grow 100 meters tall. The planet, too, depends on tropical forests; they hold as much as half of all the living carbon on land and continually suck in carbon dioxide (CO_2) emitted into the atmosphere by human activities. "If you start to see trees die, that's going to matter hugely," says Christopher Doughty, an ecologist at Northern Arizona University.

Plants rely on photosynthesis for energy, and one worry is that as temperatures ramp up, their photosynthetic machinery won't be able to cope. The enzymes and membranes that are key to capturing photons and creating sugar are generally tuned to the range of temperatures a species experiences. If the plants get much hotter or colder, photosynthesis becomes less efficient.

In tropical species, photosynthesis starts to decline around 31°C. By

Steam rises from the geothermally heated Boiling River in eastern Peru, where scientists are studying trees that manage to grow in the unusually hot environment.

PHOTO: RILEY FORTIER

47°C, the photosynthetic machinery begins to be permanently damaged, according to measurements by Slot and colleagues. Eventually, cellular membranes start to leak and enzyme activity grinds to a halt. Chlorophyll, the green photosynthetic pigment, breaks down and leaves turn brown and die. Upper canopy leaves in tropical forests currently exceed 47°C only 0.01% of the time, but that number could jump to 1.3% with only a few degrees of warming, according to a 2023 paper published in *Nature*.

Another problem is that as temperatures increase, cellular concentrations of highly reactive oxygen-bearing molecules begin to build up, which can cause oxidative damage and disrupt cellular function. One reason for the proliferation is that heat increases the speed of the chemical reactions that generate these troublesome molecules.

Plants do have ways to cope with heat stress, however (see graphic, p. 1128). Increasing the production of isoprene, a volatile organic chemical found in many plants, for example, can protect cells against reactive oxygen molecules. Plants can also tune their photosynthetic machinery to perform better at higher temperatures. One way is for leaf cells to make more saturated fatty acids to stabilize parts of the chloroplast, the organelle in photosynthetic cells where sugar is produced, that would otherwise become less efficient.

Plants can lower their own temperature as well. One strategy is to grow smaller, thinner leaves, which can dissipate heat more effectively than larger, thicker ones. Another involves opening tiny pores in their leaves, called stomata, which allows water to evaporate. That can be an effective strategy if the plant isn't deprived of water. But during droughts, plants often keep their stomata closed to prevent water loss. "When you combine the higher temperature with droughts, then things get really stressful," Slot says.

TO FIND OUT whether any of those strategies might help tropical forests survive a hotter future, scientists have turned to a spot that offers a preview: the Boiling River in eastern Peru, locally known as the Shanay-timpishka. Across a few square kilometers, superheated groundwater rising from deep faults pushes temperatures as much as 11°C above those of nearby areas. The air is so hot and humid near the river, which drains the scalding groundwater, that visitors are at risk of heat stroke.

Walking into the hottest part of the forest for the first time, ecophysiologist Alyssa Kullberg, now a postdoc at the Swiss Federal Institute of Technology, was struck by how oddly uniform it looked. Far fewer tree species grew there than in nearby, cooler parts of the forest, she thought—an impression that was later confirmed by a study in *Global Change Biology*. And the soil was strangely bare, lacking lush plants and wildlife. "You don't see a lot of animals, and you barely hear birds," she says. Some-

times dead frogs and snakes could be seen through the steam wafting off the river, floating by after presumably falling in and dying from the heat. "It's a weird forest, but very interesting."

Over the past several years, Feeley, Kullberg, and colleagues have studied six tree species that are able to live near the river. They didn't see obvious adjustments for heat. Compared with individuals of the same species growing elsewhere, the trees in the hottest locations don't have smaller, thinner leaves or other physical traits that might help them dissipate heat, such as a greater density of stomata, they reported in *New Phytologist* in 2023.

But the team did find signs of heat tolerance after clipping fingernail-size disks from leaves and probing their photosynthetic ability. At the tourist lodge where they were staying, Kullberg and colleagues immersed each disk in water ranging from 28°C to 68°C. Then they probed each sample with a handheld fluorometer, which shines light on the leaf and detects the photons emitted in response by its chlorophyll.

The team found that for three of the six species, photosynthesis in leaves from trees growing near the Boiling River could tolerate higher temperatures. For example, in *Cecropia membranacea*, a small tree known locally as the *yarumo*, they found an increase in a measure known as T50—the point at which photosynthesis is impaired by 50%. Leaf samples taken from a cooler part of the forest could tolerate only 47°C before reaching T50, but a leaf from the hottest area had a higher T50, 51°C. That's a good sign, because it means the species has the ability to respond to future warming. "Maybe it won't have a significant impact on their vulnerability," Kullberg says.

Feeley's team is continuing to study the forest, trying to figure out exactly how the toughest species tolerate the extreme environment and what their resilience bodes for the same species elsewhere as temperatures rise. "I think we're just really scratching the surface," he says.

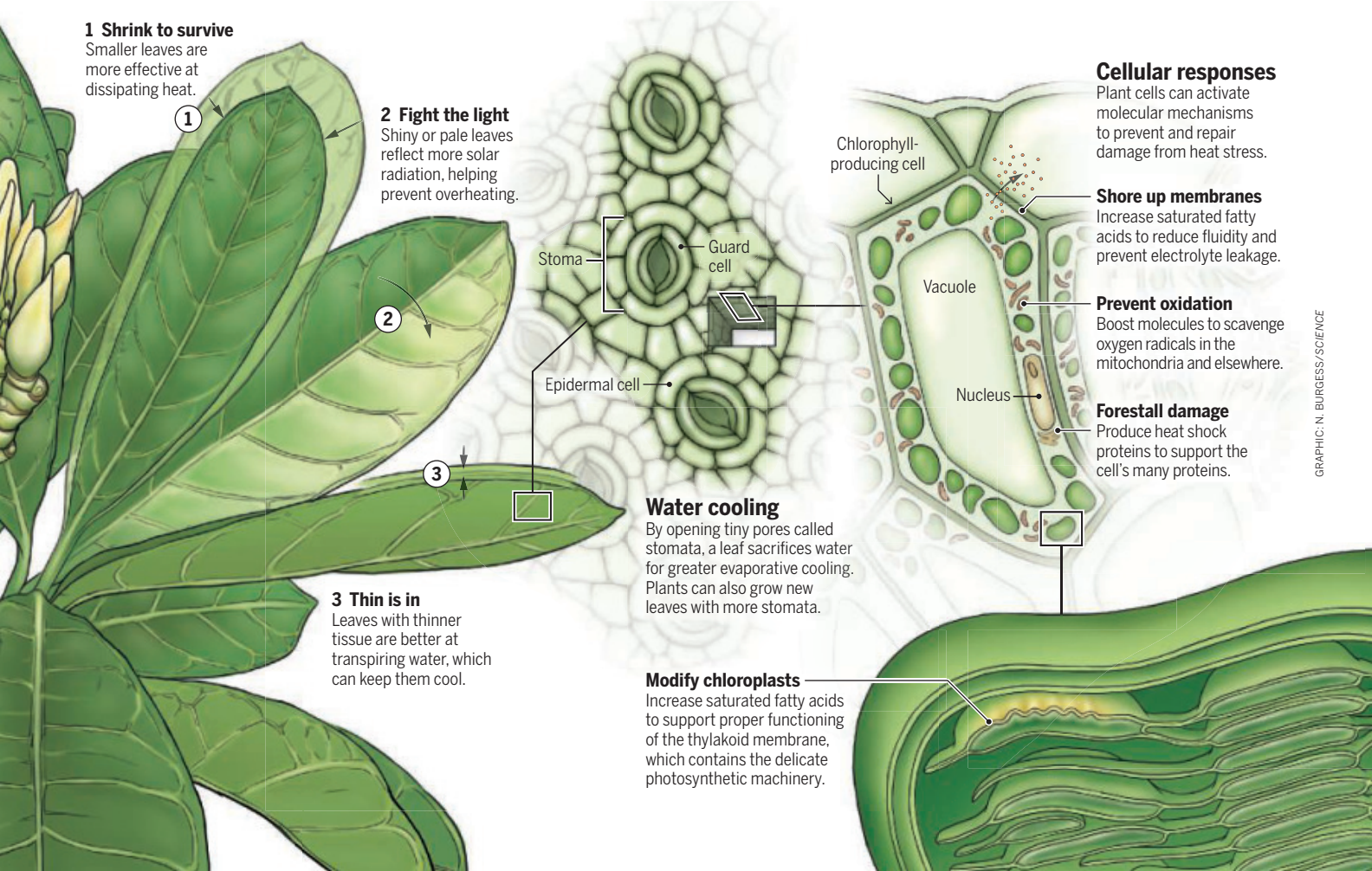
THE FOREST NEAR the Boiling River has had thousands of years to adjust to the geothermal heat. Tropical forests elsewhere are warming on a far faster timescale. What happens to species that are suddenly thrust into heat? Researchers in several countries are seeking answers by planting trees adapted to cooler, high-elevation conditions in lower elevation areas. Uddling and a large team did that in Rwanda, adding rain shields and irrigation pipes to ensure that moisture in the new sites matched the species' normal habitat.

When the trees were a few years old, the team measured their rate of photosynthesis in the field under a range of temperatures. The key data came from a portable gas analyzer, which when clamped onto a living leaf determines the amount of CO_2 flowing into its tissue. Two higher elevation species showed no increase in their optimal temperature for photosynthesis after growing in the lower plots, which are 5°C warmer on average, the researchers reported in *Global Change Biology* in 2021—a result that is contrary to the hints of photosynthetic acclimation seen along the Boiling River. "That's not that good news," says Myriam Mujawamariya, a plant ecophysiologist at the University of Rwanda who works on the project.

A later study on 16 species showed that in slower growing species, which

How plants stay cool

Changes in leaf structure (below) can help plants cope with heat, but they require new growth. Cells have other mechanisms (right) to help them react more quickly.



GRAPHIC: N. BURGESS/*SCIENCE*

typically make up the larger trees in older forests, photosynthesis was 30% lower for plants in the hotter plots compared with those at higher elevation. The finding, reported in *Tree Physiology* in 2023, doesn't bode well for the future of old-growth tropical forests, which store more carbon than any other forest type.

Elsewhere, scientists have been actively applying heat to natural forests. In the Luquillo Experimental Forest in Puerto Rico, Tana Wood, an ecologist with the U.S. Forest Service, and colleagues set up infrared panels to warm 4-meter-wide plots of soil and understory plants by 4°C. The experiment began in 2016; since then, the team has monitored the growth of seedlings to see whether the composition of species changes in the heated plots. They're also periodically collecting leaf samples to look for evidence that the plants are adjusting to the heat.

So far, they've seen mixed signs of acclimation. In one species the optimum temperature for photosynthesis increased somewhat, but in another it decreased, the team reported in *Frontiers in Forests and Global Change* in 2020. The researchers also extended the experiment to the canopy, which is too high to be warmed by panels near the ground, by adding small heaters to individual leaves near the treetops. They found a similar lack of photosynthetic acclimation there, as they reported in 2021 in *Plant, Cell & Environment*.

Kristine Crous, a plant ecophysiologist at Western Sydney University, has made similar measurements with leaf heaters in tropical forests in Australia. In 2024, she reported in *New Phytologist* that after 8 months of being heated 4°C above ambient temperatures, the leaves of four species had decreased their photosynthesis by as much as 35%. "Unfortunately, the tree species that are longest lived in old growth rainforest, they're the sensitive ones," she says.

More encouraging findings come from Biosphere 2, a massive set of greenhouses completed in 1991 outside Tucson, Arizona. Inside the largest space, tropical trees grow up to 27 meters high.

In a 2020 paper in *Nature Plants*, a team led by Marielle Smith, an ecosystem ecologist now at Bangor University, examined changing CO_2 levels inside the greenhouse under a range of temperatures. The researchers found that even when temperatures reached 43°C, photosynthesis was not impaired. They concluded that the biosphere's plants could withstand the heat because they had plenty of water, allowing them to cool their leaves by opening their stomata. It's unclear whether the plants would fare as well in natural environments, where heat often brings water stress.

Still, Joost van Haren of the University of Arizona, who directs rainforest research at Biosphere 2, sees "a positive story so far." Last summer, he and colleagues raised temperatures inside the greenhouse even more by turning off the cooling system. After clipping tiny fluorometers onto leaves and measuring photosynthesis at various air temperatures through the canopy, they found that species appear to be adjusting their photosynthetic capacity to withstand higher temperatures, van Haren says.

AS RESEARCHERS SIFT through evidence from various studies that have challenged the heat tolerance of tropical trees, a key question looms large: "How much is too much and when does the tree die?" Crous asks.

Several species appear to have the ability to bounce back, even after a severe heat wave. During van Haren's experiment in Biosphere 2, one part of the canopy reached 55°C, much hotter than during the conditions examined by Smith and colleagues. That was more than the leaves could handle, and they died. But they have since grown back. "We're seeing that there might be more resilience at Biosphere 2 than people have given tropical forests credit for," van Haren says.

Duration is probably an important factor. Wood suspects tropical species can cope with heat in the short term. The plants under the heating panels in Puerto Rico, she notes, don't die more often. But problems could mount over time if photosynthesis remains depressed. "It's like you run out of fuel to keep managing," she says. In her plots, she and her colleagues found that after a few years of warming, the understory plants weren't growing as tall.

Researchers also point out that heat doesn't strike in isolation. More intense storms and drought are also stressing the forests, which could compound the effects of warming. In Puerto Rico, for example, plant roots in the hotter plots recovered more slowly after a hurricane ravaged the forest. "That was a light bulb moment for me," Wood says, because it suggests warming makes plants less able to cope with additional disturbances.

The fate of tropical forests in a warmer future remains uncertain. One thing is not in doubt: The planet's lushest ecosystems are all now in Sachs's bell jar, with humanity, an unwitting experimenter, pumping in the heat. □

At the top of Biosphere 2's greenhouse, some leaves turned brown and died during an experimental heat wave. But the trees recovered.

PHOTO: AARON BUGAJ

PLANTS FACING THE HEAT

More from this special issue on p. 1146.

COMMENTARY

PERSPECTIVES

ECOLOGY

Global nitrogen budget revisited

Bottom sediments are important for nitrogen production in inland and coastal waters

Patrick L. Brezonik

Nitrogen is essential for the growth of all organisms. Although it is the most abundant element in Earth's atmosphere, most organisms cannot use atmospheric nitrogen gas (N_2) directly. It must be converted to usable forms, such as ammonium. Certain microorganisms perform this conversion using enzymes through the process of biological nitrogen fixation. This process is also important for Earth's carbon budget because faster plant growth leads to more carbon sequestration. On page 1205 of this issue, Fulweiler *et al.* (*1*) report that coastal and inland water habitats are important in the global nitrogen budget, despite their relatively small areal coverage compared with that of the oceans. This finding could stimulate further studies of biological nitrogen fixation in coastal and inland waters, particularly in habitat types that have been poorly studied.

The presence of nitrogen-fixing microorganisms in natural waters has been documented for more than 100 years. However, fixation rates—the amounts of nitrogen gas converted into usable forms per unit area or volume over time—were not reported for phytoplankton in surface waters until 1959 (*2*) and for sediments and deep waters (*3*, *4*) until about a decade later. Since then, many studies have been conducted in both marine and freshwater environments. Investigations of nitrogen fixation in inland waters were slow at first, with a minor peak in the late 1970s, and have increased more rapidly over the past two decades. Considerable attention has been given to nitrogen fixation in the open ocean (*5*) because of its large area of coverage (~70% of Earth's surface). In addition, the conventional view that primary

PHOTO: DAVID MCNEW/GETTY IMAGES

Harmful algal blooms, such as the one in Lake Elsinore, California, are often associated with nitrogen-fixing species of cyanobacteria.

production (that is, photosynthetic production of organic matter) is limited by the low availability of fixed nitrogen in the oceans has promoted the idea of nitrogen-fixing organisms in oceans having a competitive advantage over nonfixing organisms. Contributions of open-ocean biological nitrogen fixation to the global nitrogen budget (*6*, *7*) have been estimated in efforts to assess the impact of human activities on the global nitrogen cycle. By contrast, contributions of inland and coastal waters in global nitrogen budget estimates have received scant attention.

Fulweiler *et al.* compiled more than 4500 biological nitrogen fixation rate measurements in coastal and inland waters published between 1990 and 2022. Inland waters, such as lakes, rivers, and wetlands, were estimated to produce 24.1 teragrams (10^{12} g) of fixed nitrogen per year through biological fixation. This is 9.4% of open-ocean nitrogen contributions based on the average of 20 estimates and 8.7% of an earlier estimate of oceanic biological nitrogen fixation (*6*). Considering that the global area of inland waters is only ~3.8% of the area of oceans, inland waters thus behave as "hotspots" for biological nitrogen fixation, which agrees with previous studies on the importance of lakes in carbon cycling processes (*8*). Nitrogen produced from inland waters accounts for ~10% of the global annual fixed nitrogen by all natural processes—including lightning and biological nitrogen fixation in marine and terrestrial environments—and ~5% of combined anthropogenic and natural processes (*6*).

The study of Fulweiler *et al.* shows the dominance of nitrogen fixation in benthic regions (near or in the bottom sediments) compared with the water column. In coastal habitats, the amount of nitrogen fixed in the water column was negligible except in the continental shelf habitat. Fixed nitrogen produced in surficial sediments of all six coastal habitat types accounted for ~97% of the total coastal nitrogen fixed. Benthic nitrogen fixation in continental shelf regions alone accounted for 90% of the nitrogen fixed in coastal habitats. Likewise, benthic biological nitrogen fixation accounted for 96% of nitrogen fixed in inland waters.

By contrast, biological nitrogen fixation in water columns contributed only ~15% of the total nitrogen fixation in lakes. This finding is unexpected given the common association of nitrogen fixation with cyanobacteria (blue-green algae) blooms in lakes. Such blooms are still important for nitrogen inputs to individual lakes (*9*), but according to one study, they are insufficient to alleviate nitrogen deficits in phosphorus-rich lakes (*10*). Fulweiler *et al.* also conclude that biological nitrogen fixation in lake water columns is a minor component of global fixation.

Although the study of Fulweiler *et al.* may cover most fixation rate data, the number of sites in the dataset for a given habitat, such as lakes, is very small relative to their global abundance. Additional biological nitrogen fixation data exist for coastal and inland waters (*9*–*14*) that encompass the range of aquatic habitat types. Direct comparison of the Fulweiler *et al.* results with those from a previous data summary (*9*) is difficult because of the differences in data organization, but annual volumetric or areal nitrogen fixation rates are generally compatible.

Given the many sources of uncertainty involved in global-scale biological nitrogen fixation estimates (*1*, *6*), expansion of the inland and coastal water dataset to include pre-1990 and more recent data (post-2022) would be useful. Large sources of uncertainty from the huge spatial and temporal variabilities in biological nitrogen fixation activity within the same habitat type further support the importance of large data collection efforts. In addition, aquatic biological nitrogen fixation rates are measured by at least five different procedures, and each involves different assumptions and limitations (*1*, *10*). For example, the acetylene reduction method (*15*), which accounts for 64% of the data in Fulweiler *et al.*, is based on the finding that nitrogen-fixing enzymes reduce acetylene to easily measured ethylene. However, the appropriate factor to convert the amount of ethylene produced to the amount of nitrogen fixed has some uncertainty. The authors minimized uncertainties in their global estimates by using median values that are less sensitive to outliers than mean values. However, the relative paucity of data remains a concern in nearly all global budgeting efforts. □

REFERENCES AND NOTES

1. R. W. Fulweiler *et al.*, *Science* **388**, 1205 (2025).
2. R. Dugdale, V. Dugdale, J. Neess, J. Goering, *Science* **130**, 859 (1959).
3. P. L. Brezonik, C. L. Harper, *Science* **164**, 1277 (1969).
4. M. A. Keirn, P. L. Brezonik, *Limnol. Oceanogr.* **16**, 720 (1971).
5. J. P. Zehr, D. G. Capone, *Marine Nitrogen Fixation* (Springer, 2021).
6. D. Fowler *et al.*, *Phil. Trans. R. Soc. B* **368**, 20130164 (2013).
7. P. M. Vitousek *et al.*, *Ecol. Appl.* **7**, 737 (1997).
8. J. J. Cole, N. F. Caraco, G. W. Kling, T. K. Kratz, *Science* **265**, 1568 (1994).
9. R. W. Howarth, R. Marino, J. Lane, J. J. Cole, *Limnol. Oceanogr.* **33**, 669 (1988).
10. N. M. Hayes *et al.*, *Freshw. Biol.* **64**, 269 (2019).
11. J. J. Goering, J. C. Neess, *Limnol. Oceanogr.* **9**, 530 (1964).
12. V. A. Billaud, *J. Fish. Res. Board Can.* **25**, 2101 (1968).
13. A. Patoine, M. D. Graham, P. R. Leavitt, *Limnol. Oceanogr.* **51**, 1665 (2006).
14. H. Halm *et al.*, *Environ. Microbiol.* **11**, 2190 (2009).
15. W. D. P. Stewart, G. P. Fitzgerald, R. H. Burris, *Proc. Natl. Acad. Sci. U.S.A.* **58**, 2071 (1967).

10.1126/science.ady3520

Department of Civil, Environmental, and Geo- Engineering, University of Minnesota, Minneapolis, MN, USA. Email: brezonik@umn.edu

INFECTIOUS DISEASE

Rethinking jaundice

A buildup of unconjugated bilirubin may be a protective response to malaria

Joachim Kloehn *and* **Dominique Soldati-Favre**

Malaria continues to be among the major devastating infectious diseases in the world (*1*). It is caused by a protozoan parasite of the genus *Plasmodium* that is transmitted by mosquitoes in the *Anopheles* genus. Unfortunately, resistance to drugs and insecticides, as well as climate change, are threatening progress in malaria prevention, control, and treatment over the past decades. Certain human genetic traits confer some protection against malaria and are prevalent in malaria-endemic regions. On page 1181 of this issue, Figueiredo *et al.* (*2*) report a genetic trait that may provide protection from the disease—the decreased expression of a bilirubin conjugation enzyme. As a result, unconjugated bilirubin concentrations increase, which has antimalarial effects.

Natural lines of defense against malaria have evolved in malaria-endemic regions over millions of years. These include genetic traits that confer partial resistance or protect from severe malaria. For example, mutations in the gene encoding glucose-6-phosphate dehydrogenase results in an enzyme deficiency that increases oxidative stress in red blood cells. This has a negative effect on the parasite (the *Plasmodium* life cycle includes an asexual phase in host red blood cells). A decrease in hemoglobin production by red blood cells caused by mutations in genes encoding the globin subunits of hemoglobin (a condition known as thalassemia) makes these cells less hospitable to *Plasmodium* and increases their clearance from the host (*3*). Mutations in the gene that encodes glycophorin cause Dantu, a condition characterized by high surface tension in red blood cells that makes it difficult for malaria parasites to invade and multiply. And red blood cells that lack Duffy antigens, the receptors for one species of *Plasmodium*, hinder the parasite's ability to invade these cells (*3*). These genetic traits are often associated with substantial host trade-offs, such as the development of sickle cell disease or an increased risk of anemia, and are therefore not predominant in the global population. However, the protective effect against malaria has made these traits prevalent in malaria-endemic regions, where coevolution of the parasite and its host has shaped their genomes (*4*).

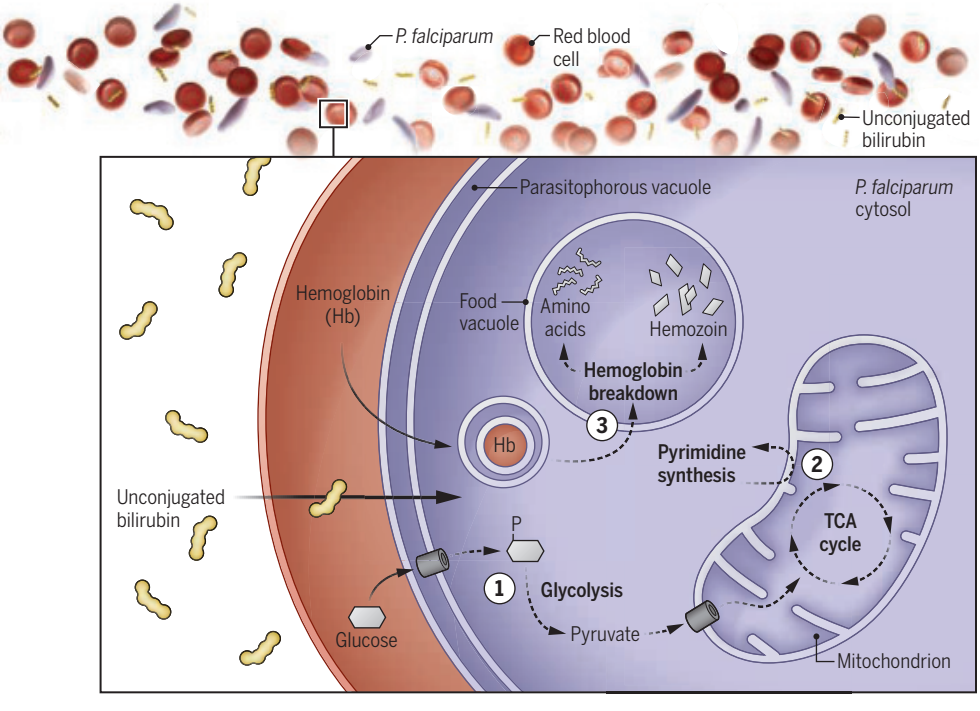
Increased amounts of unconjugated bilirubin are a characteristic of Gilbert's syndrome, a benign liver condition caused by a polymorphism in the promoter region of the gene encoding the enzyme UDP glucuronosyltransferase family 1 member A1 (UGT1A1), resulting in its reduced transcription (*5*). UGT1A1 catalyzes the addition of glucuronic acid to bilirubin, a yellowish pigment that is a by-product of red blood cell breakdown. This modification occurs in hepatocytes and transforms insoluble unconjugated bilirubin into a water-soluble conjugated form that is excreted through bile and urine. Hence, repressing UGT1A1 causes unconjugated bilirubin to accumulate, which manifests as mild jaundice. Jaundice, the yellowing of the skin and eyes, is a common feature in malaria and has been associated with hemolysis or liver dysfunction and considered a marker of disease severity (*6*). However, Figueiredo *et al.* suggest, through experiments in cultured infected red blood cells and rodent models and by examining human data, that unconjugated bilirubin serves a host-beneficial function during *Plasmodium* infection. Increasing unconjugated bilirubin amounts, through direct administration to infected red blood cells, impaired parasite development and halted disease progression. Conversely, mice genetically engineered to lack bilirubin production were hypersensitive to *Plasmodium* infection, a vulnerability that was reversed by administration of unconjugated bilirubin. Additionally, asymptomatic humans with *Plasmodium falciparum* parasitemia had increased amounts of unconjugated bilirubin relative to parasite burden compared with symptomatic infected individuals. Consistent with a protective effect of unconjugated bilirubin, the UGT1A1-repressing polymorphism is overrepresented in populations of African origin, where malaria is endemic (*7*).

Increased unconjugated bilirubin has been frequently observed in people suffering from malaria. However, Figueiredo *et al.* present a paradigm-shifting view of bilirubin-associated jaundice in the context of *Plasmodium* infection, challenging the conventional notion that the accumulation of unconjugated bilirubin in plasma is solely a consequence of liver dysfunction. Instead, the authors propose that this metabolic alteration constitutes a protective, adaptive host response that limits malaria severity.

Host metabolism has long been thought to play a major role in the defense against pathogens (*8*). This includes the withdrawal of nutrients or other factors to starve pathogens. Changes in host metabolism are also a line of defense against malaria. The protein ferroportin hinders *Plasmodium* development by exporting iron from red blood cells and limiting the parasite's access to this critical factor. Indeed, a mutation in ferroportin that prevents its degradation is positively selected for in human populations located in malaria-endemic regions (*9*). Similarly, labile heme that is released from red blood cells that are lysed during parasite infection restricts hepatic glucose

The antimalarial effects of unconjugated bilirubin

In one phase of its infectious life cycle, the malaria parasite *Plasmodium falciparum* resides in its parasitophorous vacuole within a red blood cell. Unconjugated bilirubin in the host plasma disrupts several metabolic processes in the parasite, including (**1**) glucose catabolism in the cytosol and (**2**) the tricarboxylic acid (TCA) cycle and pyrimidine synthesis in the mitochondrion. Furthermore, (**3**) hemoglobin in the pathogen's food vacuole is also affected, including its breakdown to obtain amino acids and detoxification into hemozoin.



GRAPHIC: N. BURGESS/*SCIENCE*

production by repressing glucose-6-phosphatase catalytic subunit 1. Inhibiting the parasite's access to sugar diverts its development from asexual replication to sexual differentiation (*10*). This switch reduces virulence, as the asexual stage drives rapid replication and disease, whereas the sexual stage is required for transmission to mosquitoes.

Host metabolic defenses also include the generation of metabolites or metabolic by-products with antimicrobial activity. These include nitric oxide, reactive oxygen species, lactate, succinate, and itaconate (*8*). Itaconate is produced by macrophages and inhibits multiple bacterial enzymes (*11*). Although these metabolites can have systemic immune modulatory functions (*12*, *13*), they typically act at the site of infection to exert an antimicrobial effect while avoiding damage to the host. Figueiredo *et al.* now add unconjugated bilirubin to the list of metabolites with antimicrobial properties. Red blood cells are broken down by macrophages in the spleen, liver, and lymph nodes. In contrast to the localized production of other antimicrobial metabolites, unconjugated bilirubin, a by-product of red blood cell destruction, is released from macrophages and can enter the circulation and preferentially accumulate in *Plasmodium*-infected red blood cells. The authors show that its protective effect does not extend to other infectious diseases, such as polybacterial sepsis and influenza A infection, but it is unknown whether unconjugated bilirubin exhibits activity against other hemolytic disease–causing pathogens, such as *Babesia* and *Bartonella*.

The findings of Figueiredo *et al.* suggest a multifactorial antimalarial mode of action for unconjugated bilirubin. The authors show that unconjugated bilirubin impairs the parasite's energy metabolism, pyrimidine biosynthesis, and hemozoin crystallization (which converts toxic heme to harmless hemozoin) and also reduces mitochondrial membrane potential and volume (see the figure). However, it is unclear whether these processes are targeted directly by unconjugated bilirubin or are downstream effects of the actual target (or targets), which has yet to be identified. Also, it is unclear whether the antimalarial properties of unconjugated bilirubin are limited to direct effects on the parasite or have broader effects, including immune modulatory functions. That unconjugated bilirubin is not just a by-product of *Plasmodium*-induced hemolysis but acts as a host-derived antimalarial agent reshapes the notion of jaundice in individuals with malaria. The study of Figueiredo *et al.* paves the way for new views on disease resistance, therapeutic development, and evolutionary biology. □

REFERENCES AND NOTES

1. World Health Organization (WHO), "World malaria report 2024: Addressing inequity in the global malaria response" (WHO, 2024).
2. A. Figueiredo *et al.*, *Science* **388**, eadq6741 (2025).
3. S. N. Kariuki, T. N. Williams, *Hum. Genet.* **139**, 801 (2020).
4. P. W. Hedrick, *Heredity* **107**, 283 (2011).
5. L. M. Grant, T. W. Faust, V. Thoguluva Chandrasekar, S. John, in *StatPearls* (Treasure Island, 2025).
6. A. C. Anand, P. Puri, *J. Gastroenterol. Hepatol.* **20**, 1322 (2005).
7. E. Beutler, T. Gelbart, A. Demina, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8170 (1998).
8. S. Kreimendahl, L. Pernas, *Trends Cell Biol.* **34**, 496 (2024).
9. D. L. Zhang *et al.*, *Science* **359**, 1520 (2018).
10. S. Ramos *et al.*, *Cell Metab.* **34**, 1183 (2022).
11. M. Priya *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **122**, e2423114122 (2025).
12. V. Lampropoulou *et al.*, *Cell Metab.* **24**, 158 (2016).
13. A. Llibre *et al.*, *Immunity* **58**, 535 (2025).

ACKNOWLEDGMENTS

J.K. and D.S.-F. acknowledge support from the University of Geneva and CARIGEST SA.

10.1126/science.ady7161

Department of Microbiology and Molecular Medicine, University of Geneva, Centre Médical Universitaire, Geneva, Switzerland. Email: dominique.soldati-favre@unige.ch

BIOENGINEERING

Synthesizing vascular trees at speed

A computational algorithm can render a complex artificial vascular structure in minutes

Yan Yan Shery Huang[1,2] *and* **Lining Arnold Ju**[3,4]

Engineering a biologically functional structure from living cells can solve global health challenges by supplying abundant artificial tissues and organs for drug screening, disease modeling, and transplantation (*1*–*4*). A vascular system is an essential component for the survival of thick tissues. This tree-like network delivers nutrients to cells, regulates blood pressure, and provides resilient blood flow through a hierarchical structure (*5*). Designing a micro- to millimeter-scale vascular architecture for a centimeter-scale artificial tissue requires extensive computational resources (*6*). Consequently, engineered vascular networks often have a simple lattice structure that is easy to fabricate but exhibits nonuniform pressure gradient and flow distribution. On page 1198 of this issue, Sexton *et al.* (*7*) report a computational approach that can design a complex vascular tree network at more than 230 times the speed of traditional methods. This enables the design of anatomically realistic vascular networks for organ-scale biofabrication.

Biofabrication—an automated production of biologically functional structures—has primarily focused on developing tools that organize biological building blocks (such as cells and proteins) in a three-dimensional (3D) space (*3*, *4*). As advanced bioprinting methods allowed precise cell placements with good viability, the goal has shifted toward generating human-scale, physiologically relevant tissues. Reproducing the dynamics of blood flow in an artificial tissue is important for its survival. However, designing a complex vascular tree requires substantial computational labor and cost. Although previous lattice-based perfusion designs (*8*) are sufficient for tissues at low cell densities (<10^6 cells/ml), this structure is not sufficient to meet the metabolic demand at clinically relevant cell densities (>10^8 cells/ml). Computational methods such as constrained constructive optimization—building and optimizing a hierarchical network step by step on the basis of a predefined constraint—can design vascular structures, but it often takes several days to build a complex model.

To enable rapid modeling of organ-scale synthetic vasculature, Sexton *et al.* incorporate four different algorithms. Partial binding optimization technique—solving a complex problem by breaking it down into smaller subproblems—reduces computational complexity by temporarily storing intermediate structures at each vessel addition. Partial implicit volume representation slices 3D volume into smaller portions and enables efficient handling of nonconvex geometries, such as vascular branches and heart valves, without integrating a computationally expensive mesh (a discrete representation of modeling space). Hierarchical collision detection uses a tree structure to graphically simulate and screen potential intersections between two vessels. Multi-objective optimization makes simultaneous decisions to generate a closed-loop vascular circuit.

This multifidelity simulation approach enables automatic selection of appropriate computational methods on the basis of a prescribed vessel diameter and flow regime. For example, large vessels (>150 μm in diameter) that tend to exhibit nonlaminar blood flow are simulated using 3D finite element models. By contrast, small vessels (<100 μm in diameter) with laminar flow are analyzed using more efficient one-dimensional or zero-dimensional approximations. The approach of

Sexton *et al.* generated complex branching architectures with up to a million terminal vessels in minutes. The produced vascular design had nonconvex geometries, which reflect natural vascular systems such as a biventricular heart, with vessel diameters ranging between 1.2 mm and 5 μm. The authors fabricated physical constructs of selected vascular designs using various 3D printing methods.

Democratizing virtual representation of vasculature networks could potentially transform biofabrication by allowing evaluation of perfusion efficiency prior to production rather than through a resource-intensive trial-and-error method. However, physiological vasculature serves various critical biological functions beyond nutrient distribution, and local properties of vessel wall could substantially influence blood flow dynamics in a vascular tree. For example, depending on the physiological state of a vessel, endothelial cells at the inner surface of a vessel may respond differently to shear stress from blood flow. Angiogenic sprouting—growth of capillary vessels out of an existing one— could gradually change functions of the vascular network from the initial ones (*9*). Nutrient requirements for local tissues evolve with tissue growth, remodeling, and regeneration because the spatial distribution of cell density and cell type changes. Integrating fluid mechanics with computational models that use an agent to represent time-dependent cellular behaviors could account for these variables but are only suited for simple tissues. To design a large-scale vascular tree with biological complexity, combining experiments and computational approaches will be important, including data-driven and machine learning models that learn from existing patterns of experimental observations to propose an optimal vascular design meeting tissue-specific requirements (*10*).

The ability to simulate complex vasculature offers exciting possibilities for modeling pathological vascular conditions, such as microvascular thrombosis from infections (*11*), and predicting thrombosis risk in mechanical circulatory support devices (such as ventricular assist devices and extracorporeal membrane oxygenation) (*12*). Vascular tree models can have selective occlusions (blocked vessels) or altered rheological (flow) properties. Thus, downstream effects of these perturbations to tissue perfusion and potential therapeutic targets could be identified. Such models may also help explain the improved collateral artery performance observed during cardiac regeneration (*13*), providing insights into vascular remodeling mechanisms. □

REFERENCES AND NOTES

1. J. F. Murphy *et al.*, *Biodes. Manuf.* **7**, 825 (2024).
2. "Imagining Futures of 3D Bioprinting," *WHO Global Health Foresight Series* (World Health Organization, 2024).
3. M. S. Huang, F. Christakopoulos, J. G. Roth, S. C. Heilshorn, *Nat. Rev. Bioeng.* **3**, 126 (2025).
4. R. Levato *et al.*, *Nat. Rev. Methods Primers* **3**, 47 (2023).
5. D. Ribezzi *et al.*, *Adv. Mater.* **35**, 2301673 (2023).
6. W. Schreiner *et al.*, *Med. Eng. Phys.* **28**, 416 (2006).
7. Z. A. Sexton *et al.*, *Science* **388**, 1198 (2025).
8. J. S. Miller *et al.*, *Nat. Mater.* **11**, 768 (2012).
9. S. S. S. Aye, Z. Fang, M. C. L. Wu, K. S. Lim, L. A. Ju, *Biomater. Sci.* **13**, 1131 (2025).
10. P. Du, X. Zhu, J.-X. Wang, *Phys. Fluids* **34**, 081906 (2022).
11. M. C. L. Wu *et al.*, *Nature* 10.1038/s41586-025-09076-x (2025).
12. A. Sun, A. Nasser, N. A. Yap, R. Gao, L. A. Ju, *Curr. Opin. Biomed. Eng.* **33**, 100576 (2025).
13. S. Anbazhakan *et al.*, *Nat. Cardiovasc. Res.* **1**, 775 (2022).

ACKNOWLEDGMENTS

Y.Y.S.H. acknowledges support from the European Research Council (grant no. ERC-StG 758865) and the UK Research and Innovations Biotechnology and Biological Sciences Research Council (grant no. BB/W014564/1). L.A.J. is a National Heart Foundation Future Leader Fellow Level 2 (105863) and a Snow Medical Research Foundation Fellow (2022SF176). Y.Y.S.H. used Microsoft Copilot to prepare this Perspective (see supplementary materials).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ady6122

10.1126/science.ady6122

[1]Department of Engineering, University of Cambridge, Cambridge, UK. [2]The Nanoscience Centre, University of Cambridge, Cambridge, UK. [3]School of Biomedical Engineering, The University of Sydney, Darlington, NSW, Australia. [4]Heart Research Institute, Newtown, NSW, Australia. Email: yysh2@cam.ac.uk

MICROBIOLOGY

Bacteria poison and eat their neighbors

Bacteria leverage a secretion system to kill and scavenge nutrients from nearby competitors

Carey D. Nadell[1,2] *and* **Christopher J. Marx**[3]

Bacteria have evolved numerous ways to kill nearby cells, often by secreting diffusible toxins, secondary metabolite antibiotics, or toxic metabolic by-products (*1*). Many bacteria also possess mechanisms for direct, contact-dependent killing. For example, the type VI secretion system (T6SS) extends a molecular needle into cells to deposit toxins (*2*, *3*). Bacteria bearing such poison needles carry antidotes to their own toxins to avoid killing themselves or other cells of their type. The T6SS provides an advantage by removing competitor cells that might otherwise consume the same resources (*4*, *5*). On page 1214 of this issue, Stubbusch *et al.* (*6*) report that T6SS-bearing cells can grow and replicate using the released contents of neighboring cells that they kill. This suggests that the T6SS—and potentially other mechanisms of bacterial antagonism—could be used not only to create space by eliminating competitors but also as a predatory nutrient-scavenging mechanism.

Using a combination of custom microfluidic cell culture, high-resolution imaging, stable isotope tracing, and computational modeling, Stubbusch *et al.* demonstrate that when placed under starvation conditions, T6SS-bearing bacteria (genus *Vibrio*) can grow using substrates released from lysed competitor cells. Such predatory behavior had been proposed previously for T6SS but was never demonstrated experimentally. Lending additional weight to the notion that T6SSs can be linked to a predatory lifestyle, the authors analyzed publicly available genomes of the *Vibrio* clade and documented that the presence of T6SS machinery was associated with the loss of genes related to primary and secondary carbohydrate metabolism and amino acid metabolism. This form of predation may be particularly advantageous in contexts in which nutrients released by the lysed cells do not immediately diffuse away. Consistent with this idea, a larger-scale genomic analysis by Stubbusch *et al.* revealed that T6SS is more prevalent in bacteria found in environments where there is a high density of spatially constrained bacteria growing on surfaces, such as in soil and on roots.

Notably, the analysis of *Vibrio* genomes by Stubbusch *et al.* showed that T6SS is broadly distributed in the genus but is often either present or absent in closely related lineages. This suggests that predation can be quite labile on evolutionary timescales. This pattern of T6SS occurrence across the *Vibrio* genus raises the question of whether a reduction in metabolic self-sufficiency in favor of eating the contents of neighboring cells occurs before, in concert with, or after the acquisition of T6SS, which is known to mobilize by means of horizontal gene transfer (*7*–*9*). More broadly, the evolutionary instability of T6SS machinery in the *Vibrio* clade also suggests that antagonism is not always consistently maintained by selection over long periods. This idea is supported

One underlying factor at the microbial community scale is that the attack-and-consume strategy is likely to be inherently self-limiting.

by studies of bacteria within animal gut microbiomes that have documented T6SS loss, both within a host over its lifetime and over longer evolutionary time scales (*8*, *10*, *11*).

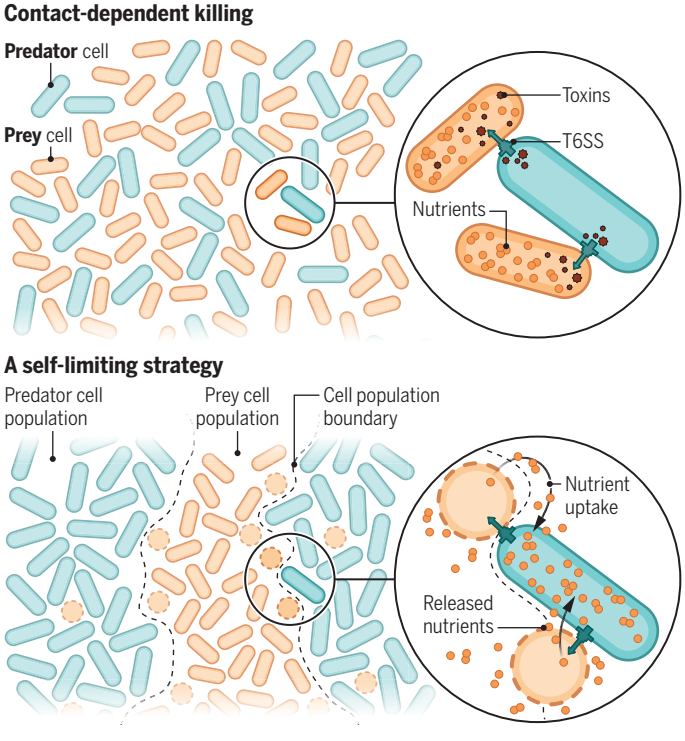
The metagenomic analysis by Stubbusch *et al.* indicates that T6SSs are widespread but remain at moderate to low frequency across environments. One underlying factor at the microbial community scale is that the attack-and-consume strategy is likely to be inherently self-limiting. T6SS-mediated antagonism becomes less effective as a resource-gathering mechanism as the relative abundance of potential prey cells decreases over time. Furthermore, although spatial structure and diffusion limitation are important for contact-dependent killing and nutrient gathering (*12*), spatial structure can also generate barriers to prey access. T6SS activity inherently tends to cause the segregation of groups of attacking and target cells, restricting the nutrient-scavenging potential of T6SS to the boundaries of the resulting cell groups (*4*, *5*).

Perhaps unexpectedly, there are striking parallels between T6SS-driven cell killing and consumption and mutualistic metabolic exchange, in which both partners gain an advantage from cross-fed metabolites (*13*, *14*). In both scenarios, nutrients are released into the extracellular space, where they are accessible to all nearby cells. Thus, for both neighbor-killing and mutualism, spatial structure is critical to success through privileged access to the released nutrients (*12*, *15*). Furthermore, because both scenarios entail releasing nutrients into the environment—either by killing other cells or directly secreting valuable compounds—they can each be exploited. Genotypes or even other species that do not pay the cost of generating the released nutrient pool can still consume the resources made available by other cells and prevent them from holding a majority in the population (*15*). This may also help to explain why T6SS-mediated secretion as a nutrient-scavenging strategy can be self-limiting.

The study of Stubbusch *et al.* also emphasizes an important general principle: Cells that are unable to access growth-limiting substrates through their native metabolism—or in the case of auxotrophs, cells unable to produce essential cellular building blocks—do not necessarily need a cross-feeding partner. Instead, if a cell possesses a T6SS or another mode of lysing other cells, it can simply kill and eat its neighbors. An area for future work will be to explore the similarities and differences between the conditions that promote neighbor-consuming predation versus mutualistic metabolic exchange as bacterial lifestyles. The ubiquity, yet low frequency, of T6SS-based predation across metagenomes suggests that this lifestyle plays an important role in an array of diverse environments. □

A neighbor-eating strategy

Bacteria that possess a type IV secretion system (T6SS) inject toxins directly into neighboring cells that may be competing for local resources. Toxins break down the cell wall of susceptible cells, causing the release of nutrients that the attacking bacteria can consume for growth and proliferation. However, this can cause spatial segregation of the bacterial cell population, restricting attacks to the boundaries of predator cell groups and prey cell groups. This might limit the number of cells that can leverage the T6SS system for predation.



GRAPHIC: A. FISHER/*SCIENCE*

REFERENCES AND NOTES

1. E. T. Granato, T. A. Meiller-Legrand, K. R. Foster, *Curr. Biol.* **29**, R521 (2019).
2. A. B. Russell, S. B. Peterson, J. D. Mougous, *Nat. Rev. Microbiol.* **12**, 137 (2014).
3. M. Basler, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20150021 (2015).
4. L. McNally *et al.*, *Nat. Commun.* **8**, 14371 (2017).
5. W. P. J. Smith *et al.*, *PLOS Biol.* **18**, e3000720 (2020).
6. A. K. M. Stubbush *et al.*, *Science* **388**, 1214 (2025).
7. M. J. Coyne, K. G. Roelofs, L. E. Comstock, *BMC Genomics* **17**, 58 (2016).
8. B. D. Ross *et al.*, *Nature* **575**, 224 (2019).
9. B. Jana, K. Keppel, C. M. Fridman, E. Bosis, D. Salomon, *mSystems* **7**, e0072322 (2022).
10. M. I. Steele, N. A. Moran, *mSystems* **6**, e00063-21 (2021).
11. S. Robitaille *et al.*, *Nat. Ecol. Evol.* **7**, 2092 (2023).
12. C. D. Nadell, K. Drescher, K. R. Foster, *Nat. Rev. Microbiol.* **14**, 589 (2016).
13. E. J. Culp, A. L. Goodman, *Cell Host Microbe* **31**, 485 (2023).
14. C. R. Evans, C. P. Kempes, A. Price-Whelan, L. E. P. Dietrich, *Trends Microbiol.* **28**, 732 (2020).
15. W. R. Harcombe, J. M. Chacón, E. M. Adamowicz, L. M. Chubiz, C. J. Marx, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 12000 (2018).

ACKNOWLEDGMENTS

We thank B. Ross for feedback on the manuscript. C.D.N. acknowledges support from National Institute of General Medical Sciences grant R35GM151158, and C.J.M. acknowledges support from National Science Foundation grant DBI-2320667.

10.1126/science.ady7008

[1]Department of Biological Sciences, Dartmouth, Hanover, NH, USA. [2]Department of Microbiology and Immunology, Geisel School of Medicine at Dartmouth, Hanover, NH, USA. [3]Department of Biological Sciences, University of Idaho, Moscow, ID, USA.
Email: carey.d.nadell@dartmouth.edu; cmarx@uidaho.edu

BIOMEDICAL TECHNOLOGY

A liability framework for high-risk neural devices

A no-fault compensation scheme may help balance innovation and patient protection

A. Rotenberg[1], M. Gunning[1,2], R. Magistro Nadler[1,3], Z. H. T. Kiss[4], J. Illes[1]

High-risk medical devices are integral to modern health care, but unexpected adverse events from performance failures or malfunctions may leave patients irreversibly harmed. Consumers in the US who incur injuries from products can ordinarily seek relief under the product liability laws of their state. However, a 2008 US Supreme Court decision against patient complainants restricted this legal pathway, raising concerns of patient abandonment. With the emergence of implantable brain-computer interfaces (iBCIs) exposing vulnerable patients to a new suite of unknowns and potential physical and neurocognitive impacts, policy-makers must revisit the commitment to justice at the forefront of innovation. Focusing on the case example of intracranial neurotechnology, we describe a no-fault compensation scheme that may offer relief to injured patients in the US. Its potential advantages may also hold relevance in countries where civil litigation remains the only pathway to patient recourse.

Our proposal for a no-fault compensation scheme rests on arguments that reasonable assurances of safety and effectiveness should be accompanied by relief for unreasonable harms and that investments in the best interests of patients are contributions to—not detractions from—medical innovation. Although recent developments in neurotechnology motivate our focus, the underlying regulatory and compensatory justice challenges implicate all high-risk medical devices.

ASYMMETRIC LIABILITY FRAMEWORK

In 1974, the US Congress passed legislation that empowered the US Food and Drug Administration (FDA) to regulate medical devices. Included in the new set of laws was a clause that barred the states from undercutting the FDA by creating different regulatory requirements. Soon afterward, medical device manufacturers began defending themselves in product liability lawsuits by asserting that the clause preempted, or effectively granted them immunity from, all state-law claims. This interpretation of the law resonated with the US Supreme Court. In *Medtronic* v. *Riegel* (2008), the Court held that FDA requirements imposed through the rigorous premarket approval (PMA) process for evaluating device safety and efficacy broadly preempted state-law claims that "challeng[ed] the safety or effectiveness" of a medical device [(*1*), p. 312]. The decision was later applied to devices used under investigational device exemption (IDE) designations; such devices have only been approved for experimental use in the relevant clinical context (*2*). These factors focused the scope of preemption on high-risk devices that are subject to IDE and PMA requirements and used when patient vulnerability is greatest.

The decision of the Supreme Court indicated that state-law claims could overcome the preemption defense if they were grounded in violations of FDA policy. However, this pathway is narrow and fraught with challenges (*3*, *4*). Federal courts maintain stringent submission requirements, affecting cases that begin in state courts. Inconsistent judicial interpretations often preclude lawsuits. As a result, most injuries resulting from high-risk devices leave patients without meaningful legal redress, even when the devices cause considerable harm. The results of post-*Riegel* preemption decisions in medical device lawsuits current to 10 April 2025 are as follows: Of 706 decisions, 65% (n = 458) favored manufacturers, 23% (n = 162) yielded mixed outcomes, and 12% (n = 86) favored complainants (*5*) (see table S1). These preempted claims do not proceed to discovery or trial, and their merits are not weighed by a jury. Thus, for more than half of patients purportedly injured by high-risk and investigational devices, *Riegel* and related decisions eliminate the possibility of seeking legal recourse.

The exclusive pertinence of *Riegel* to high-risk and investigational devices raises an important question of accountability. If it is true, as Justice Ginsburg noted, that "FDA regulation of a device cannot anticipate and protect against all safety risks" [(*1*), p. 337], then who should shoulder responsibility for unanticipated harms? The context provided by innovative intracranial neurotechnology illustrates the need to address this challenge.

INTRACRANIAL NEUROTECHNOLOGY

Intracranial neurotechnology encompasses high-risk medical devices, such as deep brain stimulation (DBS) and responsive neurostimulation, that are placed within the skull or brain to monitor or manipulate neural activity. The positioning and purpose of these devices give rise to profound risks such as seizures, stroke, hemorrhage leading to neurologic impairment, and death (*6*). These outcomes are rare, and most treated patients achieve a substantial degree of relief (*7*). Yet the severe risk profile of intracranial neurotechnology limits its use to a treatment of last resort. Patients only become candidates after exhausting less invasive alternatives (*8*). Emergent harms are therefore absorbed by vulnerable individuals pursuing their last lines of hope. For example, in *Millman* v. *Medtronic*, a patient with a debilitating motor condition sued to recoup the costs of his fourth DBS implantation procedure in an 8-year period (*9*). Allegedly, a representative of the manufacturer had advised repeated replacements that led to a deterioration of the patient's health. The case was dismissed in its entirety under *Riegel*, and the veracity of the allegations was never scrutinized.

Patients may understand and accept the potential risks of treatment, but unless informed consent discussions delve into the complex intersections of federal and state law, they may not be aware of the full implications of those risks. This raises ethical and legal concerns for researchers and physicians (*9*). It also imperils participants in clinical trials for IDE devices, which have not yet met the PMA standard of safety and effectiveness. Nonetheless, federal courts have held that the restrictions on patient-initiated lawsuits imposed by *Riegel* apply in the IDE context (*2*). This has direct im-

plications for participants with quadriplegia or amyotrophic lateral sclerosis, for example, who have consented to participate in the 18 ongoing iBCI trials registered at clinicaltrials.gov (*10*). These include trials sponsored by companies such as Neuralink and Synchron that have received substantial attention through coverage in the media.

If participants in these clinical trials are injured, the costs may be theirs alone to bear. Recognition of this asymmetry, prompted by, for example, a high-profile lawsuit, could provide a compelling disincentive for potential participants, with clear repercussions for human trials. Thus, to promote fairness, safeguard the future of clinical innovation, and recognize the societal debt owed to patients and research participants, systems must be created to provide support when worst-case scenarios arise.

PATIENT COMPENSATION PROGRAM

The federal government can balance societal interests by opening legal pathways to recourse when high-risk devices manufactured in compliance with FDA policies cause harm. A no-fault compensation program may be the suitable and responsive approach compared to alternatives (see the table). A no-fault compensation program would offer an extrajudicial alternative to dispute resolution wherein complainants may be remunerated without raising accusations of legal wrongdoing. This provides an accessible and relatively nonadversarial pathway to recourse. New Zealand and the Nordic countries, for example, maintain public no-fault compensation programs for iatrogenic harms that broadly absorb the costs of medical hazards. To fill the void left by *Riegel* in the US, the narrower National Vaccine Injury Compensation Program (NVICP) may serve as a model (*11*).

NVICP is a specialized program under the Department of Health and Human Services that adjudicates claims of vaccine-related injuries. It was created by Congress in 1986 to balance societal interests by curtailing the extraordinary judgments awarded against vaccine manufacturers in product lability lawsuits, thereby preserving the national vaccine supply. Key features distinguish the NVICP from traditional courts. For example, the judge-jury-defense triad is replaced by a special master who reviews petitions, assesses eligibility, and calculates damages. The program further expedites proceedings by granting a presumption of causality to harms listed in a vaccine injury table. Claims of these harms are not subject to the usual causation analysis seen in tort law; it is enough to show, for instance, that anaphylaxis occurred within 4 hours of receiving the seasonal influenza vaccine. Initially, petitioners could elect to reject decisions and pursue their cases through the court system, but the Supreme Court closed this option in 2011 based on their interpretation of the law.

Through a small excise tax collected from manufacturers on each dose, the NVICP has provided support to many petitioners. Between 2006 and 2022, more than 5 billion vaccinations were administered in the US, and 12,593 petitions were filed with the NVICP. Of these, 9124 were deemed compensable; average awards exceeded $500,000 to cover a wide range of expenses (*12*). These data suggest that injured patients may be reliably compensated under the NVICP framework. They also attest to the feasibility of a distributed funding structure. The wholesale adoption of this model for device-related harms would improve upon the status quo; modest adjustments could improve upon it even further.

Whereas the NVICP addresses diverse types of injuries, a medical device compensation program for iBCIs ought to focus first on severe or long-term impairments. Given high associated personal health care costs and their variance from treatment expectations, patient abandonment by way of preemption may have the steepest consequences in these cases. A device injury table for life-altering harms would expedite these claims. Such a table may grant a presumption of causation, for example, to neurological deficits that emerge if an iBCI system migrates within a patient's brain over time. Unless the government provides another theory about cause, a petitioner who substantiates claims to this effect would automatically be entitled to compensation. Should causality be challenged, a panel of experts may need to be consulted.

As in *Riegel*, the decision of the Supreme Court that restricted post-NVICP civil litigation was supported by interpretations of text (*13*). To avoid this risk, policy-makers must ensure that any law involving a medical device compensation program will still preserve existing pathways for civil litigation. This is important because no-fault programs are not concerned with legal wrongdoing and therefore eschew protected discovery phases that are standard in civil proceedings and represent vital mechanisms of accountability that complement regulatory aims (*14*). A program for medical devices may preserve the role of civil litigation by implementing a dual-entrance structure. One entryway upstream of civil litigation would allow expedient compensation decisions, whereas another entryway downstream of civil litigation could capture preempted cases that have been rejected by the courts.

Potential pathways to recourse

Compared with alternatives, a no-fault compensation program may be more suitable and responsive

Approach	Benefits	Drawbacks
Instituting a no-fault medical device compensation program	• Efficient relief distribution • Equitable for patients • Distributes collective risks	• Requires act of Congress • Raises costs of market entry • May shield manufacturers from liability
Narrowing the 2008 *Riegel* decision to permit liability claims	• Increases manufacturer accountability to patients • Complements regulatory mechanisms • May yield large jury judgments	• Requires decision of Supreme Court or act of Congress • Inequitable for resource-limited patients • May disincentivize critical industries
Relying on third-party insurance models	• Efficient relief distribution • Distributes collective risks	• Inequitable for resource-limited patients • May only provide narrow coverage

The excise tax funding structure used by the NVICP depends on high production volume and may not be scalable to low-volume medical devices. A potential funding scheme must therefore be flexible to account for different commercial realities. One solution may be to adopt a modified form of the insurance program for nuclear energy facilities laid out in the Price-Anderson Act of 1957. Under Price-Anderson, operators of nuclear facilities must carry private liability insurance as the first source of compensation for victims of accidental harm. If damages exceed the primary insurance limit, compensation is then provided through a secondary financial protection pool, which is funded by mandatory contributions from all operators. In the context of medical devices, the layers of compensation would be reversed such that the primary layer—a no-fault program—would be funded by mandatory contributions. The federal government could subsidize contributions for smaller manufacturers as well as those developing devices that align with priority therapeutic goals, thereby promoting innovation and aligning commercial incentives with national research strategies. Contributions would begin during the IDE application process to provide security for clinical trial participants. A secondary layer of private insurance would complement risk-spreading with risk-based pricing, promoting safety through the same mechanisms as automobile insurance programs. Private insurance would support industry accountability, expand compensation capacity, and relieve taxpayers of their pres-

ent obligation to pay the full costs of device-driven health care crises (*11*). Taken together, a carefully structured no-fault injury program for iBCIs and other high-risk medical devices could advance patient justice, preserve incentives for innovation, and broadly promote public interests.

RESPONSIVE REGULATORY TOOLS

Excitement for the future of neurotechnology sparked by demonstrations of its potential benefits has encouraged thousands of individuals to apply for participation in clinical trials for experimental implants (*15*). They and future patients deserve a safety net when results fall injuriously short of promises. Yet crucial means of legal recourse for patients and participants injured by approved and investigational medical devices have been restricted by the US Supreme Court. The implementation of a no-fault compensation program for medical device injuries, enacted by the federal government and supported by experts from relevant medical specialties, will benefit vulnerable patients as well as research participants by addressing both device-related harms and the resources needed for continued device use once trials end. Anticipatory changes that align responsive regulatory tools to the needs of the modern commercial landscape will bring balance to the fore of medical innovation and justice where it breaks down. □

REFERENCES AND NOTES

1. *Riegel* v. *Medtronic Inc.*, 552 U.S. 312 (2008).
2. *Day* v. *Howmedica Osteonics Corp.*, 2015 U.S. Dist. LEXIS 189268 (D. Colo. 2015).
3. D. A. Simon, *Wis. L. Rev.* **2024**, 1079 (2023).
4. D. A. Simon, A. S. Kesselheim, in *Digital Health Care Outside of Traditional Clinical Settings: Ethical, Legal, and Regulatory Challenges and Opportunities*, I. G. Cohen *et al.*, Eds. (Cambridge Univ. Press, 2024), pp. 109.
5. J. M. Beck, "(New) medical device PMA preemption scorecard," Drug & Device Law Blog, 21 August 2008.
6. M. Ward *et al.*, *Br. J. Neurosurg.* **35**, 625 (2021).
7. J. K. Krauss *et al.*, *Nat. Rev. Neurol.* **17**, 75 (2021).
8. J. Illes, *AJOB Neurosci.* **3**, 65 (2012).
9. *Millman et al.* v. *Medtronic Inc.*, No. 1:14-cv-01465 (D.N.J. 2015).
10. National Library of Medicine, Search results; https://clinicaltrials.gov/search?cond=brain%20computer%20interface&aggFilters=status:not%20rec&viewType=Card&term=Implant [accessed 2025].
11. D. Chang, *Hastings Law J.* **65**, 283 (2013).
12. Health Resources and Services Administration, "National Vaccine Injury Compensation Program data report" (Department of Health and Human Services, 2024); https://www.hrsa.gov/vaccine-compensation/data.
13. S. T. Berns, *Univ. Miami Int. Comp. Law Rev.* **29**, 62 (2021).
14. H. Rodgers, *Fla. State Univ. Law Rev.* **45**, 861 (2018).
15. R. Winkler, "Elon Musk's Neuralink gets FDA green light for second patient, as first describes his emotional journey," *Wall Street Journal*, 20 May 2024.

ACKNOWLEDGMENTS

We acknowledge the helpful comments of editors and reviewers. We acknowledge support from the Canadian Institutes of Health Research and ERA-NET NEURON grant ERN-174185 (J.I., Z.H.T.K.), the North Family Foundation (J.I.), and the University of British Columbia Distinguished University Scholars Program (J.I.). M.G. and R.M.N. contributed equally to this work.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adt8110

10.1126/science.adt8110

[1]Neuroethics Canada, Division of Neurology, Department of Medicine, University of British Columbia, Vancouver, BC, Canada. [2]Interdisciplinary Studies Graduate Program, Faculty of Graduate and Postdoctoral Studies, University of British Columbia, Vancouver, BC, Canada. [3]Allard School of Law, University of British Columbia, Vancouver, BC, Canada. [4]Hotchkiss Brain Institute, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada.
Email: jilles@mail.ubc.ca; zkiss@ucalgary.ca

BIOMEDICAL TECHNOLOGY

New case law and liability risks for manufacturers of medical AI

Recent case law can shape how innovation unfolds

Sara Gerke[1,2] *and* **David A. Simon**[3]

Medical devices increasingly incorporate artificial intelligence and machine learning (AI/ML-enabled medical devices). Manufacturers of such devices must consider various factors that can shape the trajectory of research and technology development and deployment, including liability risks. Although there is currently no direct case law on liability when using medical AI (*1*), a recent products liability suit in Louisiana—*Dickson* v. *Dexcom Inc.*, 2024, 2:24-CV-00121—is particularly relevant for AI/ML manufacturers. Although *Dickson* does not involve an AI/ML-enabled medical device, many tools similar to those in the case are currently being developed by using AI/ML. This decision thus addresses an important question relevant to AI/ML-enabled device manufacturers: Can they block a broad spectrum of personal injury and products liability claims when the US Food and Drug Administration (FDA) authorizes their devices through the newer and increasingly popular "De Novo" pathway? We discuss this case below and analyze its potential implications for AI/ML-enabled device manufacturers.

Although this article focuses on US law, it also contains valuable lessons for manufacturers outside the United States that may want to bring their AI/ML-enabled medical products to the US market. It also provides lessons for regulators to better understand liability risks for AI/ML manufacturers and reevaluate whether their current framework needs adjustment to promote AI/ML-enabled medical device innovation and adequate remedies for harmed patients.

PATHWAYS TO MARKET

Medical devices are classified by the FDA according to risk. Under the Federal Food, Drug, and Cosmetic Act (FDCA) [21 USC § 360c(a)(1)(C)], a high-risk (Class III) device cannot reach the market without "Premarket Approval" (PMA), which requires "provid[ing] reasonable assurance of its safety and effectiveness," typically by successfully completing clinical trials. The vast majority of devices that reach the market, however, are classified as moderate- to low-risk devices (Classes II and I, respectively). For these devices, manufacturers can often seek "510(k) Premarket Notification" or "clearance," by which the FDCA [21 U.S.C. § 360c(i)(1)] requires them to demonstrate that the device is "substantially equivalent" to another legally marketed device in the United States ("predicate device"). Most of the Class I and some Class II devices are 510(k)-exempt, the FDA having determined that 510(k) is not required to ensure reasonable safety and effectiveness. Class I devices are subject to "general controls" (such as device reporting, labeling, and registration), whereas Class II devices are also subject to "special controls" (such as performance standards) if they are available (*2*). Special controls under the FDCA [21 U.S.C. § 360c (a)(1)(B)] are "necessary to provide adequate assurance of safety and effectiveness" and can include postmarket surveillance, performance standards, and patient registries.

Almost all nonexempt Class I or Class II devices reach the market through the 510(k) process. Unlike Class III PMA devices, the FDA typically does not require applicants to prove reasonable safety and effectiveness by submitting clinical trial data (*2*). More recently, an increasing number of devices have reached the market through a different

pathway: De Novo Classification Request. This pathway is meant for devices that do not fit well within the 510(k) pathway because they lack a predicate device but present lower risks than those devices that typically need a PMA (*3*). Because the De Novo pathway is a middle course for new moderate- to low-risk devices, manufacturers may provide clinical data (*4*). The FDA usually requires special controls if the De Novo device is classified into Class II (*5*).

Despite the attractiveness of the De Novo pathway, it still remains underused compared with the 510(k) pathway. The number of De Novo–authorized AI/ML-enabled medical devices (32) has so far been about 3.27% of the number of 510(k)-cleared AI/ML-enabled medical devices (980) (as of 25 March 2025) (*6*, *7*). This number seems low but still shows a trend of bringing AI/ML-enabled medical devices to market through the De Novo pathway. For example, data show that between October 2017 and August 2023, the overall number of all De Novo–authorized medical devices (26) compared with the number of all 510(k)-cleared devices (2929) was 0.89% (*8*). During the same period, the number of De Novo–authorized AI/ML-enabled medical devices (19) was about 2.88% of the number of 510(k)-cleared AI/ML-enabled medical devices (659) (*6*). Manufacturers of AI/ML-enabled devices successfully used the De Novo pathway about three times more often than did most device manufacturers.

LIABILITY FOR INJURIES

When a medical device causes injury to a patient, manufacturers may be liable for damages under state laws prohibiting devices that contain manufacturing, design, or marketing defects. When a product defect injures a patient, the patient can bring a lawsuit under state law. But these lawsuits face several obstacles. One key potential roadblock is a defense that federal law [21 U.S.C. § 360k(a)] "expressly preempts" or blocks tort claims: Manufacturers can argue that a state law requires manufacturers to do something "different from, or in addition to" what federal law requires concerning the device's safety and effectiveness.

However, an "express preemption" defense is not a silver bullet against litigation. It applies to manufacturers of PMA devices and not to manufacturers of 510(k) devices (*5*, *10*). According to the US Supreme Court, the primary reason for this is that PMA devices undergo a more rigorous premarket review than do 510(k) devices (*3*). Whereas PMA devices must provide valid scientific evidence, 510(k) devices focus on demonstrating substantial "equivalence [between two devices], not [the] safety" of the manufacturer's device [*Medtronic, Inc.* v. *Lohr*, 518 U.S. 470, at 492–94 (1996)]. This means that device manufacturers may face a choice between cheaper and faster market access [510(k)] and greater liability protection (PMA).

Until recently, only a few courts had addressed the issue of preemption for De Novo devices—for example, *Tuttle* v. *Dexcom, Inc.*, no. 1:20-CV-4744-LMM (2021)—and none had specifically opined on the express preemption question when the FDA had codified risks and mitigation measures for a device in federal regulations. *Dickson* is the first case to hold that federal law can expressly preempt personal injury lawsuits under state law when the FDA codifies regulations applicable to a Class II device authorized through the De Novo pathway. *Dickson* is important because it previews the key issue: How attractive this pathway becomes in the future may depend on whether manufacturers can successfully assert express preemption as a defense to personal injury lawsuits under state law.

The *Dickson* case

In *Dickson*, the plaintiff alleged she was injured when the defendant's continuous glucose monitor—authorized as a Class II medical device through the De Novo pathway—failed to alert her to dangerously low blood glucose levels. Because she was not warned about her levels, she had a hypoglycemic event while driving and crashed her car. She sued on various theories, including design defect, breach of express warranty, and failure to warn under the Louisiana Products Liability Act. In its defense, the manufacturer argued that the case should be dismissed because, among other reasons, the FDCA expressly preempted the plaintiff's claims under 21 U.S.C. § 360k(a).

After issuing two opinions, the court ultimately agreed. In its first opinion (15 July 2024), the court held that the FDA triggered a preemption analysis by codifying into federal regulations "device-specific label and design requirements" (special controls), entitled "Integrated Continuous Glucose Monitoring System" (21 C.F.R. § 862.1355). The court then dismissed the plaintiff's design defect and breach of express warranty claims, holding that they are expressly preempted under 21 U.S.C. § 360k(a) because they establish requirements "in addition to" the ones established in the federal regulations.

However, the court left open a possible alternative path for a claim. It noted that for PMA devices, "failure to warn" claims are not preempted when they allege that the manufacturer failed to update labeling under federal "Changes Being Effected" (CBE) regulations, which require the manufacturers to submit label changes on the basis of "newly acquired [risk] information" [21 C.F.R. §§ 814.39(d)(2), 314.70(c)(6)(iii)]. For the purposes of argument only, the court then assumed that a similar mechanism—perhaps through a 510(k) submission—might exist for devices initially authorized through the De Novo pathway. On the basis of this assumption, it allowed the plaintiff to refile her complaint with additional evidence that the manufacturer had "newly acquired [risk] information" that would require a label update.

In its subsequent opinion (25 September 2024), the court reaffirmed its previous preemption ruling and considered only the failure to warn claim and the issues of whether the CBE regulation applies, and if so, whether newly acquired information required a label update. The court first stated that the CBE regulation applies only to PMA devices (not to De Novo devices). The court then assumed, for purposes of argument only, that even if "a mechanism like the CBE regulation" applied to De Novo devices, the plaintiff had not demonstrated sufficient evidence to support her claim of newly acquired risk information requiring the manufacturer to update the label.

IMPLICATIONS FOR MANUFACTURERS

Dickson highlights the uncertainty about liability risks for manufacturers of digital health devices and underscores why they must pay close attention to preemption cases when strategizing device development. The more courts follow *Dickson*, the more we expect manufacturers to market AI/ML-enabled moderate-risk devices as De Novo Class II devices to secure express preemption protection from state law liability (*3*) rather than be creative by claiming in a 510(k) that there is a predicate device that is substantially equivalent to their device. For example, several AI/ML-enabled medical devices that have 510(k) clearance are based on predicates that are based on other predicates, and so on, even down to a predicate that was launched before 28 May 1976, at a time when the safety and effectiveness of devices had not yet been assessed (*2*). Although the De Novo pathway is more time-intensive and costly than a 510(k), those resources would be well spent and possibly amortized if it offers preemption.

Additionally, a *Dickson*-like posture by courts is likely to encourage manufacturers to invest more heavily in moderate-risk (rather than high-risk) devices—for example, AI/ML-enabled surgical software, that provide substantial returns because the De Novo pathway offers the same liability protection as, but is less rigorous and less costly than, a PMA. *Dickson* also reinforces that a single decision does not foreclose creative lawyers from making new arguments to nullify or avoid preemption. For example, lawyers may argue that the CBE regulation—that allows lawsuits against manufacturers of PMA devices when they fail to update the device's label on the basis of newly acquired risk information—also applies (or should apply) to De Novo devices. Without such application, the argument goes, the law would command a perverse result: PMA devices would require more evidence and review than those for De Novo devices but would receive less protection from the express

preemption doctrine than that of De Novo devices. This would be inconsistent with Supreme Court precedents [*Medtronic, Inc.* v. *Lohr*, 518 U.S. 470 (1996) and *Riegel* v. *Medtronic, Inc.*, 552 U.S. 312 (2008)] that have conditioned preemption protection on how thoroughly the FDA reviewed the device in question.

Relatedly, technological innovation in De Novo devices may face a considerable pull-back if a post-*Dickson* court rejects the preemption argument and a manufacturer faces substantial litigation costs. In the face of uncertainty, manufacturers are likely to make decisions according to their appetite for risk, which may depend on factors such as the device type, profitability, first-mover advantage, market segment, and insurability. *Dickson* also has limited reach because the nature of the judicial process is hierarchical and diverse. *Dickson* was decided by a federal district court and is not binding on any other court or parties. For example, the *Dickson* court concluded that the FDA's codification of "special controls" into federal regulations could preempt state law claims. But it is unclear whether other courts will see this differently.

However, other courts are likely to be sympathetic to the *Dickson* court's analytical orientation. The *Dickson* court thought, on the basis of *Riegel*, that the FDA's codification of special controls constituted "requirements" applicable to the device. Even if the *Dickson* court was wrong about the importance of this FDA action, future courts are likely to be sensitive to how rigorous the De Novo process is. Depending on whether and to what extent courts follow *Dickson*, manufacturers could avoid liability by emphasizing the complexity of regulation applicable to their devices.

The uncertainty raised by the *Dickson* decision should be weighed by device manufacturers against other factors. For example, a De Novo pathway allows manufacturers to set the standards for a new type of device, from which it can build in successive 510(k) filings (*11*). Because the manufacturer controls many aspects of the De Novo submission, the manufacturer can influence the ability of competitors to enter the market with a similar device (*11*). Because many AI/ML-enabled medical devices run on data and software that are undisclosed but also protectable as trade secrets, manufacturers may attempt to ossify standards that are inaccessible without the trade secret owner's permission. Even if only moderately successful, this strategy could prevent or delay competitor market entry. Manufacturers could demand licensing revenue or refuse to license trade secrets altogether, which could deter firms from entering the market. If the market becomes less competitive as a result, the overall pace of innovation might slow, which is precisely the opposite of what the preemption doctrine is supposed to encourage.

On the other hand, the first De Novo device still opens the door for competitors with similar AI/ML-enabled devices to enter the market through the less time-consuming and less expensive 510(k), allowing market competition. The possibility of successfully relying on express preemption in cases of a De Novo device could also be interpreted as a reward for first-to-market innovations and for establishing a new device type, which may also lead to an uptake in using the De Novo pathway in the long run and ultimately to safer devices (*2*).

Last, even if manufacturers can successfully assert a preemption defense, they still must litigate the defense. Although the argument will decrease litigation costs by disposing of the case early, it will not eliminate them entirely. Thus, until the issue of preemption for De Novo devices is resolved more completely, manufacturers should allocate funds to litigate personal injury claims.

REDUCING UNCERTAINTY

This considerable regulatory and legal uncertainty facing manufacturers may decrease investment in AI/ML-enabled medical devices. To reduce uncertainty, AI/ML manufacturers and other stakeholders could push for legislative or regulatory changes. Congress could enact a new provision or section in 21 U.S.C. § 360k(a) clarifying to what type of medical devices FDCA's express preemption provision can apply. For example, Congress could clearly state that preemption is applicable to PMA but not De Novo devices, or the reverse. Or it could create a default no-preemption rule for De Novo devices but allow firms to apply for exemptions from liability. There is some precedent for the exemption approach. Existing law allows for states to apply to the FDA for an exemption from preemption when stricter requirements are "required by compelling local conditions" [21 U.S.C. § 360k(b)]. To reduce uncertainty, Congress could specify the general conditions under which the FDA could impose preemption, perhaps based on its assessment of a particular device. The FDA could then develop regulations to implement these provisions, and particular manufacturers could apply for the exemption.

Alternatively, the FDA could reduce uncertainty by issuing new regulations that create a clear path for personal injury claims. For example, it could implement a CBE regulation for De Novo devices similar to the one that exists for PMA devices. Under existing precedent, this would allow plaintiffs to bring the type of claims alluded to by the *Dickson* court. Manufacturers could still be liable under a failure to warn claim even if the plaintiff's other claims are expressly preempted. This could be particularly important for AI/ML-enabled medical devices because they may deploy algorithms that are not "locked" in the future. The continuous learning of such algorithms would preferably improve their performance, but it is also possible that they may worsen over time for some patient groups (*12*). Newly acquired risk information of the AI/ML-enabled medical device would require manufacturers to update their device's label.

This latter strategy is not foolproof, especially in light of the Supreme Court's decision in *Loper Bright Enterprises* v. *Raimondo*, 603 U.S. 369 (2024), which overturned the precedent requiring courts to defer to reasonable agency interpretations of statutes. After *Loper Bright*, the FDA's decision to enact a CBE regulation could be overturned by a court, eliminating the cause of action the regulations created. Uncertainty about how courts will decide this issue may also reduce investor appetite for AI/ML technologies, slowing the pace of innovation (*13*, *14*). □

REFERENCES AND NOTES

1. M. Duffourc, S. Gerke, *Stanf. Technol. Law Rev.* **27**, 1 (2024).
2. S. Gerke, *Yale J. Health Pol. Law Ethics* **20**, 432 (2021).
3. D. A. Simon, C. Shachar, I. G. Cohen, Northwest. Univ. Law. *Rev. Online* **119**, 137 (2024).
4. J. L. Johnston, S. S. Dhruva, J. S. Ross, V. K. Rathi, *JAMA Intern. Med.* **180**, 1701 (2020).
5. W. N. Price, S. Gerke, I. G. Cohen, *Research Handbook on Health, AI and the Law*, B. Solaiman, I. G. Cohen, Eds. (Edward Elgar, 2024), pp. 150.
6. FDA, "Artificial intelligence and machine learning (AI/ML)-enabled medical devices" (2025); https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices.
7. S. Gerke, *Emory Law J.* **74**, (2025).
8. M. Aboy, C. Crespo, A. Stern, *npj. Digit. Med.* **7**, 1 (2024).
9. A. Kumar *et al.*, *Health Aff.* **42**, 1559 (2023).
10. D. A. Simon, A. S. Kesselheim, in *Digital Health Care Outside of Traditional Clinical Settings: Ethical, Legal, and Regulatory Challenges and Opportunities*, I. G. Cohen, D. B. Kramer, J. Adler-Milstein, C. Shachar, Eds. (Cambridge Univ. Press, 2024), pp. 109–122.
11. J. Sherkow, M. Aboy, *Nat. Biotechnol.* **38**, 2018 (2020).
12. B. Babic *et al.*, *Science* **366**, 1202 (2019).
13. D. A. Simon, M. J. Young, *JAMA* **332**, 1325 (2024).
14. S. Gerke, D. A. Simon, *Health Aff.* 10.1377/forefront.20250101.818091 (2025).

ACKNOWLEDGMENTS

The views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union (EU) or the Health and Digital Executive Agency. Neither the EU nor the granting authority can be held responsible for them. The authors acknowledge support by EU grant 101057321. The funder had no role in the preparation, review, or approval of the manuscript or decision to submit the manuscript for publication. S.G. received grants from the EU (101057099), the National Institute of Biomedical Imaging and Bioengineering (NIBIB) and the National Institutes of Health Office of the Director (NIH OD) (1R21EB035474-01), the University of Illinois Urbana-Champaign's (UIUC) Cancer Center at Illinois, and the UIUC's EU Center through support from the US Department of Education's Title VI NRC Program. S.G. is also a Research Fellow at the University of Copenhagen, Faculty of Law, supported by a Novo Nordisk Foundation Grant for a scientifically independent International Collaborative Bioscience Innovation & Law Programme (Inter-CeBIL programme; NNF23SA0087056). D.A.S. owns Simon Law Group, a private law practice.

10.1126/science.adu4932

[1]College of Law, University of Illinois Urbana-Champaign, Urbana-Champaign, IL, USA. [2]European Union Center, University of Illinois Urbana-Champaign, Urbana-Champaign, IL, USA. [3]School of Law, Northeastern University, Boston, MA, USA. Email: gerke@illinois.edu

SCIENCE AND SOCIETY

Dark matters

A new film celebrates the subterranean

Nathaniel J. Dominy

Underland
Robert Petit, director
Sandbox Films, 2025.
79 minutes

British nature writer Robert Macfarlane's acclaimed writings pivot around themes of language and landscape and their enduring legacies in myth, literature, and memory—or "matters of the human heart," as he puts it (*1*). In 2019, Macfarlane published *Underland: A Deep Time Journey*, which took a darker turn. A bestselling work of transcendent beauty, *Underland* is a tale of personal adventure that explores and complicates humanity's relationship with subterranean spaces, the ephemeral nature of time and place, and, perhaps above all, the fragility of all we are and all we create.

The film adaptation of *Underland* premiered on 5 June 2025 at the Tribeca Film Festival. Poetic and mesmerizing, the film echoes the book in many essential ways, but it follows different storylines. It will resonate with readers of *Science* for focusing on the work of three intrepid researchers: Fátima Tec Pool, a Mexican archaeologist; Bradley Garrett, an American geographer; and Mariangela Lisanti, an American theoretical physicist. *Underland* joins these scientists underground to shine a light on their daring quests for knowledge. It may seem paradoxical—that darkness is a medium for vision, a means of insight—but it is effective. After all, the verb "to understand" hints at the importance of passing beneath something to comprehend it.

The 20th-century scholar Joseph Campbell argued that stories everywhere recount adventurous rites of passage, following a narrative structure known as the hero's journey (*2*). His template was complex, involving 17 events organized into three main stages, but he summarized the basic pattern as follows: "A hero ventures forth from the world of common day into a region of supernatural wonder: fabulous forces are there encountered and a decisive victory is won: the hero comes back from this mysterious adventure with the power to bestow boons on his fellow man." Germane here is Campbell's characterization of the middle stage as an "inmost cave" where heroes battle with "powers of darkness in order to return with the gift of knowledge."

Underland ventures into several inmost caves, both real and metaphorical. The first is a vast limestone cenote in Mexico. It invites comparison with Xibalba, the Mayan underworld, and the scene of many adventures in Mayan mythology, especially those of the Hero Twins Hunahpu and Xbalanque. Another inmost cave is a 200-mile labyrinth of storm drains beneath Las Vegas, a refuge and lair for hundreds of marginalized people. Built to accommodate sudden downpours, the tunnels bear witness to other excesses above, along with dashed hopes and tattered dreams. A third inmost cave is an ultraclean laboratory within SNOLAB, a research complex in the operational Vale Creighton nickel mine near Sudbury, Canada. Physicists there work 2 km underground to detect dark matter and understand its physical properties.

The film's scenes from SNOLAB are especially poignant for ending in failure; the experiment documented by the filmmakers did not detect a signature of dark matter. Lisanti's disappointment here is familiar, representing a different kind of inmost cave—one of self-doubt. But our hero perseveres. Impelled by a fear of never knowing, she asks, "Shall we start again?" although her words are more credo than question. This scene is important. Failure and resolve are essential elements of science; they are success in disguise. But such moments of emotional reckoning are seldom acknowledged by scientists or represented in popular culture, which is a great pity because they are pivotal to every hero's journey. "Asking why," muses Lisanti, "is the act of refusing to accept the state of not knowing."

For the final act, the filmmakers visit a sacred space, drawing attention to Mayan rites of passage. The rock walls they capture feature a gallery of extraordinary hand stencils, both positive and negative; and falling on these images are the researchers' own shadows, as if darkness is speaking with the past.

As I watched, I could not help but wonder whether this moment of revelation and shadow play was intended as a subtle reference to Plato's allegory of the cave, a dialogue that equates caves and shadows with ignorance and sunlight with knowledge. *Underland* certainly underscores the value of light—there are vivid scenes of lasergrammetry and uranium fluorescence—but the central message is to invert and complicate Plato's allegory, to call attention to the profound insights that lurk beneath the skin of the earth, where darkness thickens and sounds stir. □

REFERENCES AND NOTES

1. R. Macfarlane, *Underland: A Deep Time Journey* (Hamish Hamilton, 2019).
2. J. Campbell, *The Hero with a Thousand Faces* (Pantheon Books, 1949).

ACKNOWLEDGMENTS

This essay benefited from conversations with L. Fannin and C. Hobaiter.

10.1126/science.ady5476

The reviewer is at the Department of Anthropology and the Department of Biological Sciences, Dartmouth College, Hanover, NH, USA. Email: nathaniel.j.dominy@dartmouth.edu

PHOTO: COURTESY OF SANDBOX FILMS

ECOLOGY

North America's last prairies

The region's grasslands—vital ecosystems and carbon sinks—have been farmed and ranched beyond recognition **Sarah Boon**

Sea of Grass: The Conquest, Ruin, and Redemption of Nature on the American Prairie
Dave Hage and Josephine Marcotty
Random House, 2025. 400 pp.

Grasses sequester carbon, helping to mitigate global warming.

If asked to name endangered landscapes, many people would say the Amazon rainforest, oceans, or coral reefs. But what about grasslands? "Feared by pioneers, shunned by tourists, dismissed today as a wasteland best viewed from thirty thousand feet, the North American prairie is nonetheless one of the richest ecosystems on Earth," write journalists Dave Hage and Josephine Marcotty in their new book, *Sea of Grass*.

Rivaling the biodiversity of a tropical rainforest in their natural state, North American grasslands have today been farmed and ranched almost beyond recognition, with only islands of true prairie—characterized by grasses that exhibit most of their growth underground in root systems that can be up to 13 miles long—remaining. They sequester carbon in those roots and in their leaves; converting grasslands to other uses "is the climate change equivalent of adding 11.2 million cars to the road every year," note Hage and Marcotty. "Plowed grassland [can release] more than one-quarter of its stored soil carbon, most of that in the first five years," they add.

North American prairie farming began in the late 1800s, when the US government gave out parcels of land to European settlers looking to start a new life. (This land, of course, was not theirs to give, as it was populated by Indigenous nations who relied on it for food and the necessities of life and who were killed and/or relocated to make way for settler colonization.) Many settlers ended up on the eastern plains—in Iowa, Minnesota, and eastern Kansas, where wet and boggy tallgrass prairie proved difficult to farm at first.

Eventually, new technologies made farming more viable for the region. John Deere's steel plow and the introduction of drain tiles underneath farm fields made the prairie soil drier and easier to work. More recently, the introduction of herbicide-resistant seeds and oversized farm equipment has made planting easier and helped crops thrive. Even on the plains' western edge, where ranches outnumber farms because the summer has traditionally been too dry and the winter too long and cold to farm, shortgrass prairie is falling to the plow as seeds are engineered to require less water and less time to germinate. The North American tallgrass prairie disappeared in the 19th century, readers learn, and only 40% of the shortgrass prairie remains.

The most influential advance that has changed the region's prairie landscape has been the introduction of nitrogen and phosphorus–based fertilizer. Although it was originally harvested from bird guano, fertilizer is now produced in laboratories and targeted to meet the needs of specific crops. But it is not without environmental costs. Fertilizer contributes to nitrous oxide air pollution, and excess nutrients flow out of Midwestern fields via tile drains and into the Mississippi River. By the time the river reaches the Gulf of Mexico, it carries enough nutrients to form an ocean "dead zone"—the nutrients cause algae to bloom, which then die and fall to the bottom of the ocean, where they decay, depleting the water of oxygen and killing fish and other ocean species. Nitrogen also forms nitrate, which must be removed from municipal water supplies before the water can be used.

Pesticides are another key factor contributing to the changing prairie ecosystem. Monocultures of wheat and corn do not provide the diversity that many insects need to survive, and those seen as pests develop immunity to pesticides over time. Meanwhile, the number of insects continues to decline, in general.

It also does not help that the prairie and its deep root systems have been replaced by shallow-rooted crops, which leads to erosion during wind and rainfall events. This and other factors, such as drought, caused the severe dust storms of the 1930s as well as more recent storms. Since the 1950s, overall soil depth has decreased considerably in the region as topsoil is lost. Potential solutions to erosion include no-till agriculture and planting cover crops, both of which improve soil fertility, moisture, and structure. These kinds of solutions, however, have proven deeply unpopular with many farmers in the US. "Farmers are more likely to adopt a practice if they hear about it from a neighbor than if it comes from an expert," observe the authors, referencing research that has found as much. Governments have nonetheless advanced initiatives intended to both improve the environment and benefit farmers, subsidizing the creation of "buffer strips," which reduce fertilizer runoff and soil erosion and provide insect habitats, for example.

PLANTS FACING THE HEAT
More from this special issue on p. 1146.

The book ends with an exploration of bison on the prairie landscape, linking the species' presence to healthy grasslands and Native heritage. Prairie is meant to be grazed, Hage and Marcotty argue, and bison are the best evolved to graze it. They are also a critical part of Indigenous heritage. The Fort Peck Reservation in northeastern Montana, for example, has the biggest bison operation in the northern Great Plains, which it uses for both teaching and rewilding the landscape. "Having relentlessly sacrificed…the land, its Indigenous people, and its creatures…to achieve remarkable national prosperity, perhaps we're rich enough to restore some of what we've taken and heal the scars of our history," urge the authors in the book's closing line. □

10.1126/science.adw8576

The reviewer is a freelance writer and editor from Vancouver Island, Canada, and the author of the forthcoming book *Meltdown: The Making and Breaking of a Field Scientist* (Univ. of Alberta Press, 2025). Email: snowhydro1@gmail.com

PHOTO: JIM BRANDENBURG/MINDEN PICTURES

Illegal gold mining near Ecuador's Cerro Plateado Biological Reserve has caused deforestation and disrupted the riverbed.

PHOTO: DIANA SZÉKELY

The hidden costs of gold mining in Ecuador

A growing demand for gold (*1*) has triggered a rapid expansion of artisanal, small-scale, and often illegal mining in South America (*2*). Ecuador, a hotspot for biodiversity (*3*, *4*), defends the rights of nature through its constitution (*5*). Strong environmental laws regulate mining activities (*5*), and in a 2023 referendum, citizens chose nature protection rather than resource extraction (*6*). However, illegal mining activities are threatening the environment and Indigenous communities. The Ecuadorian government must protect its people by enforcing its environmental laws.

The environmental effects of gold mining are severe and wide-ranging. Deforestation caused by gold mining steadily increased between 2019 and 2024, including in protected areas and Indigenous territories (*7*). Mercury, cyanide, arsenic, and heavy metals associated with mining have polluted Ecuador's soil, water, and air, leading to the loss of ecosystem services, impaired water quality, and biodiversity decline (*8*).

The lack of adequate governmental response has provided unspoken permission for illegal mining activities (*9*), which have led to increased corruption and the involvement of organized crime syndicates that incorporate the mining industry into their money laundering schemes and drug and human trafficking activities (*10*). Indigenous communities, facing structural poverty and a long history of state neglect, must contend with a rapidly degrading environment that severely limits their access to traditional resources and tourism development prospects (*11*). The dangerous working conditions for local miners and the increase in criminality pose additional threats to these communities (*10*).

The extractivist approach is not a solution to the sociopolitical and environmental challenges that Ecuador faces. Instead, the country should advance the rights of nature framework beyond its conceptual foundations and recognize it as a binding legal instrument (*6*). Moreover, imposing stricter rules on gold trade worldwide, especially in countries that import gold from Ecuador, could limit the extent of small-scale mining. These national and international actions would mitigate the effects of mining, allow Ecuador and importing countries to adapt to a constantly changing socioecological reality, and increase ecosystem resilience in mining areas.

Dan Cogălniceanu[1] and Diana Szekely[1,2]

[1]Faculty of Natural and Agricultural Sciences, Universitatea Ovidius din Constanta, Constanta, Romania. [2]Museo de Zoologia, Universidad Tecnica Particular de Loja, Loja, Ecuador. Email: dcogalniceanu@univ-ovidius.ro

REFERENCES AND NOTES

1. World Gold Council, "Gold Demand Trends: Full Year 2024" (2025); https://www.gold.org/goldhub/research/gold-demand-trends/gold-demand-trends-full-year-2024.
2. G. P. Asner, R. Tupayachi, Environ. Res. Lett. 12, 094004 (2017).
3. T. Raz *et al.*, *J. Zool.* **322**, 126 (2024).
4. F. M. Sabatini *et al.*, *Nat. Commun.* **13**, 4683 (2022).
5. C. Mestanza-Ramón *et al.*, *Land* **10**, 1220 (2021).
6. C. A. Gallegos-Riofrío *et al.*, *Earth Syst. Gov.* **24**, 100253 (2025).
7. Monitoring of the Andes Amazon Program, "MAAP #184: Avance de La Actividad Minera En La Provincia de Napo (Ecuador)" (2023); https://www.maapprogram.org/es/mineria-napo-ecuador-2/.
8. I. M. Dossou Etui *et al.*, *Ecotoxicology* **33**, 484 (2024).
9. P. Cisneros, *Lat. Am. Policy* **12**, 6 (2021).
10. L. Wagner, M. Hunter, in *Illegal Mining: Organized Crime, Corruption, and Ecocide in a Resource-Scarce World*, Y. Zabyelina, D. van Uhmpp, Eds. (Springer, 2020), pp. 77–104.
11. R. A. Miserendino *et al.*, *Resour. Policy* **38**, 713 (2013).

10.1126/science.adx3004

Address Arctic shipping crisis at COP30

As a result of human activities, Arctic summer sea ice may disappear by 2030 (*1*). As the window for navigation in the Arctic shipping lanes has lengthened, the number and frequency of ships attempting to traverse the region have increased (*2*). Ships in the Arctic release black carbon, which accelerates ice melting, and the reduction in sea ice allows more ships to enter the Arctic for shipping activities (*3*). To mitigate the damage, attendees of the 30th United Nations Climate Change Conference in November (COP30) should commit to a legally binding Arctic shipping convention.

No meetings or guidelines have adequately addressed the climate and environmental risks of Arctic shipping. The International Code

of Safety for Ships Operating in Polar Waters lacks mandatory emission reduction standards (*4*). The Arctic Council's environmental initiatives have stalled, and recent United Nations Climate Change Conferences have not established goals to mitigate damage by ships in the Arctic (*5*). The International Maritime Organization ship carbon intensity index does not take into account the ecological vulnerability of ice navigation (*6*). The vague provisions in Article 234 of the United Nations Convention on the Law of the Sea (*7*), which gives coastal states environmental jurisdiction, has been overused by Russia, Canada, and other countries, resulting in fragmented international navigation rights, uneven safety standards, and the uncoordinated management of routes.

The expansion of shipping has stimulated the exploitation of Arctic resources, leading to the release of methane from the thawing of permafrost (*2*, *8*)—a feedback mechanism that could prematurely void the Paris Agreement's 2°C target (*9*). The asymmetric melting of the Antarctic ice sheet further amplifies the instability of the global climate system (*10*), and the Arctic ecosystem is more sensitive to pollution from shipping than expected (*11*). Less ice also leads to more floating ice and icebergs in shipping lanes, increasing the risk of shipping.

The COP30 Arctic shipping convention should clearly stipulate environmental protection standards, technical specifications for ships, and regulatory responsibilities of each country. Environmentally friendly ship technology, such as advanced antipollution equipment and more efficient ice navigation technology, should be required. An Arctic shipping carbon budget, aligned with the Intergovernmental Panel on Climate Change's target of a 40 to 70% reduction in emissions by 2050 (*12*), should be determined. The convention should require Arctic nations to improve international cooperation, build a transnational ice early warning system, and establish a governance body—informed by the practical experience of the C40 Cities Climate Leadership Group—that includes representation from Arctic port cities.

Zhicheng Du[1], Changyue Liu[1], Zhaotian Xie[2], Hui-Yan Luo[3], Wentao Zhang[2], Xin He[2]

[1]Institute of Biopharmaceutical and Health Engineering, Shenzhen International Graduate School, Tsinghua University, Shenzhen, China. [2]Institute of Materials Research, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China. [3]Department of Medical Oncology, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University, Guangzhou, China. Email: duzc21@mails.tsinghua.edu.cn

REFERENCES AND NOTES

1. A. Jahn, M. M. Holland, J. E. Kay, *Nat. Rev. Earth Environ.* **5**, 164 (2024).
2. Protection of the Arctic Marine Environment (PAME), "The increase in Arctic Shipping: 2013-2024" (2025); https://oaarchive.arctic-council.org/items/01ddf449-9048-4d6a-a056-65303831bb63/full.
3. H. Lee *et al*., "IPCC, 2023: Summary for Policymakers," in *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Core Writing Team, H. Lee, J. Romero, Eds. (IPCC, 2023), pp. 1–34.
4. H. Deggim, in *Sustainable Shipping in a Changing Arctic*, L. P. Hildebrand, L. W. Brigham, T. M. Johansson, Eds. (Springer, 2018), pp. 15–35.
5. T. Jiang *et al*., *Innovation* **5**, 100559 (2024).
6. International Maritime Organization (IMO), "EEXI and CII - Ship carbon intensity and rating system" (2023).
7. M. H. Nordquist, S. N. Nandan, Eds., *United Nations Convention on the Law of the Sea 1982: A Commentary - Volume VII* (Martinus Nijhoff Publishers, 2011).
8. J. C. Stroeve *et al*., *Science* **387**, 616 (2025).
9. United Nations, "Paris Agreement: Report of the Conference of the Parties to the United Nations Framework Convention on Climate Change" (21st session, 2015).
10. Z. An *et al*., *Science* **385**, 560 (2024).
11. X. Qi *et al*., *Ocean Coast. Manag.* **247**, 106936 (2024).
12. IPCC, Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, P. R. Shukla *et al*., Eds. (Cambridge Univ. Press, 2022).

10.1126/science.adx4305

US tariffs jeopardize medicine supply chains

In April, the Trump administration announced sweeping tariffs on key trade partners, and it is poised to include pharmaceuticals next (*1*). Tariffs increase input costs for domestic drug manufacturers, particularly producers of generic medications, who operate on thin margins (*2*, *3*). Moreover, because the US produces only a fraction of the generic medications that it needs (*4*), such drugs could serve as a cheap yet potent geopolitical tool for retaliation. Rather than expanding tariffs to pharmaceuticals, the US should build resilience in its medicine supply chains through targeted investments and international coordination.

Effective health care depends on uninterrupted access to critical medicines, including antibiotics, painkillers, and anesthetics (*5*). Such medications are predominantly low-margin generics produced through opaque, highly fragmented supply chains with low redundancy (*3*, *6*). With manufacturing concentrated in only a few facilities worldwide, disruptions—whether from natural disasters or geopolitical tensions—can rapidly trigger downstream shortages (*7*).

The US is particularly vulnerable to these disruptions because of the trade asymmetry in generic drug sources and the relative ease with which trade partners could limit the US supply. Although generics account for more than 80% of global prescriptions, they represent only 20 to 30% of market value (*8*). Therefore, countries targeted by US tariffs, even those unrelated to pharmaceuticals, could retaliate by restricting exports of critical medicines, with limited effects on their own revenues. A sudden drop in US pharmaceutical imports could trigger shortages within weeks (*9*), affecting hundreds of thousands of patients.

These tensions should prompt a broader reassessment of pharmaceutical supply chain resilience. Short-term emergency responses may delay the onset of shortages by optimally redistributing existing stocks (*9*, *10*). Long-term industrial strategies, such as reshoring pharmaceutical manufacturing, may improve supply chain robustness by addressing their root causes (*4*, *11*). However, both short- and long-term solutions depend on a thorough understanding of national distribution networks and global pharmaceutical supply chains. Without reliable, granular data on production capacity, sourcing patterns, and inventory levels, policy-makers cannot accurately assess vulnerabilities or prioritize interventions. Most important, robust, data-driven methods to assess supply chain resilience are not available. Policy-makers and researchers must work together to foster a systemic understanding of pharmaceutical supply chain risks.

Giona Casiraghi, Georges Andres, Frank Schweitzer, Luca Verginer

Department of Management, Technology, and Economics, ETH Zürich, Zürich, Switzerland. Email: lverginer@ethz.ch

REFERENCES AND NOTES

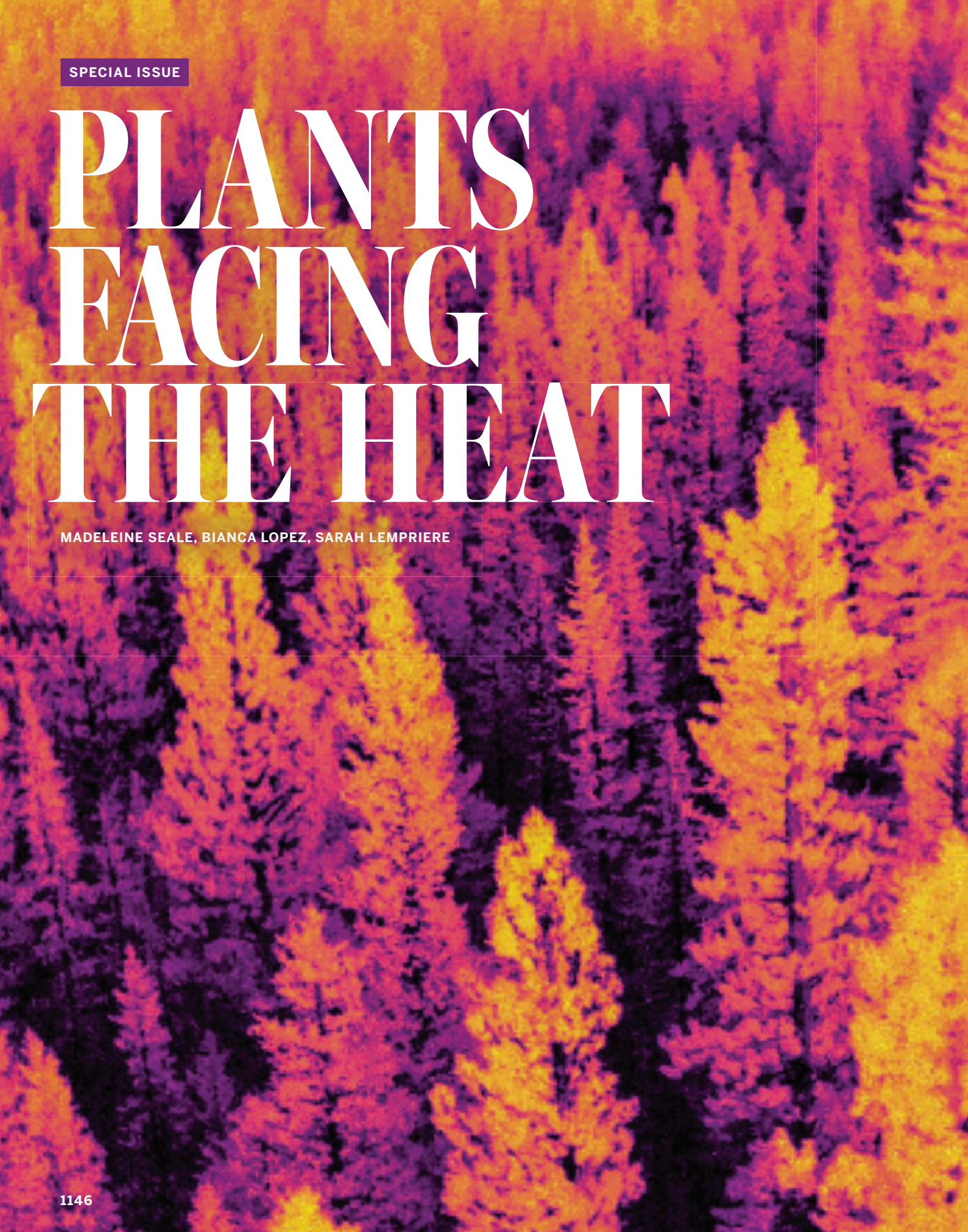
1. Department of Commerce, Bureau of Industry and Security, "Notice of Request for Public Comments on Section 232 National Security Investigation of Imports of Pharmaceuticals and Pharmaceutical Ingredients," 90 FR 15951 (2025); https://federalregister.gov/d/2025-06587.
2. Association for Accessible Medicines, "Association for Accessible Medicines comments on new tariffs" (2025).
3. National Academies of Sciences, Engineering, and Medicine, *Building Resilience into the Nation's Medical Product Supply Chains* (National Academies Press, 2022).
4. US Food and Drug Administration, "Drug Shortages: Root Cause and Potential Solutions," technical report (2019).
5. European Medicines Agency, "First version of the Union list of critical medicines agreed to help avoid potential shortages in the EU" (2023).
6. T. de Jongh *et al*., "Future-proofing pharmaceutical legislation – Study on medicine shortages" (European Commission, Directorate General for Health and Food Safety, 2021).
7. F. A. Miller, S. B. Young, M. Dobrow, K. G. Shojania, *BMJ Qual. Saf.* **30**, 331 (2021).
8. G. Siotis, C. Ornaghi, M. Castanheira, "Market definition and competition policy enforcement in the pharmaceutical industry" (ECARES, 2020).
9. A. Amico, L. Verginer, G. Casiraghi, G. Vaccario, F. Schweitzer, *Sci. Adv.* **10**, eadj1194 (2024).
10. A. Amico, L. Verginer, F. Schweitzer, *Sci. Data* **11**, 774 (2024).
11. B. Roehr, *BMJ* **370**, m3393 (2020).

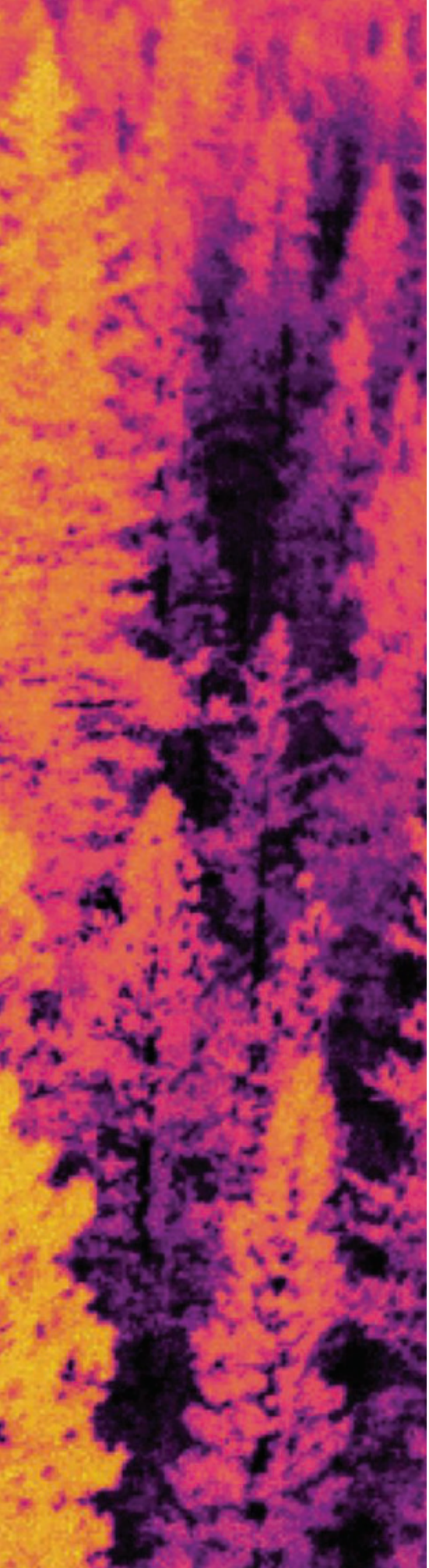
10.1126/science.adx7154

SPECIAL ISSUE

PLANTS FACING THE HEAT

MADELEINE SEALE, BIANCA LOPEZ, SARAH LEMPRIERE





CREDIT: ANDREW D. RICHARDSON ADAPTED BY M. HERSHER/*SCIENCE*

As continued greenhouse gas emissions turn up the world's heat, the plants in our varied natural ecosystems and agricultural land increasingly feel the impact of high temperatures. Plants have a toolbox of cellular and physiological heat responses to promote cooling and repair damage, as well as population responses through adaptation and dispersing to track climate shifts. The ability of plants to respond to heat scales up to influence the productivity and resilience of ecosystems and food production systems.

This Special Issue examines how heat affects plants at multiple levels. Researchers have uncovered many ways in which plants sense heat and integrate temperature into signaling pathways. Biophysically, heat affects the rates of reactions, structural stability, and phase separation within cells. With photosynthesis underlying the majority of life's carbon fixation, its sensitivity to heat is of particular importance. Efforts to optimize crop architecture, water use, and photosynthetic efficiency under high temperatures offer promising paths forward to mitigate some of the effects of heat on crop production. Beyond tweaking known heat-tolerance mechanisms, wild plant populations harbor diverse strategies to cope with temperature change. A better understanding of these mechanisms through sequencing and analysis of plant genomes may provide alternative ways to help plants adapt to high heat. However, predicting whole-ecosystem responses to warming requires recognizing that heat responses at the leaf and plant scales do not always apply. At larger scales, vegetation structure and interactions with other species, such as plant-associated microbes, play an important role.

These layers of regulation and response to high temperatures hang in a fine balance, and many ecosystems—such as prairie grasslands, tropical forests, and croplands—are suffering as climate change leads to increasing temperature extremes. Understanding how plants respond to higher temperatures across scales is critical to predict and mitigate the impacts of warming on vegetation and, ultimately, the people who depend on it. With our global food system and ecosystems at stake, we turn our attention to how plants are facing the rising heat.

10.1126/science.adz0700

Thermal imagery captures variation in temperature across a forest canopy in Colorado, US. How plants experience and cope with heat varies across spatial, temporal, and biological scales.

Wild relatives of cultivated sunflowers have evolved a diverse range of strategies for tolerating climate stress.

PERSPECTIVE

Can wild plant adaptations help crops tolerate heat?

Wild plant species harbor a vast but largely unknown diversity of temperature stress solutions

Sam Yeaman

Climate change is causing global heating and pronounced shifts in precipitation regimes, posing a substantial evolutionary challenge to life on Earth. Plants are particularly sensitive to climate, as they cannot relocate to mitigate its impact. To cope with historical environmental extremes, plants have evolved a range of strategies, from physiological changes that increase water use efficiency to shifting flowering time to avoid periods of drought. All such changes affect the energy available for growth, and understanding these trade-offs is particularly important in crop breeding, where the goal of maximizing yield must be balanced against the need to survive stress during growth. With some models predicting reductions in crop yield of 7 to 23% as a result of climate change (*1*), offsetting these effects will require inventive solutions. The diversity of adaptive strategies found in wild plant species may provide some of the answers.

Biotechnology and genomics have produced a wide array of methods for improving crop yields and stress tolerance traits, from genomic selection that complements traditional breeding practices, to mutagenesis, to using genome-wide association studies (GWASs) to identify causal variants that yield desirable characteristics. With the advent of CRISPR-Cas9 genome editing technology, it is now possible to edit specific genetic variants into breeding lines, enabling changes that might be exceedingly slow or even impossible to achieve with conventional breeding. Genome editing techniques have been experimentally deployed on a wide range of crop species, often inactivating or removing (knocking out) target genes to yield increases or decreases in stress tolerance traits (*2*). However, improvements in stress tolerance can have negative side effects on other traits, so agronomically useful outcomes may require editing of many loci, consistent with the highly polygenic nature of traits such as heat stress tolerance. If they only focus on a target agricultural species, such approaches are limited to using the existing pool of genetic variation present in these species and the additional space accessible by mutagenesis or knockouts. It has long been recognized that wild relatives of crops provide an additional source of genetic diversity (*3*); however, this still represents a relatively small pool compared with the vast array of adaptive strategies and trait variation present across

thousands of wild species of plants. Studying the diversity of ways that wild plants adapt to extreme climates is enabling the discovery of pathways, genes, and variants that exist outside the small pool of agricultural species, providing new targets for biotechnology.

Continued improvements in genome sequencing are substantially increasing the number of species with sequenced genomes. The Earth BioGenome Project is coordinating the assembly of the genomes of all known complex species on Earth, including more than 10,000 plants. However, a single genome does not reveal much about how a species adapts to its environment, even when compared to the genomes of thousands of other species. Phylogenetic methods use comparisons among many species to identify genes with particularly fast rates of change in the amino acid sequence of the proteins they encode. But such methods are only sensitive to detecting evolutionary change that proceeds by the accumulation of many amino acid substitutions over a sustained period. As such, they often identify genes involved in an evolutionary arms race—for example, genes involved in the immune system (*4*). If adaptation has proceeded by changes in regions of the genome that do not encode proteins or by just a few important changes in amino acids, it can be difficult or impossible to detect with phylogenetic methods.

...improvements in stress tolerance can have negative side effects on other traits...

To reveal the secrets of how a species thrives in extreme climatic conditions, hundreds or thousands of individuals must be sampled from across the range of environmental conditions that it inhabits. This enables the use of methods that search for signatures of association between the frequency of an allele in a population and the environment where the population is found (e.g., alleles associated with temperature or amount of precipitation). These methods apply approaches similar to GWASs but search for associations with the environment inhabited by each individual plant, instead of the phenotype. However, although hundreds of studies have been published using such methods, it is unclear how to compare results among them to identify genes that are particularly important for tolerance to environmental stressors such as extreme heat. In any given species, it is notoriously difficult to parse the signal of natural selection from the noisy background of random fluctuations in allele frequency (genetic drift). Different methods that correct for the effect of genetic drift over space (population structure) can yield markedly different results, with some better at reducing false positives at the expense of missing some true signals and others prioritizing the reverse (*5*). Comparison of the sets of putatively adaptive genes identified by different studies is therefore complicated by heterogeneity of these analysis methods.

A promising approach is to combine many single-species datasets together and test for genes that have strong signatures of adaptation across many independently evolving species (repeated adaptation). This also avoids the thorny problem of correcting for population structure, as genetic drift is a random process that is unlikely to generate chance associations in multiple species. A recent study of 25 distantly related plant species found many genes with signatures of repeated adaptation to spatial variation in temperature and precipitation (*6*), including well-known candidate genes such as *FLOWERING LOCUS T* (*FT*), which affects response to seasonal changes in day length in the model species *Arabidopsis thaliana* (*7*). Interestingly, little is known about some of the genes with the strongest signals of repeated adaptation, such as AT5G16780 and AT3G03010, which exhibited strong association to precipitation variables in 13 of 25 species. No single gene showed involvement in all species, and most of the 108 candidates had strong signatures in only ~4 to 8 species (*FT* was only strongly associated with precipitation, and in just 3 species). Presumably, gene families with members like ubiquitin-related protein 3 (*RUB3*), which exhibited strong associations to temperature or precipitation variables in 14 of 25 species, are prime candidates for biotechnology applications aimed at improving tolerance to stress associated with changes in these environmental conditions.

However, like other genome scan methods, this approach provides little scope to identify the causal nucleotide changes driving adaptation, as many neighboring neutral variants exhibit strong but spurious associations owing to chromosomal proximity (genetic linkage) with the variant under selection. Because identifying causal variants requires labor-intensive manipulation and functional validation, bioinformatic approaches can be used to generate a shortlist of top candidates most likely to have phenotypic effects. Many tools use phylogenetic conservation to identify amino acid changes that are rarely observed across the tree of life, suggesting that they are more likely to have substantial effects on survival or reproduction (*8*). Recently developed methods such as AlphaMissense use artificial intelligence to model protein folding to predict the impact of an amino acid substitution (*9*). Such tools provide promising ways to both improve understanding of adaptation and refine the search for variants of agronomic importance prior to functional validation.

Although finding genes involved in repeated adaptation can reveal promising candidates for biotechnology, studying why adaptation is nonrepeated is just as important for understanding how species will respond to climate change. For example, whereas some studies in wild *A. thaliana* have found large effects of *FT* driving changes in the timing of flowering among populations inhabiting different environments, others have implicated other loci of large effect such as *FRIGIDA* (*FRI*) and *FLOWERING LOCUS C* (*FLC*) (*10*) and also typically find many alleles of small effect (*7*). These results suggest considerable redundancy in the number of ways a species can adapt to a given stress. Focused experimental work can help clarify how such variants affect growth rate and fecundity, providing information on the nature of trade-offs that will be critical to manage in an agronomic context, where both yield and resistance to stressors such as extreme heat are desirable traits.

Although repeated adaptation has a relatively clear explanation, a lack of repeated signal may occur for many reasons (*11*). A recent meta-analysis of the repeatability of genes underlying phenotypic variation in both plants and animals found that closely related species have more overlap than distant ones (*12*). Also, pairs of species that occupy similar environments exhibit higher signatures of repeated adaptation than those that inhabit more different environments, as shown in stick insects (*13*). In other cases, populations of the same species that inhabit similar environments may exhibit strikingly different phenotypes, a phenomenon that has been studied for drought escape (speeding up development to complete a life cycle before drought) versus avoidance (reducing water loss) in *A. thaliana*, sunflower (*Helianthus anomalus*), monkeyflower (*Mimulus guttatus*), and others (*14*). However, even for species evolving the same trait in response to the same environment, high redundancy in the mapping from genotype to phenotype to fitness can also result in nonrepeated adaptation at the level of the gene. High redundancy occurs when a trait is very highly polygenic, with many more loci that can mutate to cause changes in trait values than would be needed to yield an optimal change in phenotype. Therefore, as little is known about the extent of redundancy in tolerance to climate stress, it remains a key area for further study. Bigger projects are now underway exploring repeatability across many species, with the California Conservation Genomics consortium analyzing 81 plant species across the range of the state and the RepAdapt working group combining data from 164 plant species worldwide. As more examples of the comparative genomics of adaptation accumulate, much more will be learned about the fundamental factors that shape how the interplay between genetics and evolution shapes stress tolerance traits.

How does studying repeated adaptation translate back into knowledge useful for genetic improvement of agricultural crops? Conventional breeding and biotechnology will undoubtedly continue to improve crop yield and stress tolerance, but insights from wild plants will help both to identify new targets for biotechnology and to determine which combination of approaches is likely to yield optimal results for a given trait of interest. Studying the diversity of adaptive solutions in wild species can reveal which kinds of traits are so highly polygenic and redundant that they will be more suited to conventional breeding approaches or genomic selection rather than targeted gene editing. For such traits, which exhibit minimal repeated adaptation, the effect sizes of individual mutations tend to be so small that it is exceedingly difficult to identify causal alleles, and any focused editing would require changes at too many loci to be practical, given current methods. By contrast, traits that adapt through changes in a few key genes in a repeated way across many species likely involve mutations of larger effect. Although such traits may also be polygenic, they will be more amenable to biotechnology approaches that target these key genes, for example, by adding genes (transformation), gene editing of existing sequence (by CRISPR-Cas9), and mutagenesis paired with screening. As gene editing methods improve and enable simultaneous targeting of many loci (*15*), it will become increasingly important to find good targets.

Understanding the genetic basis of plant adaptation to heat stress and other shifts associated with climate change will help improve outcomes in agriculture and forestry, but it is unclear if it will be enough. Given the unprecedented pace of current climate change, temperature extremes are expected to extend well beyond historical limits, which may push crops beyond improvements in heat resistance that can be afforded by breeding and engineering. Genetic improvement of crops therefore constitutes only one piece of what must be a many-faceted strategy for adapting agriculture to climate change (*1*). Increased rates of crop failure associated with climate change may substantially reduce yields of conventional crop varieties when averaged over both good and bad growing seasons. In such conditions, stress-tolerant genotypes with lower yields per year may achieve higher average yields over multiple years if crop failure is included in the calculation. Inevitably, the nature of trade-offs between growth and stress tolerance needs to be understood to find optimal genotypes, and the diversity of the natural world provides a veritable library of solutions to be studied. □

REFERENCES AND NOTES

1. E. E. Rezaei *et al.*, *Nat. Rev. Earth Environ.* **4**, 831 (2023).
2. M. A. Aziz *et al.*, *Front. Plant Sci* **13**, 1027828 (2022).
3. H. Dempewolf *et al.*, *Crop Sci.* **57**, 1070 (2017).
4. A. F. Moutinho, F. F. Trancoso, J. Y. Dutheil, *Mol. Biol. Evol.* **36**, 2013 (2019).
5. S. Hoban *et al.*, *Am. Nat.* **188**, 379 (2016).
6. J. R. Whiting *et al.*, *Nat. Ecol. Evol.* **8**, 1933 (2024).
7. G. A. Auge, S. Penfield, K. Donohue, *New Phytol.* **224**, 55 (2019).
8. D. Wang, J. Li, Y. Wang, E. Wang, *NAR Genom. Bioinform.* **4**, lqab122 (2022).
9. J. Cheng *et al.*, *Science* **381**, eadg7492 (2023).
10. A. Fulgione *et al.*, *Nat. Commun.* **13**, 1461 (2022).
11. D. I. Bolnick, R. D. H. Barrett, K. B. Oke, D. J. Rennison, Y. E. Stuart, *Annu. Rev. Ecol. Evol. Syst.* **49**, 303 (2018).
12. M. Bohutínská, C. L. Peichel, *Trends Ecol. Evol.* **39**, 396 (2024).
13. S. Chaturvedi *et al.*, *Nat. Ecol. Evol.* **6**, 1952 (2022).
14. N. J. Kooyers, *Plant Sci.* **234**, 155 (2015).
15. A. Berman *et al.*, *Nat. Commun.* **16**, 4111 (2025).

ACKNOWLEDGMENTS

S.Y. thanks S. Verma, A. Xenos, and T. Booker for comments and suggestions. S.Y. acknowledges support by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant.

10.1126/science.adw3673

Department of Biological Sciences, University of Calgary, Calgary, AB, Canada.
Email: samuel.yeaman@ucalgary.ca

PERSPECTIVE

Plant microbiomes feel the heat

Rising temperatures change the structure and function of plant microbial communities

Anna-Liisa Laine *and* **Sara Leino**

Rising global temperatures—a key component of climate change—are reshaping biodiversity, yet their effects on the plant microbiome remain poorly understood. This microbiome, composed of diverse bacteria, fungi, and other microorganisms, forms dynamic communities in above- and belowground plant tissues and surrounding soil. Microbes influence plant growth and resilience through interactions that range from beneficial (e.g., nutrient provisioning) to antagonistic (e.g., disease causing). Microbial communities are affected by their plant hosts as well as microbe-microbe interactions and environmental conditions. These factors often interact in dynamic ways, making it challenging to disentangle their individual contributions, including the confounding effects of drought and temperature-induced stress—the two major features of climate change that profoundly influence plants. Understanding the roles of biotic and abiotic factors in shaping microbiome-mediated plant performance is vital for predicting ecological change and maintaining ecosystem functions in the face of rising temperatures.

...advanced methods are enabling plant microbiome research to move beyond descriptive studies...

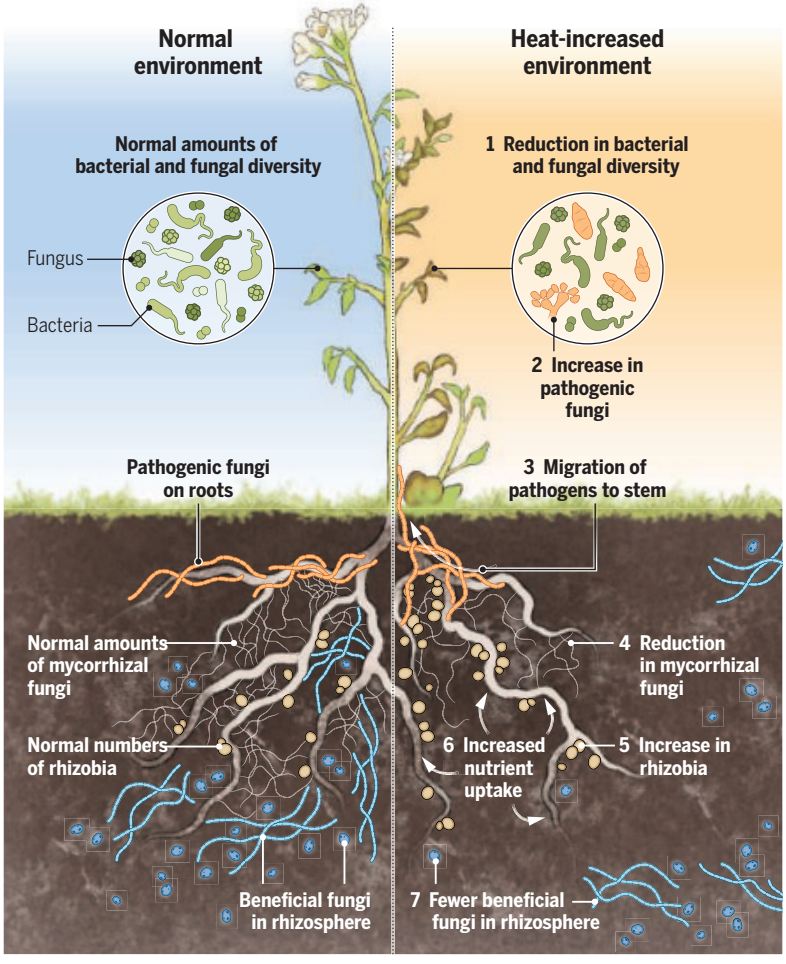
The plant microbiome is complex and diverse and varies among different plant genotypes, populations, and species (*1*–*3*). This makes deciphering the dynamics involved in plant microbiome formation and function technically challenging. Nevertheless, advanced methods are enabling plant microbiome research to move beyond descriptive studies toward uncovering the intricate mechanisms that govern how these microbial communities respond to their environment. For example, microcosms—simplified, laboratory-based systems constructed using microbes of known origin—allow the effects of specific environmental stressors (such as drought, flooding, or temperature changes) on microbial communities to be studied under controlled and reproducible conditions. This approach removes confounding influences, such as soil heterogeneity and the complexity introduced by metabolically inactive microbes found in natural ecosystems, and provides a powerful framework for investigating the effects of biotic and abiotic stressors (*4*). Sterile systems, in which the ge-

netics of plants and microbes as well as environmental conditions can be tightly controlled, are allowing researchers to isolate and observe the effects of rising temperatures on plant-microbe interactions. Synthetic communities are a type of sterile system in which well-characterized microbial strains are cocultured on a plant host. This allows targeted manipulation of microbiome composition and observation of microbial activities, such as root colonization, as well as interactions with the plant host that range from cooperation to competition (*5*, *6*). Moreover, the development of molecular methods for precise tracking of microbial strains has enabled the differentiation, quantification, and monitoring of synthetic community members under various complex conditions (*7*). Despite being a simplified representation of natural microbial communities, these methodological innovations provide essential tools for unraveling complex host-microbe dynamics and are being used to explore how rising temperatures affect plant-associated microbial communities.

Microbial taxa are adapted to distinct climatic niches, and consequently, warming can drive shifts in microbial richness (the number of different microorganisms present) and phylogenetic diversity to varying degrees in different taxa (*8*). This is exemplified by the differential responses of plant-associated fungal and bacterial communities to rising temperatures, with fungi often exhibiting greater compositional turnover compared with bacteria (*9*, *10*). Fungal communities respond strongly to climate stressors, such as increased temperature, with warming substantially reducing fungal richness, altering species composition, and decreasing the abundance of key functional groups such as mycorrhizal fungi—fungi that form nutrient-exchanging associations with plant roots—which ultimately weakens plant-microbe interactions and reduces ecosystem resilience (*8*, *9*). Bacterial communities appear to be less affected by temperature changes, although they remain closely linked to fungal communities through trophic interactions (*11*), which highlights the interconnected nature of plant-associated microbial networks. There is also evidence of distinct responses to temperature among different families of viruses that cohabit the same plant individuals (*3*). These findings underscore the need to consider taxon-specific and network-level responses when assessing the impacts of warming on plant-associated microbiomes.

Temperature alters plant-microbe interactions

The plant microbiome is complex and diverse and varies among different plant genotypes, populations, and species. Advances in methods are increasing understanding of the effect of heat on the plant microbiome. Key recent findings include heat-induced declines in leaf fungal and bacterial diversity (**1**), greater presence of pathogenic fungi on leaves (**2**), and increased migration of pathogenic fungi from the root to the stem (**3**). Warming also decreases the abundance of symbiotic mycorrhizal fungi (**4**) and can increase the activity of nitrogen-fixing bacteria that form nodules (rhizobia) on the roots of legumes (**5**), which has implications for plant nutrition and the global nitrogen cycle (**6**). In addition, higher temperatures reduce the migration of beneficial fungi, such as Ascomycota and yeasts, from the bulk soil to the soil surrounding the roots [rhizosphere (**7**)].



GRAPHIC: A. FISHER/*SCIENCE*

Changing temperatures may alter not only the taxonomic diversity of the plant microbiome but also its ecological functions that range from beneficial to disease causing. For example, in *Schima superba* seedlings, a 3°C temperature increase reduced the migration of beneficial fungi, such as Ascomycota, soil saprotrophs, and yeasts, from bulk soil to the rhizosphere (soil surrounding the roots) while increasing the movement of fungal plant pathogens from roots to stems (*8*). Warming also favored plant pathogens more generally, increasing their richness and phylogenetic diversity, with species of fungi such as *Alternaria*, *Fusarium*, and *Botrytis* thriving under elevated temperatures. Another recent study explored the effects of heat stress on microbial plant symbionts in the soil and found that mycorrhizal fungi decreased by an average of 59% in relative abundance in response to a 3.7°C temperature increase (*9*). This reduction could impair plant nutrient uptake and stress resilience. By contrast, rhizobia—nitrogen-fixing bacteria that form symbiotic nodules with legumes—showed increased diversity and abundance in response to heat, suggesting a potential rise in nitrogen-fixing activity in legume-associated plants, with implications for plant nutrition and the global nitrogen cycle under future climate conditions (see the figure). Temperature also influences the ecological functions of microbiomes by altering the balance between generalist and specialist microbial taxa within plant microbiomes—although generalists can thrive across a wide range of conditions and hosts, specialists are adapted to narrower ecological niches or specific plant partners (*2*). Hence, warming is likely to favor generalists over specialists, potentially reducing functional specialization and altering the stability and adaptability of plant microbiomes under changing environmental conditions.

The effects of warming on the microbiome vary across plant compartments, with some studies reporting more pronounced shifts in the diversity and structure of the leaf-associated microbial communities (phyllosphere) compared with those interacting with plant roots. In the phyllosphere, warming tends to reduce fungal and bacterial diversity, whereas in the rhizosphere, major changes in overall diversity are less apparent, although microbial community composition shifts as a result of species turnover (*10*). Disentangling the effects of a warming environment remains challenging because experimental warming

may not result in uniform temperature changes above- and belowground. Many studies simulate ambient temperature changes in controlled climate chambers without directly measuring temperature changes in specific plant compartments (*4*, *8*, *10*). Nevertheless, differences in the responses of above- and belowground microbial communities to changing temperatures can create trophic "mismatches," whereby temperature-sensitive leaf-associated microbes disrupt plant resource availability for higher trophic levels, such as herbivores, whereas the more stable rhizosphere microbiome continues to support belowground consumers, including detrivores, which feed on dead organic matter (*12*). It has been hypothesized that this potential imbalance could, over time, alter key ecosystem processes and affect the stability of terrestrial food webs (*12*). Understanding the long-term consequences of such mismatches under warming conditions is an important direction for future research.

In the phyllosphere, warming tends to reduce fungal and bacterial diversity...

Belowground, temperature changes can indirectly alter plant-microbiome interactions by modifying soil nutrient dynamics, thus generating feedback mechanisms that can influence plant growth, species coexistence, and competition (*11*). Heat can increase nitrogen availability in soil by altering the activity of nitrogen-transforming bacteria (*13*). This could affect the initiation and strength of symbiotic interactions with ectomycorrhizal and arbuscular mycorrhizal partners. For example, in *Populus* species, the formation of associations with ectomycorrhizal fungi decreased under nitrogen addition, whereas associations with arbuscular mycorrhizal fungi were unaffected by the nitrogen treatment (*11*). Furthermore, shifts in soil mineral nitrogen concentrations under heat stress can, along with changes in microbial communities, interactively influence outcomes such as plant-plant competition (*11*, *13*). For example, increased nitrogen cycling by bacteria can enhance nutrient uptake by plants (*13*), which could intensify competition among plants. As climate change progresses, microbial interactions are expected to play an increasingly central role in mediating plant competition.

Plant defense pathways, which detect and respond to microbial threats, can influence which microbes plants associate with (*3*, *6*). A key component of plant defense pathways is respiratory burst oxidase homolog protein D (RBOHD), a membrane-bound enzyme in the nicotinamide adenine dinucleotide phosphate (NADPH) oxidase family that produces reactive oxygen species (ROS), which are crucial signaling molecules involved in plant stress and immune responses. Heat-induced reductions in ROS signaling are expected to alter microbial interactions (*14*). In a recent *Arabidopsis thaliana* synthetic community study, the bacterium *Xanthomonas* was found to suppress other bacterial pathogens when its virulence genes were inhibited by ROS (*6*). Heat stress also suppresses systemic acquired resistance (SAR), reduces salicylic acid accumulation, and increases pathogen susceptibility of plants (*14*). For instance, in a sterile system, thermopriming *A. thaliana* by exposing plants to sublethal heat (40°C for 2 hours) was found to inhibit the activation of SAR by pathogens. SAR is a form of induced defense that provides long-lasting protection against a broad range of microorganisms (*14*). Therefore, understanding how plants regulate their microbiomes in fluctuating temperatures will be critical for enhancing resilience to both heat and disease in a changing climate.

Temperature-induced shifts in plant-microbe interactions have the potential for cascading effects on ecosystem stability, yet their consequences remain poorly understood. A key priority for future research is to identify microbial taxa or traits most sensitive to warming, particularly those mediating nutrient acquisition or disease resistance. Establishing long-term field experiments that track shifts in microbial and plant community composition under changing temperatures and link these to functional outcomes is essential. Another promising area is leveraging existing common garden experiments—where plants of different genotype, species, or origin are grown in the same environment—or warming experiments by adding microbial sampling to assess changes in composition, diversity, and function. Wide geographical coverage is essential because global warming does not follow a uniform global pattern: The extent and direction of temperature change can vary substantially across regions. Developing rapid, standardized protocols for soil and root microbiome monitoring could also help integrate microbes into ongoing ecosystem surveillance programs. These approaches should be coupled with controlled experiments (such as mesocosms or synthetic community systems) and molecular analyses to dissect how warming alters host-microbe signaling, microbial gene expression, and community formation.

Applied microbiome science offers innovative solutions for addressing agricultural and environmental challenges. Beneficial microbes could be harnessed through targeted applications such as microbiome engineering, which involves modifying or managing microbial communities to promote plant health, and biostimulants, microbial products that enhance plant growth and stress tolerance. This could strengthen plant resilience, reduce reliance on chemical inputs such as synthetic fertilizers and pesticides, and improve food security in a warming world. Additionally, optimizing management practices—such as increasing plant diversity to support soil microbial communities (*15*)—could enhance ecosystem resilience to warming in a sustainable manner. These approaches align with the One Health framework, which recognizes the interconnectedness of plant, human, and environmental health. However, their success is entirely dependent on a better understanding of how microbes influence plant responses to warming. Integrating microbial findings into ecosystem management and land-use strategies will be essential for developing climate-adaptive solutions that support biodiversity and ecosystem function in both natural and managed landscapes. □

REFERENCES AND NOTES

1. J. Chamard, M. Faticov, F. G. Blanchet, P.-L. Chagnon, I. Laforest-Lapointe, *Commun. Biol.* **7**, 360 (2024).
2. R. T. Nethmini *et al.*, *Commun. Biol.* **8**, 5 (2025).
3. A. Norberg *et al.*, *Curr. Biol.* **33**, 1665 (2023).
4. C. G. Knight *et al.*, *Nature* **636**, 690 (2024).
5. T. R. Northen *et al.*, *Nat. Microbiol.* **9**, 2774 (2024).
6. F. Entila, X. Han, A. Mine, P. Schulze-Lefert, K. Tsuda, *Nat. Commun.* **15**, 456 (2024).
7. B. Jorrin *et al.*, *Microbiome* **12**, 81 (2024).
8. X. Wu, M. D. MacKenzie, J. Yang, G. Lan, Y. Liu, *Mol. Ecol.* **34**, e17652 (2025).
9. J. A. Boyle, B. K. Murphy, I. Ensminger, J. R. Stinchcombe, M. E. Frederickson, *Ecosphere* **15**, e70077 (2024).
10. D. Hoefle *et al.*, *Environ. Microbiome* **19**, 62 (2024).
11. M. E. Van Nuland, P.-J. Ke, J. Wan, K. G. Peay, *J. Ecol.* **111**, 564 (2023).
12. M. P. Thakur, *Biol. Lett.* **16**, 20190770 (2020).
13. W. Lin *et al.*, *Ind. Crops Prod.* **199**, 116764 (2023).
14. A. Nishad, J. K. Gautam, I. Agarwal, A. K. Nandi, *Plant Cell Environ.* **48**, 3352 (2025).
15. L. A. Domeignoz-Horta *et al.*, *Nat. Commun.* **15**, 8065 (2024).

ACKNOWLEDGMENTS

The authors acknowledge financial support from the Research Council of Finland (grant 362242) and the European Research Council (AdG 101097545 Co-EvoChange) to A.-L.L.

10.1126/science.adw3659

Research Centre for Ecological Change, Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland. Email: anna-liisa.laine@helsinki.fi

REVIEW

Safeguarding crop photosynthesis in a rapidly warming world

Carl J. Bernacchi[1,2,3,4]*†, Stephen P. Long[2,3,4]*†, Donald R. Ort[2,3,4]*†

Continued greenhouse gas emissions will accelerate global warming and intensity of heat waves, which already harm crop productivity. From the stability of key enzymes to canopy processes, photosynthesis is affected by temperature. All crops suffer declines in photosynthetic rate when temperatures cross critical thresholds, with irreversible losses typically occurring above 40° to 45°C. Protective measures within plants can be induced by growth at elevated temperatures but not from the sudden temperature elevation of heat waves. Strategies to improve the heat resilience of photosynthesis include modifying surface energy balance, optimizing canopy architecture, improving enzymatic heat tolerance, and (re)engineering key metabolic pathways for greater efficiency or to remove bottlenecks. This Review summarizes present knowledge on the major mechanisms that underlie high-temperature inhibition of photosynthesis and explores opportunities for breeding and biotechnological interventions to overcome them.

Greenhouse gas emissions from anthropogenic activities are driving an upward shift in global mean temperatures, which have already surpassed the preindustrial baseline by 1.5°C in 2024 (*1*) and are projected to rise a further 1.2°C by 2050 under the "business-as-usual" scenario SSP5-8.5 of the Intergovernmental Panel on Climate Change (IPCC) (*2*). This reflects a global average that does not account for spatial and temporal variability. Terrestrial surfaces are warming faster than ocean surfaces, and temporal variability is driving a large increase in short-duration extreme temperature events (heat waves) (*3*). For example, Coordinated Regional Downscaling Experiment (CORDEX) model ensembles (*2*) predict that for Southeast Asia, the number of days per year with a heat index exceeding 41°C will increase from around 50 in the period 1995–2014 to 160 for 2041–2060 and 300 by 2081–2100. By mid-century, the southern United States can expect around 50 more days at these temperatures per year and central and northern Brazil some 100 more days. All crop-growing regions of the world will see an increased number of days with such extreme temperatures (*2*). Rising average global temperatures, coupled with these increasingly frequent heat waves, already pose a major threat to agricultural productivity worldwide (*4*).

Temperature plays a dominant role in all facets of crop physiology, and photosynthesis, the primary means by which carbon is assimilated for growth and yield, is of particular importance (*5*). Many years of field experiments studying elevated atmospheric [CO_2] (*6*) have shown the close link between net photosynthesis and crop productivity. Any decline in photosynthetic efficiency with temperature increase has the potential to substantially affect yields. High temperatures affect photosynthesis at multiple levels, from altering energy balance and gas exchange to impairing enzyme function and altering gene expression (*7*). Crops may be divided into two photosynthetic types: C_3 and C_4 (see Box 1 for an explanation of these and other terms). Because it uses a different metabolic process, C_4 photosynthesis has a greater degree of resilience to warmer temperatures (*8*) than C_3 photosynthesis (Fig. 1). Substantial declines in the net rate of CO_2 uptake per unit of leaf area (A) occur when temperatures surpass the relatively low thermal optimum (T_{opt}) of around 25°C in most C_3 crops, whereas C_4 crops typically show a T_{opt} of about 35°C (*9*, *10*). Beyond a critical leaf temperature threshold, typically in the range of 40° to 45°C, photosynthetic rate drops sharply in both C_3 and C_4 crops (Fig. 1). Yet as outlined above, days with these air temperatures will become far more common over the next two to three decades. This underscores the urgent need to develop heat-resilient crops through a combination of physiological insights and breeding or bioengineering strategies to meet present and growing global demands for food, feed, fiber, bioenergy, and other bioproducts (*11*). Recent advancements provide promising avenues that could mitigate the impacts of high temperatures on photosynthesis, at least to some extent, in order to avoid substantial yield losses.

Plant energy balance and heating

Plants are efficient at absorbing solar radiation, which is foundational for supplying the energy needed for photosynthesis. However, in full sunlight, the energy absorbed by leaf chlorophylls is in excess of that needed to drive photochemistry (*12*), and this excess must be dissipated to minimize temperature elevation and photoinhibition (*13*, *14*). The major fates of absorbed energy (*15*) are loss through evaporation of water (latent heat loss), convection (sensible heat loss), and radiation (thermal heat loss) (Fig. 2). Latent heat transfer from transpiration serves to lower leaf temperature. If the rate of latent cooling is inadequate to balance the absorbed radiant and thermal energy, leaves will warm. Convective and radiative heat losses from the leaf to the surrounding environment will also lower leaf temperature, but unlike transpiration, these cannot, during conditions when photosynthesis occurs, lower leaf temperature below that of the ambient air.

Plants experience extreme heat stress when both air temperature and solar radiation are high. This is exacerbated by drought, which limits transpiration and latent heat cooling. Even when soil moisture is not limiting, stomatal opening is negatively correlated with atmospheric vapor pressure deficit (VPD). VPD is a measure of how "dry" the air is relative to its maximum capacity for holding moisture. As temperature increases, the moisture-holding capacity of air rises exponentially, which can be predicted from the thermodynamic properties of air. The large increase in VPD that can result from a temperature increase is illustrated in the following example. At a leaf temperature of 25°C and 60% relative humidity in the surrounding air, the water vapor pressure of that air (e_a) will be 1.90 kPa and the saturation vapor pressure (e_s) 3.17 kPa. The difference between the two is VPD ($e_s - e_a$), in this case, 1.27 kPa, which quantifies the drying power of the air. If leaf temperature is now increased to 35°C and e_a remains constant at 1.90 kPa, VPD nearly triples (3.72 kPa). In practice, warming also elevates the humidity of the air (e_a) such that relative humidity might remain constant. Even so, the VPD at 35°C would still be 2.25 kPa and almost double its value at 25°C. This sharp rise in VPD doubles evaporative demand and simultaneously reduces stomatal conductance (g_s) (*16*). Prolonged exposure to high VPD stress slows cell expansion, reduces leaf area, and decreases stomatal aperture and density, compounding limitations on CO_2 uptake and lowering water-use efficiency (WUE) and overall plant productivity (*17*, *18*). These are therefore indirect effects of rising temperature that limit photosynthesis and potentially yield.

The strong coupling between VPD and temperature makes it difficult to isolate their individual effects on stomatal conductance as temperatures rise. As VPD increases, a continued lowering of g_s would potentially lead to complete stomatal closure. However, studies in which temperature is increased under constant VPD show that the

[1]Global Change and Photosynthesis Research Unit, USDA-ARS, Urbana, IL, USA. [2]Department of Crop Sciences, University of Illinois Urbana-Champaign, Urbana, IL, USA. [3]Department of Plant Biology, University of Illinois Urbana-Champaign, Urbana, IL, USA. [4]Carl R. Woese Institute of Genomic Biology, University of Illinois Urbana-Champaign, Urbana, IL, USA. ***Corresponding author. Email: bernacch@illinois.edu (C.J.B.); slong@illinois.edu (S.P.L.); d-ort@illinois.edu (D.R.O.)** †These authors contributed equally to this work.

VPD-g_s relationship can uncouple at extreme high temperatures (*19*), allowing increased transpiration that may prevent temperatures lethal to photosynthesis. However, if soil water is scarce or if plant hydraulic conductivity is insufficient, this potential for evaporative cooling diminishes. Field and controlled-environment studies indicate that many major crops fail to sufficiently enhance hydraulic conductivity at high temperatures, limiting their ability to meet rising evaporative demand (*20*). Variability in hydraulic conductance within the germplasm of major crops has not been extensively evaluated but could provide an important avenue to improving capacity for latent heat cooling.

Rather than dissipate absorbed energy, adaptation in some plants has involved decreasing the amount of solar energy absorbed. Because leaves receive more energy in full sunlight than they can use in photosynthesis, reflecting more light does not affect carbon gain yet will both cool the leaf and lower photoinhibiton and photodamage to photosynthetic capacity. Several leaf properties can increase reflectivity, including surface hairs (*21*), surface waxes (*22*), and leaf chlorophyll content. Given variability in each within crop germplasm, these properties could all be used in breeding crops with more reflective leaves. Altered leaf angles may also decrease radiation interception around solar noon. Leaves that are more vertical intercept less direct radiation when solar elevation is high, decreasing thermal load around the warmest time of the day (*23*) (Fig. 2). A more vertically oriented canopy decreases the total direct solar radiation incident upon leaf surfaces (*24*). A further mechanism for achieving reduced light absorption, particularly during water shortage, is paraheliotropism, which lowers light interception by dynamic changes in leaf orientation. This change in leaf orientation occurs through the action of the pulvinus, an enlarged section at the base of a leaf petiole (stalk) that causes the leaf to move as it swells or shrinks according to its water content. For example, soybean leaflets move under the control of such pulvini. In a soybean cultivar, strong paraheliotropic movement during mild drought and high sunlight was shown to significantly lower leaf temperature, transpiration, and water stress (*25*).

Photosynthetic gas-exchange responses to high temperature

Measured net leaf CO_2 uptake (A) reflects the balance between photosynthetic CO_2 uptake and CO_2 released from mitochondrial respiration and photorespiration. There is some plasticity in temperature tolerance resulting from crop growth temperature. Both the T_{opt} and thermal maximum (T_{max}) for A can increase with a period of growth at higher temperatures. This can be substantial for T_{opt} (>10°C) in evergreen perennials but is small in C_4 species and in C_3 annuals, which includes most food crops (*9*). What underlies reduction in A at temperatures above T_{opt}?

Within the chloroplast, there is a wide array of photosynthetic processes that are temperature sensitive (*26*, *27*). In addition to increases in membrane fluidity, the photosynthetic apparatus and pigments are affected by the increased generation of reactive oxygen species above T_{opt}. Inhibition of oxygen evolution as well as photosystem I function can occur at high temperatures, but well above T_{opt}. However, field experiments and controlled environment studies (Fig. 3) implicate carbon metabolism as the most important and physiologically meaningful cause of high-temperature inhibition of photosynthesis, and this is therefore the focus of this and the following sections (*9*, *28*).

The probability of an oxygenation event at the active site for the enzyme Rubisco (see Box 1), and therefore of photorespiration, increases with temperature. Based on the conserved kinetics of Rubisco, it can be predicted that the loss of photosynthetic carbon gain to photorespiration is 28% at 25°C but 48% at 35°C in C_3 plants. The rate

Box 1. Explanation of major terms and abbreviations.

A The net rate of CO_2 uptake per unit of leaf area

CBB cycle Calvin-Bassham-Benson cycle. Carboxylation of the five-carbon RuBP results in two molecules of the C3 compound glycerate-3-phosphate (G3P), which are reduced to glyceraldehyde-3-phosphate. This triose phosphate is then cycled through a series of isomerase-, bisphosphatase-, transketolase-, and phosphorylase-catalyzed reactions to regenerate RuBP. The cycle is autocatalytic, potentially releasing one triose phosphate, for every three carboxylations, for onward synthesis of all organic constituents of the plant and substrates for respiratory metabolism.

C_3 Refers to plants in which Rubisco and the CBB cycle are in all green cells of the leaf and the first product of CO_2 assimilation is the C_3 compound G3P. Most crops are C_3 plants, including all woody crops, rice, wheat, brassicas, and legumes.

C_4 Refers to plants in which Rubisco and most of the CBB cycle is confined to chloroplasts within the bundle sheath—large cells that surround the vascular bundles. These in turn are surrounded by mesophyll cells with chloroplasts that lack Rubisco. CO_2 is first assimilated into the C_4 compound oxaloacetate, a dicarboxylate, within the mesophyll via the carboxylation of phospho-enol-pyruvate (PEP) catalyzed by PEP carboxylase. The resulting dicarboxylates diffuse to the Rubisco-containing bundle sheath cells, where they are decarboxylated to release CO_2 and pyruvate. An impermeable barrier minimizes CO_2 diffusion back to the mesophyll. The pyruvate diffuses back to the mesophyll via plasmodesmata, where it is phosphorylated to PEP, completing the C_4 photosynthetic cycle. Compared with C_3 photosynthesis, C_4 requires an extra two ATPs per CO_2 assimilated. However, this cost is typically offset by the fact that CO_2 is concentrated in the bundle sheath to a level that competitively inhibits the oxygenase reaction of Rubisco, all but eliminating photorespiration. Only a few crops are C_4, notably maize, pearl millet, sorghum, sugarcane, grain amaranths, most tropical pasture grasses, and miscanthus, a biomass feedstock.

Photorespiration Begins when Rubisco catalyzes the oxygenation of RuBP, which produces one molecule of G3P and one of 2-phosphoglycerate (2-PG). 2-PG is metabolized through a multiorganelle pathway to G3P, which reenters the CBB cycle. This is at the cost of one CO_2 emitted for every two 2-PG molecules with the consumption of four NADPH and seven ATP. The process from oxygenation to recovery of G3P is termed photorespiration, owing to its analogy to respiration in consuming oxygen and releasing CO_2; however, unlike respiration, it consumes rather than produces ATP and NADPH. It imposes a substantial penalty on A.

Rca Rubisco activase is essential to maintaining the activity of Rubisco. In the dark, inhibitory sugar phosphates inactivate Rubisco by occupying the enzyme active site. On illumination, Rca in a multimeric form uses energy from ATP hydrolysis to structurally remodel Rubisco to release the inhibitory sugar phosphates. This activity is also important throughout the course of the day to prevent the inhibition of Rubisco activity.

RuBP Ribulose-1:5-bisphosphate

Rubisco RuBP carboxylase/oxygenase. All CO_2 assimilated by plants is through the carboxylation of RuBP, which is catalyzed by this enzyme and forms the first step of the CBB cycle. However, the enzyme also catalyzes the oxygenation of RuBP, leading to photorespiration.

T_{opt} The leaf temperature at which light-saturated A is maximal

T_{max} The leaf temperature at which irreversible damage to photosynthesis begins to occur

of carboxylation is not a direct factor affecting A if ribulose-1:5-bisphosphate (RuBP) regeneration is limiting. However, there is still a reduction in A, albeit smaller, when RuBP regeneration is limiting, because increased photorespiration consumes adenosine triphosphate (ATP) and NADPH (reduced form of nicotinamide adenine dinucleotide phosphate) that would otherwise power RuBP regeneration. The probability of oxygenation of RuBP increases with temperature because the solubility of CO_2 in water (relative to that of O_2) decreases, and the activation energy requirement of the oxygenation reaction is greater than that of carboxylation, both of which cause increased photorespiration (*7*). This largely explains the higher temperature optimum of C_4 photosynthesis, in which photorespiration is minimal. The optimum temperature of A in C_3 plants increases from 25°C in normal air, of about 400 parts per million (ppm) [CO_2], to 35°C at 800 ppm [CO_2], a similar optimum to that of C_4 photosynthesis resulting from suppression of photorespiration (*7*).

Growth at elevated temperature can produce both damaging and protective effects on photosynthesis. Increasing exposure time to heat stress exacerbates the negative effects of instantaneous heat stress. High temperatures deactivate enzymes, but with continued exposure, these can become irreversibly denatured. Membrane fluidity increases with temperature, and with prolonged exposure, leakage of metabolites and inorganic ions will occur, as well as disruption of membrane organization. As the time at elevated temperature is prolonged, these changes result in a steady decline in photosynthesis and, in turn, productivity while limiting the potential for any recovery. Secondary effects also result. Because high temperatures typically coincide with high light, impairment of photosynthetic capacity and its protective mechanisms makes the apparatus more vulnerable to photodamage, particularly a loss of the labile D1 protein in the photosystem II core (*29*). If the temperature increase is gradual, then protective mechanisms can allow photosynthesis to continue at higher temperatures. In general, these mechanisms raise the temperature tolerance of photosynthesis by an extra 2° to 3°C. There are three major mechanisms by which photosynthesis can acclimate to high temperature. (i) Several photosynthetic proteins are coded by gene families or result from alternative splicing, where more thermotolerant isoforms are expressed at elevated temperatures. Rubisco activase (Rca) is one of the best-known examples (*30*). Altered patterns of gene expression can result in increased amounts of proteins that would otherwise limit photosynthesis at high temperature (*31*). This may include a rebalancing of investment in Rubisco versus RuBP regeneration, because RuBP regeneration limitation increases with temperature (*28*). (ii) Heat shock proteins (HSPs), which facilitate the correct folding of photosynthetic proteins, can also prevent enzyme inactivation, denaturation, and aggregation. The heat shock transcription factor (HSF), recognized by a conserved region on the promoters of HSP genes, is elevated with growth at high temperature, resulting in increased protection (*32*). (iii) Remodeling of the lipid composition of the thylakoid membranes by increasing diglyceride content counteracts increased fluidity that would otherwise allow ion leakage and disruption of membrane processes (*33*).

The temperature sensitivity of Rubisco catalytic efficiency and specificity

In addition to photorespiration in C_3 crops increasing with rising temperature, the Rubisco activation state declines, further contributing to decreased A and underlying the sharp drop in photosynthesis observed at about >40°C (*34*, *35*) (Fig. 1). Rubisco is deactivated by the binding of inhibitory sugar phosphates, which cannot dissociate from its catalytic sites without intervention by Rca (*36*). Temperatures above T_{opt} increase the rate of Rubisco inactivation owing to higher concentrations of inhibitory compounds. In addition, Rca is often the most thermolabile photosynthetic protein, leading to inactivation at temperatures often at or slightly above T_{opt} (*37*) and below temperatures that affect other

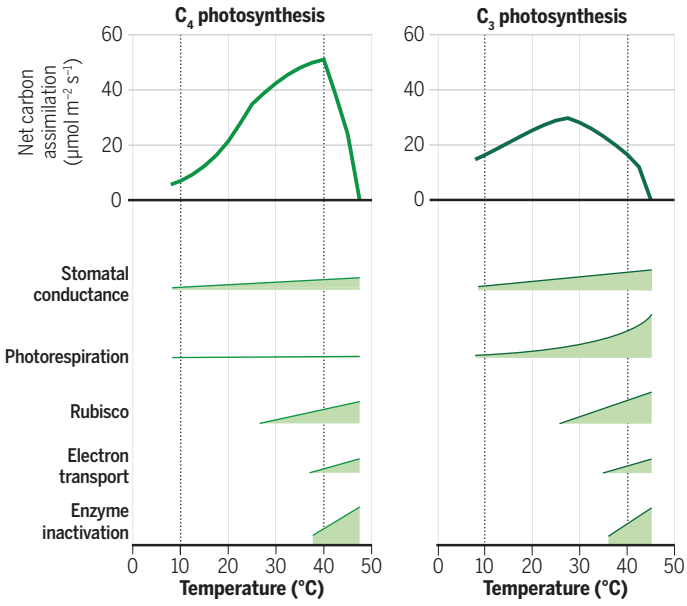


Fig. 1. Modeled temperature responses of leaf-level net carbon assimilation for C_4 and C_3 photosynthetic pathways. Graphs were created using a C_4 (*8*) and a C_3 (*76*) gas-exchange model but modified to include a rapid loss of enzymatic function above 42°C—an effect usually omitted from leaf gas-exchange models. For C_3 species, as leaf temperature increases beyond T_{opt}, photosynthesis declines while photorespiration and dark respiration increase, whereas for C_4 species, there is no photorespiration, resulting in a high T_{opt}. The average T_{opt} for net CO_2 uptake (A) is around 35°C for C_4 and 25°C for C_3 plants. Average high-temperature tolerance of photosynthesis (T_{max}), that is, the temperature at which complete irreversible damage occurs, is about 50°C for tropical plants and 42°C for plants from high latitudes (*77*). However, these vary with adaptation to the area of origin. For example, the C_4 grass *Spartina anglica* in Britain has a T_{opt} of 30°C compared with a T_{opt} of 47°C for the C_4 shrub *T. oblongifolia* in Death Valley, California (*78*, *79*). Below each graph, semiquantitative schematics depict the temperature sensitivity of key processes—stomatal conductance, photorespiration, Rubisco activity, electron transport, and enzyme inactivation. The width of each shape is aligned to the temperature axis and denotes the relative contribution of that process to limiting photosynthesis as temperature increases.

enzymes and chloroplast membrane processes. The thermolability of Rca causes the proportion of inactive Rubisco to rise with temperature.

Most plants express two Rca isoforms: a smaller β isoform (41 to 43 kDa) and a larger α isoform (43 to 47 kDa), which has a C-terminal extension containing two redox-regulated cysteine residues (*38*). Redox regulation of the Rca-α isoform by chloroplast-localized thioredoxin connects regulation of Rubisco activity to the dynamic chloroplast redox potential that changes with environmental conditions, including light intensity and temperature. Solanaceae species are distinctive in that they express only the redox-insensitive β isoform. In the absence of redox regulation in Solanaceae, light regulation via Rca appears to be mediated by changes in the adenosine diphophaste (ADP)/ATP ratio of the chloroplast (*34*, *36*).

Beyond differences in the number of Rca genes and the isoforms they produce, there is notable variation across species in the expression and relative abundance of Rca isoforms and their response to temperature (*39*, *40*). In *Arabidopsis* (*Arabidopsis thaliana*), the Rca-α and -β isoforms originate from alternative splicing of a single gene and are present at the transcript and protein levels in approximately equal amounts; their ratio is not affected by temperature. In rice (*Oryza sativa* L.), the two Rca isoforms are also produced by alternative splicing in which Rca-β is at higher levels at permissive growth temperatures but heat stress results in increased expression of Rca-α (*41*).

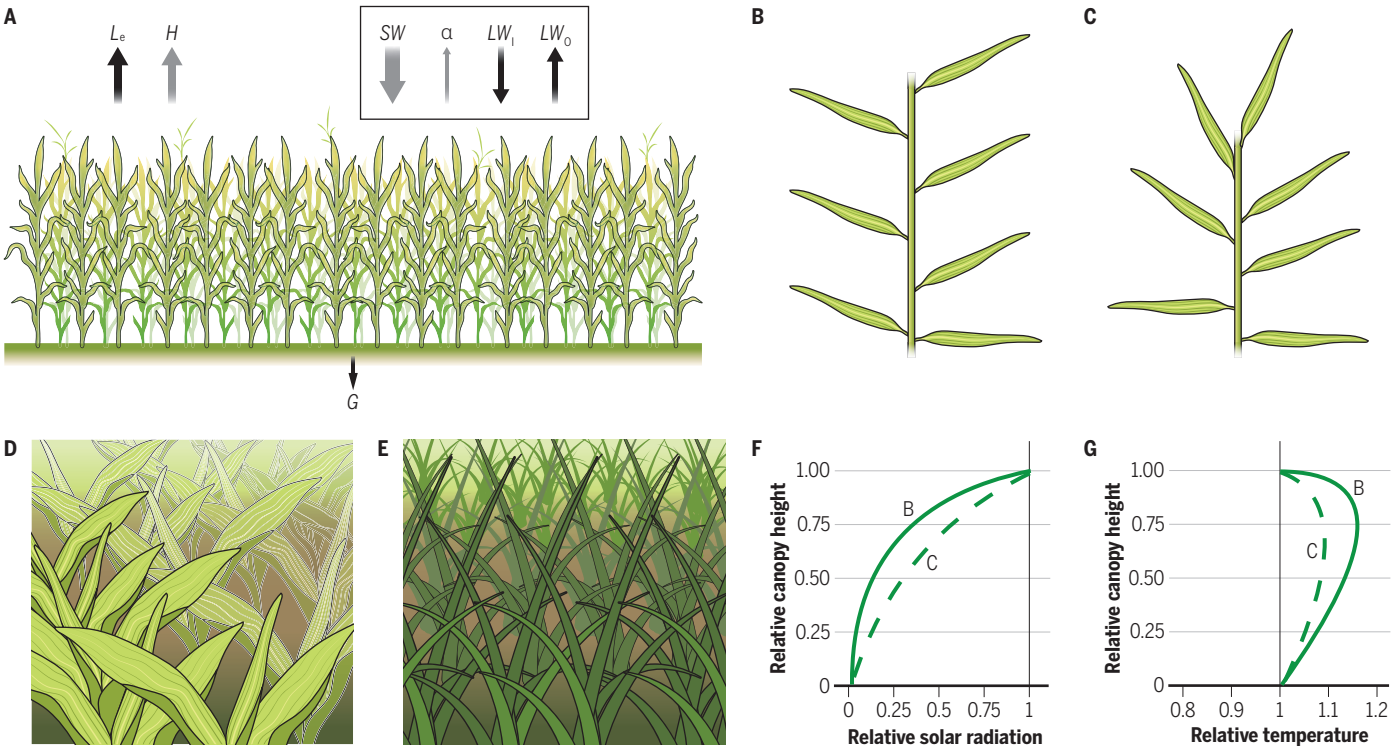


Fig. 2. Canopy energy balance and canopy architecture. (**A**) Energy budget representation of a plant canopy. The total available energy is determined by downwelling shortwave radiation from the sun (*SW*) minus the amount reflected (α) by the canopy, as well as by the difference between incoming (LW_I) and outgoing (LW_O) long-wave radiation. The balance of these radiative fluxes, outlined by a rectangular box in the figure, represents net radiation (R_n), which is the total potential energy available to a plant canopy (*15*). Most of this available energy is partitioned into latent (L_e) and sensible (*H*) heat fluxes, although a small amount of energy penetrates through the canopy into the soil (G) and a much smaller part is used in photosynthesis, typically <1%. The relative fluxes of *H* versus L_e are modulated by stomatal conductance—a higher conductance increases L_e and thus lowers *H*. Arrow widths represent the generalized magnitude of the fluxes for each component during midday, clear-sky conditions. (**B** to **G**) Idealized diagrams [(B) and (C)] and photographs [(D) and (E)] of horizontal [(B) and (D)] and vertical [(C) and (E)] plant architectures. Healthy (D) and heat and/or water-stressed (E) maize (*Z. mays*) plants at similar growth stages demonstrate adaptive responses that alter canopy energy balance under stress conditions. The relative solar radiation incident upon leaves (F) and the temperature of leaves (G) change from the canopy top (relative canopy height = 1) to the soil surface (relative canopy height = 0). The letters adjacent to the lines in (F) and (G) correspond to the respective canopy architecture panels [(B) and (C)]. Leaf rolling in stressed plants (E) creates a more vertical canopy orientation, changing the distribution of solar radiation within the canopy (F) and thereby lowering the heat load (G). In addition to modifying leaf orientation, plants use various strategies to influence their energy budgets, such as altering the leaf reflectivity using trichomes, waxes, or pigments and adjusting stomatal or boundary-layer conductances.

In the C_4 crops sorghum (*Sorghum bicolor*), sugarcane (*Saccharum officinarum*), and maize (*Zea mays* L.), two *Rca* genes are present, one each for Rca-α and Rca-β. Whereas the *Rca*-β gene is consistently expressed at normal growth temperatures, *Rca*-α expression is only induced at high temperatures (*40*). In maize, heat-induced expression of *Rca*-α occurs in seedlings but appears diminished or absent in mature plants (*42*, *43*). The induction profile of *Rca*-α expression mimics recovery of photosynthesis and the profile of Rubisco reactivation after high-temperature exposure. This association between Rca-α isoform expression and restoration of Rubisco activation at high temperature potentially supports a thermoprotective role of Rca-α in carbon fixation in C_4 grasses by sustaining Rubisco activation at high temperature. Wheat (*Triticum aestivum* L.) differs yet again. Although wheat has two *Rca* genes, one codes for the β isoform and the other produces both an α and β isoform by alternative splicing (*44*), again with α increasing with high-temperature exposure.

Developing photosynthetically thermotolerant crops

Crop architecture and energy balance

Optimizing crop canopies to achieve more even light distribution has long been hypothesized to increase net daily carbon gain. For example, selecting or engineering canopies with upright leaves in the upper canopy and more horizontal leaves in the lower canopy can increase carbon gain (*45*) and potentially enhance resilience to high temperatures (*7*) (Fig. 2). Reducing leaf chlorophyll content has also been shown to increase light penetration into the canopy without sacrificing overall photosynthetic rates (*45*, *46*). In these scenarios, upper leaves—which normally absorb more light than they can use—maintain high photosynthesis with reduced absorption while lower leaves benefit from receiving more solar radiation, thereby increasing total canopy photosynthesis (*45*). More evenly distributed solar radiation throughout the canopy will also distribute heat loads more uniformly, potentially moderating leaf temperatures. However, such traits may involve trade-offs; for instance, altering leaf reflectivity or absorptive properties could reduce total light capture and photosynthesis in milder conditions. This might be addressed using high-temperature inducible promoters. Consequently, breeders and biotechnologists must balance the benefits of temperature resilience with the risk of reduced net photosynthesis when growing conditions are not heat-stressed. Highly mechanistic modeling tools to quantitatively evaluate these trade-offs are available (*47*). Additionally, high-throughput field phenotyping of three-dimensional form and spectral properties (*48*) (Fig. 3) coupled with resequencing of hundreds of genotypes (*49*) of a crop can allow identification of advantageous alleles. Once identified, alleles underlying canopy and reflectance properties

that will improve protection of photosynthesis at high temperature could be introgressed into elite cultivars, aided by molecular marker or genomic selection.

Improving water-use efficiency

Anticipated increases in VPD due to rising temperature places added pressure on crop water resources. Presently, 40% of global crop production is irrigated, accounting for ~71% of freshwater extraction. With diminishing water resources (*50*), improved crop WUE will be vital. In principle, rising [CO_2] allows plants to maintain the same, or higher, photosynthetic carbon assimilation rate with lower stomatal conductance (g_s). Although plants naturally reduce g_s under elevated [CO_2], the reduction is typically insufficient to realize the full potential WUE gains—particularly in C_4 crops, where photosynthesis is already close to [CO_2] saturation (*51*, *52*). To address this, researchers are exploring ways to further decrease g_s through breeding and genetic engineering with little or no penalty to carbon gain (*53*). Overexpression of photosystem II subunit S, for instance, can lower g_s at all light intensities to levels that reduce transpiration without affecting photosynthesis or productivity, resulting in a 30% decrease in whole-plant water use (*54*). Another approach is reducing stomatal density, that is, the number of stomata per unit leaf area. In C_4 sorghum, a moderate reduction in stomatal density through the insertion of a synthetic epidermal patterning factor transgene lowered plant water use by ~15% with no adverse effect on photosynthesis (*55*). In addition to transgenic strategies, there is substantial natural variation in stomatal density within many crop germplasms (e.g., a 2.5-fold range across 235 rice accessions), suggesting that reduction could be achieved through breeding (*56*).

Although improving WUE through lower g_s may be useful in maximizing water availability (*55*), the lower transpirational cooling can also increase the likelihood of extreme leaf temperatures. This may be exacerbated in elevated [CO_2] through further decreases in g_s (*57*). However, the potential to lower water use with combined optimization of canopy structure and surface-energy balance can potentially overcome this hypothesized positive feedback (*47*). It will therefore be important to assess for each crop and region whether this promising approach to improving WUE risks temperature damage and how that might be alleviated by stacking traits. For example, combining altered canopy structure and reflectance with decreased transpiration has been predicted to improve WUE without elevating leaf temperature and impairing photosynthesis (*47*).

Leaf photosynthetic physiology

Natural variability in Rubisco specificity for CO_2 relative to O_2 and catalytic parameters also offer potential gains for photosynthetic efficiency and thermotolerance (*58*). High-throughput measurement of Rubisco kinetics has shown the potential to replace present crop Rubisco with faster or higher-specificity forms from other species that could boost photosynthetic rates under warm conditions and rising [CO_2] (*58*). Yet cross-species Rubisco substitution requires matching all necessary assembly and chaperone proteins—an ongoing challenge that has not yet been fully resolved (Table 1). Even though the many remaining mysteries and uncertainties concerning the regulation of Rubisco activity by Rca complicate the development of photosynthesis thermotolerance, the protection of Rubisco activity via more-thermotolerant Rca appears particularly promising (*59*).

As noted above, Rca is particularly thermolabile, and more so in crops adapted to cooler climates. Substituting Rca from a warm-climate species into a cooler-climate species has been shown to increase photosynthetic T_{opt} (*60*). An alternative approach involves manipulating Rca isoforms within a species. For instance, heat-induced expression of an Rca-α isoform in sorghum appears to confer a higher temperature tolerance of Rubisco activation (*41*). Because constitutive Rca-α expression might introduce a fitness cost at optimal temperatures, regulating its expression threshold or placing it under an inducible promoter may be advantageous. Overexpression of maize Rca in rice did not increase the Rubisco activation state or photosynthetic rate below 25°C but had a stimulatory effect at 40°C (*61*), suggesting a protective role of Rca overexpression on steady-state photosynthesis at high temperatures. Emerging evidence also suggests that a few amino acid changes can greatly improve Rca thermostability (*35*). However, the molecular basis for Rca thermolability varies across species and among isoforms. In some cases, thermolability may involve disruption of the multimeric Rca complex (*36*). Photosynthesis of *Tidestromia oblongifolia*, a native inhabitant of the floor of Death Valley, California, is clearly well adapted to high temperature (*62*), suggesting that understanding *Rca* sequence variations in this species and other desert plants may be especially informative. The diversity of Rca forms means that engineering robust, thermotolerant Rca will likely require species-specific or even cultivar-specific approaches.

Rising temperature and elevated [CO_2] can shift the primary limitation on C_3 photosynthesis from Rubisco activity to RuBP regeneration. Sedoheptulose-1,5-bisphosphatase (SBPase) is often a key bottleneck in RuBP regeneration via the CBBc (see Box 1). An open-air replicated field experiment (Fig. 3) tested soybeans with transgenic up-regulation of SBPase alongside wild-type controls under elevated temperature and [CO_2] singly and in combination (*63*). Although higher temperature reduced yield in both ambient and elevated [CO_2] conditions, the SBPase-transgenic lines maintained significantly higher yields than wild type under combined heat and elevated [CO_2], effectively matching wild-type yields in ambient conditions. Other transgenic approaches have also demonstrated potential in field settings. For example, heat stress–induced overexpression of the D1 protein significantly increased biomass and grain yield in field-grown rice (*29*).

Installing a photorespiratory "bypass" in C_3 crops provides another strategy to mitigate the effects of higher temperatures on photosynthetic efficiency (*59*, *64*). Several photorespiratory bypass designs have been proposed to recycle 2-phosphoglycolate (2-PG) with lower requirements for ATP and NADPH than in the native pathway (*65*). They generally rely on metabolizing 2-PG within the chloroplast and with fewer metabolic reactions. High-temperature field experiments (Fig. 3) using tobacco (*Nicotiana tabacum* L.) engineered with one such alternative photorespiratory pathway showed higher net photosynthetic CO_2 uptake and a 26% increase in biomass under season-long elevated temperatures (5°C above nonheated control plots) compared with wild-type plants (*66*). Similarly, genetically modified potato (*Solanum tuberosum*) expressing a bypass pathway showed increased photosynthetic capacity and daily carbon assimilation during naturally occurring heat waves and a 30% increase in tuber biomass relative to wild type (*67*). These findings support theoretical predictions that rising temperature amplifies photorespiratory losses and highlights the potential for bypass strategies to sustain or improve crop yields in a warming climate.

Photorespiration in C_3 crops could be largely eliminated by conversion to the C_4 form. A C_3-to-C_4 conversion would require C_4 compartmentation of photosynthetic enzymes and the formation of a diffusive barrier between the mesophyll and bundle sheath cells. Although this clearly requires multiple genetic changes, it is notable that nature has achieved this transition independently almost 70 times. The past two decades have seen great progress in understanding the molecular basis of what makes a C_4 leaf and in installing parts of the system into C_3 rice (*68*). This conversion would not only increase photosynthesis and WUE at all temperatures but also mean that carbon gain above 25°C would increase, rather than decrease, with temperature rise to a T_{opt} of 35°C (Fig. 1). A possibly more tractable alternative would be to convert C_3 crops to the C_2 form (*68*). C_2 photosynthesis is considered to represent evolutionary transition points between C_3 and C_4. In C_2 plants where photorespiratory CO_2 release is confined to the bundle sheath, there can be substantial recapture of this CO_2 in

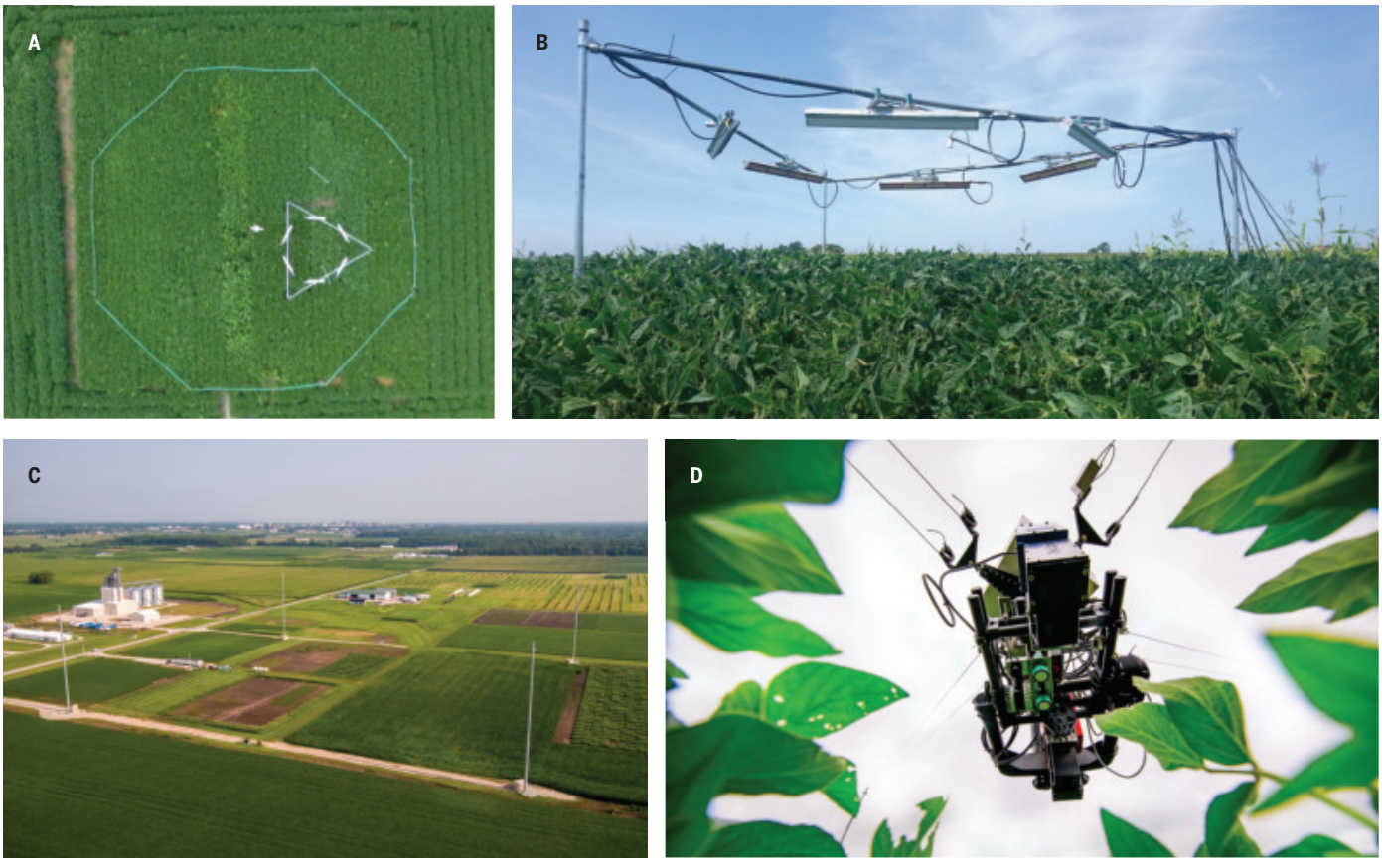


Fig. 3. Experimental techniques for understanding high-temperature impacts and tolerance in crop germplasm in farm fields. Aerial view within an elevated CO_2 plot (**A**) and lateral view (**B**) of an in-field infrared heating array used to simulate warmer growing conditions for crops at the Soybean Free Air Concentration Enrichment facility (SoyFACE) in Urbana, Illinois. The electrical current to six infrared heaters is modulated to maintain a set point canopy temperature above a nonheated reference area adjacent to the heated area, as determined from thermal imaging of the canopy. As temperatures fluctuate over the reference plot, output from the heaters is adjusted to match the target temperature increase for the heated plots. Various experiments have been undertaken to simulate global warming (*80*) and heat waves (*81*) and to test strategies to genetically improve crop resilience to temperatures (*63*,*66*). (**C**) Aerial view of the 4-Ha RIPE Aerial Plant Phenotyping System (RAPPS) located on the Energy Farm of the University of Illinois Urbana-Champaign and (**D**) a view of the sensor package, including hyperspectral imagers, light detecting and ranging (lidar), thermal photography, and RGB photography sensors, mounted on the dolly. RAPPS can move the sensor package over the 4 Ha of farmland to provide ultra-high-resolution, repeatable, semiautonomous, rapid, and high-accuracy information on more than 100,000 individual plants. RAPPS is highly versatile and can provide high-throughput information to advance breeding efforts for improved thermal resilience in crops using natural variation in temperatures or, alternatively, by integrating high-temperature treatments [(A) and (B)] into the measurement footprint.

the surrounding mesophyll as well as increased concentration around Rubisco. It would require fewer changes than conversion to C_4 but would have smaller benefits. Engineering microbial CO_2-concentrating mechanisms (CCMs) into crop chloroplasts is another strategy to improve the thermal resilience of C_3 crops. Like C_4 photosynthesis, these serve to concentrate CO_2 at Rubisco, thereby minimizing photorespiratory losses. Many cyanobacteria concentrate CO_2 in carboxysomes, which are microcompartments that contain Rubisco, Rca, and carbonic anhydrase. It is estimated that this system would increase both photosynthesis and WUE by about 60% and increase T_{opt} by about 10°C. Although carboxysomes have been assembled in C_3 crop chloroplasts (*69*), installing all the necessary ancillary components has not been achieved as of yet. Installing the pyrenoid system found in the chloroplasts of many eukaryotic algae would have similar benefits (*70*). For all these CCMs, success depends on the discovery of the set of components needed to make the system successful in C_3 crops, making the time horizon for attaining this difficult to know (Table 1).

Conclusions

The projected temperature increase between 2010 and 2050 is estimated to depress yields of the major grains by 6 to 16%, against a backdrop of a potential >50% increase in demand over this period (*11*). Table 1 shows a range of examples from cellular to whole-crop canopy changes that could safeguard photosynthesis in our warming world. Except for CCMs, all technological changes needed are known and therefore achievable. The time frames given assume that the resources and personnel are available, and this assumption has present real-world barriers. Whether advantageous alleles within the crop germplasm, edits, or transgenes are considered, all will require introgression by crop breeders into elite lines adapted and locally acceptable for different regions. This is at a time when capacity for public domain plant breeding has become substantially diminished (*11*). Several of the traits that have been shown to improve photosynthetic temperature resilience are transgenic. The time taken for a new plant biotechnology-derived genetic trait to reach commercialization during the period from 2017 to 2022 was 16.5 years at a cost of $115.0 million per transgene (*71*). Without changes in regulatory frameworks coupled with social acceptance of transgenic crops, this will remain a major impedance to progress. Many countries have accepted or are considering accepting DNA-edited crops, where no foreign DNA has been added, without regulation beyond that required of conventionally bred crops (*72*). The improvements demonstrated to date of the first

Table 1. Traits for, benefits of, timeline for, and risks of increasing the temperature resilience of crops. Estimated improvements in temperature tolerance and the time taken to achieve it. Where "years to proof of concept" are 0, the trait or invention has either been demonstrated in a single-site field trial or shown in a detailed mechanistic model, with reference number provided. Other time estimates are based on the authors' understanding of the state-of-the-art technology. "Years to farmers' fields" are minima and assume uninterrupted passage and use of winter nurseries for temperate crops. In the case of breeding, it assumes identification of the advantaged alleles and/or loci in year 1, hybridization with elites in year 2, and then three rounds of backcrossing per year and multiplication of resulting improved elite germplasm in years 4 and 5 for delivery to seed systems. Editing assumes that mutations that up-regulate (or, in the case of chlorophyll, down-regulate) expression are found and that these are then introgressed into elite cultivars, as is done for the use of natural variation. Time here will be strongly dependent on the evolving regulations around edited material. For transgenic plants, the time from discovery through development to authorization of a new plant biotechnology-derived trait for cultivation is estimated at 16 years, a number based on a survey of the major companies that produce transgenic food crops (*71*).

Trait or invention	Crop type	Predicted increase in photosynthetic thermotolerance (°C)	Years to proof of concept	Years to farmers' fields	Risks and notes
Up-regulation of SbPase	C_3	5	0 (*63*)	5*, 8†, 16‡	Sufficient variation within crop germplasm*; edit found that will up-regulate expression†
Up-regulation of Rca	$C_3(C_4)$§	3	0 (*61*)	5*, 8†, 16‡	As above
Up-regulation of HSPs	C_3 and C_4	3	0 (*32*)	5*, 8†, 16‡	As above
Increase leaf reflectivity	C_3 and C_4	5	0 (*47*)	5*, 8†, 16‡	As above
Decrease leaf chlorophyll	C_3 and C_4	3	0 (*47*)	5*, 8†, 16‡	As above
More vertical leaves	C_3 and C_4	3	0 (*47*)	5*, 8†, 16‡	As above; increased verticality has already been explored in cereals, which may leave little room for further change
Photorespiratory bypass	C_3	5	0 (*66*)	16‡	Already shown to protect yield at elevated temperature, without apparent detriment, despite decreased metabolic flux through the native photorespiratory pathway
Edited Rca	$C_3(C_4)$§	5	3	11†	Depends on advantageous edits being identified
Transplant thermally adapted Rca	$C_3(C_4)$§	10	5	20‡	Foreign Rca may not effectively bind native Rubisco
Transplant thermally adapted Rubisco	C_3 and C_4	10	15	30‡	Requires effective transformation of plastid and nucleus and effective binding to native Rca
Rubisco edited to obtain more thermostable forms	C_3 and C_4	10	3	11†	Depends on advantageous edits being identified
Add the carboxysome system	C_3	10	Uncertain¶	Uncertain¶	Would also substantially increase efficiency of light, water, and nitrogen use
Add the pyrenoid system	C_3	10	Uncertain¶	Uncertain¶	As above
Convert C_3 to C_4	C_3	10	Uncertain¶	Uncertain¶	As above; this transition has occurred in nature almost 70 times
Convert C_3 to C_2	C_3	5	Uncertain¶	Uncertain¶	As above, but likely to require fewer genes than conversion to C_4

*Alleles that improve thermotolerance identified and introgressed into elite cultivars. †DNA editing of the upstream region of a gene to increase expression. ‡Transgenic expression of a foreign gene or genes. §$C_3(C_4)$ indicates that the trait or invention is of benefit to C_3 and possibly of benefit to C_4. ¶Uncertain because more discovery is needed to determine the minimum set of genes required for this transition.

five entries in Table 1 have all concerned transgenic up-regulation or suppression of gene expression. These, however, concern genes already present in the crop. Increasingly, these changes could likely be achieved by editing the upstream noncoding region of genes, thereby producing the desired phenotype without the addition of foreign DNA (*57*). This could greatly decrease the time needed to move innovations to seed systems and farms (Table 1). New technologies, particularly those enabled by artificial intelligence, from high-throughput phenotyping to reconfiguration of key proteins coupled with DNA editing, offer hope that barriers can be reduced.

If alleles conferring improved thermal tolerance of photosynthesis are identified within the germplasm of a crop and its interfertile relatives, then these can be introduced into elite lines in a relatively short period of time (Table 1). The discovery of the rice *Sub1-1A* allele that allows rice to survive submergence with increasing flooding events and its introgression into a wide range of rice cultivars that were rapidly adopted by some of the world's poorest farmers shows how successful this approach can be in tackling climate change impacts (*73*). Until now, identifying alleles or loci that provide improved temperature tolerance of photosynthesis from hundreds or thousands of accessions or tilling populations of a crop would have been a huge undertaking, but new technologies offer a means to achieve this quickly. Full sequences of hundreds of genotypes of major crops are now becoming available at an ever-increasing pace. How might this be used in identifying alleles that could be used in adapting photosynthesis to higher temperatures? It would be challenging to screen large amounts of germplasm under a controlled elevation of temperature; however, high-throughput tools (Fig. 3) could be used to screen thousands of genotypes in the field under natural variations in temperatures, including under the increasingly frequent high-temperature events. High-throughput techniques of crops such as solar-induced fluorescence (*74*) and hyperspectral imaging to estimate different photosynthetic parameters, including Rubisco activity in vivo (*75*), facilitate the application of genome-wide association analyses to identify relevant loci and advantageous alleles that could then be introgressed into elite cultivars for improved photosynthetic temperature tolerance. Success here will depend on the existence of tolerance within the germplasm.

As highlighted here (Table 1), there are many opportunities to safeguard crop photosynthesis in a rapidly warming world. These are all technologically feasible. Whether they are achieved will depend, as with other adaptations in the food supply system, on public-domain commitment and investment.

REFERENCES AND NOTES

1. World Meteorological Organization, "State of the global climate 2024" (WMO-No. 1368, WMO, 2024).
2. Intergovernmental Panel on Climate Change (IPCC), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, V. Masson-Delmotte *et al*, Eds. (Cambridge Univ. Press, 2021).
3. D. I. V. Domeisen *et al.*, *Nat. Rev. Earth Environ.* **4**, 36–50 (2022).
4. T. Hasegawa *et al.*, *Nat. Food* **2**, 587–595 (2021).
5. J. Berry, O. Bjorkman, *Annu. Rev. Plant Physiol.* **31**, 491–543 (1980).
6. E. A. Ainsworth, S. P. Long, *Glob. Change Biol.* **27**, 27–49 (2021).
7. C. E. Moore *et al.*, *J. Exp. Bot.* **72**, 2822–2844 (2021).
8. S. von Caemmerer, *J. Exp. Bot.* **72**, 6003–6017 (2021).
9. W. Yamori, K. Hikosaka, D. A. Way, *Photosynth. Res.* **119**, 101–117 (2014).
10. R. F. Sage, D. S. Kubien, *Plant Cell Environ.* **30**, 1086–1106 (2007).
11. S. P. Long, *Philos. Trans. R. Soc. London Ser. B* **380**, 20240229 (2025).
12. X. G. Zhu, S. P. Long, D. R. Ort, *Annu. Rev. Plant Biol.* **61**, 235–261 (2010).
13. N. Zahra *et al.*, *Environ. Exp. Bot.* **206**, 105178 (2023).
14. R. M. Marchin *et al.*, *Glob. Change Biol.* **28**, 1133–1146 (2022).
15. G. S. Campbell, J. M. Norman, *An Introduction to Environmental Biophysics* (Springer, 2000).
16. J. T. Ball, I. E. Woodrow, J. A. Berry, in *Progress in Photosynthesis Research: Volume 4 Proceedings of the VIIth International Congress on Photosynthesis Providence, Rhode Island, USA, August 10–15, 1986*, J. Biggins, Ed. (Springer, 1987), pp. 221–224.
17. J. López, D. A. Way, W. Sadok, *Glob. Change Biol.* **27**, 1704–1720 (2021).
18. C. Grossiord *et al.*, *New Phytol.* **226**, 1550–1566 (2020).
19. J. Urban, M. W. Ingwers, M. A. McGuire, R. O. Teskey, *J. Exp. Bot.* **68**, 1757–1767 (2017).
20. A. M. Locke, L. Sack, C. J. Bernacchi, D. R. Ort, *Ann. Bot.* **112**, 911–918 (2013).
21. C. P. Bickford, *Funct. Plant Biol.* **43**, 807–814 (2016).
22. H. D. R. Carvalho, J. L. Heilman, K. J. McInnes, W. L. Rooney, K. L. Lewis, *Agric. For. Meteorol.* **284**, 107893 (2020).
23. X. Yang *et al.*, *Ecol. Lett.* **26**, 1005–1020 (2023).
24. A. Kadioglu, R. Terzi, *Bot. Rev.* **73**, 290–302 (2007).
25. V. S. Berg, S. Heuchelin, *Crop Sci.* **30**, 631–638 (1990).
26. R. A. Slattery, D. R. Ort, *Plant Cell Environ.* **42**, 2750–2758 (2019).
27. X. G. Zhu *et al.*, *Front. Plant Sci.* **13**, 967203 (2022).
28. C. J. Bernacchi, U. M. Ruiz-Vera, M. H. Siebers, N. J. DeLucia, D. R. Ort, *Biochem. J.* **480**, 999–1014 (2023).
29. J. H. Chen *et al.*, *Nat. Plants* **6**, 570–580 (2020).
30. G. E. Degen, D. J. Orr, E. Carmo-Silva, *New Phytol.* **229**, 1298–1311 (2021).
31. L. Z. Huang, M. Zhou, Y. F. Ding, C. Zhu, *Int. J. Mol. Sci.* **23**, 11970 (2022).
32. S. Hu, Y. Ding, C. Zhu, *Front. Plant Sci.* **11**, 375 (2020).
33. X. Zhang *et al.*, *Plant Cell Environ.* **48**, 3391–3405 (2025).
34. I. Wijewardene, G. Shen, H. Zhang, *Stress Biol.* **1**, 2 (2021).
35. A. P. Scafaro, N. Bautsoens, B. den Boer, J. Van Rie, A. Gallé, *Plant Physiol.* **181**, 43–54 (2019).
36. I. Sparrow-Muñoz, T. C. Chen, S. J. Burgess, *Biochem. Soc. Trans.* **51**, 627–637 (2023).
37. A. P. Scafaro, B. C. Posch, J. R. Evans, G. D. Farquhar, O. K. Atkin, *Nat. Commun.* **14**, 2820 (2023).
38. J. A. Perdomo, J. C. Scales, W. S. Lee, K. Kanyuka, E. Carmo-Silva, *Plant Direct* **8**, e583 (2024).
39. G. E. Degen, D. Worrall, E. Carmo-Silva, *Plant J.* **103**, 742–751 (2020).
40. J. A. Perdomo, P. Buchner, E. Carmo-Silva, *Photosynth. Res.* **148**, 47–56 (2021).
41. S. Y. Kim, R. A. Slattery, D. R. Ort, *Glob. Change Biol. Bioenergy* **13**, 211–223 (2020).
42. S. C. Stainbrook *et al.*, *Biosci. Rep.* **44**, BSR20240353 (2024).
43. Z. Ristic *et al.*, *J. Exp. Bot.* **60**, 4003–4014 (2009).
44. R. Nagarajan, K. S. Gill, *Plant Mol. Biol.* **96**, 69–87 (2018).
45. R. A. Slattery, D. R. Ort, *Plant Physiol.* **185**, 34–48 (2021).
46. B. J. Walker *et al.*, *Plant Physiol.* **176**, 1215–1232 (2018).
47. D. T. Drewry, P. Kumar, S. P. Long, *Glob. Change Biol.* **20**, 1955–1967 (2014).
48. P. Fu, K. Meacham-Hensold, M. Siebers, B. Feddersen, C. Bernacchi, *Authorea* essoar.10508739.2 [Preprint] (2021); https://doi.org/10.1002/essoar.10508739.2.
49. B. Song *et al.*, *Mol. Plant* **16**, 1252–1268 (2023).
50. K. A. Novick *et al.*, *Nat. Clim. Change* **6**, 1023–1027 (2016).
51. C. E. Salesse-Smith, Y. Wang, S. P. Long, *New Phytol.* **245**, 951–965 (2025).
52. A. Srivastava, V. Srinivasan, S. P. Long, *Plant Cell Environ.* **47**, 1716–1731 (2024).
53. T. Lawson, A. D. B. Leakey, *J. Exp. Bot.* **75**, 6677–6682 (2024).
54. B. Turc *et al.*, *J. Exp. Bot.* **75**, 3959–3972 (2024).
55. J. N. Ferguson *et al.*, *J. Exp. Bot.* **75**, 6823–6836 (2024).
56. W. Phetluan *et al.*, *Plant Sci.* **330**, 111624 (2023).
57. C. J. Bernacchi, B. A. Kimball, D. R. Quarles, S. P. Long, D. R. Ort, *Plant Physiol.* **143**, 134–144 (2007).
58. D. J. Orr *et al.*, *Plant Physiol.* **172**, 707–717 (2016).
59. R. Croce *et al.*, *Plant Cell* **36**, 3944–3973 (2024).
60. A. E. Carmo-Silva, M. E. Salvucci, *Photosynth. Res.* **108**, 143–155 (2011).
61. W. Yamori, C. Masumoto, H. Fukayama, A. Makino, *Plant J.* **71**, 871–880 (2012).
62. O. Björkman, R. W. Pearcy, A. T. Harrison, H. Mooney, *Science* **175**, 786–789 (1972).
63. I. H. Köhler *et al.*, *J. Exp. Bot.* **68**, 715–726 (2017).
64. H. Xu *et al.*, *Plant Commun.* **4**, 100641 (2023).
65. E. N. Smith, M. van Aalst, A. P. Weber, O. Ebenhoeh, M. Heinemann, *Sci. Adv.* **11**, eadt9287 (2025).
66. A. P. Cavanagh, P. F. South, C. J. Bernacchi, D. R. Ort, *Plant Biotechnol. J.* **20**, 711–721 (2022).
67. K. Meacham-Hensold *et al.*, *Glob. Change Biol.* **30**, e17595 (2024).
68. R. Furbank, S. Kelly, S. von Caemmerer, *Photosynth. Res.* **158**, 121–130 (2023).
69. N. D. Nguyen *et al.*, *Photosynth. Res.* **156**, 265–277 (2023).
70. C. Fei, A. T. Wilson, N. M. Mangan, N. S. Wingreen, M. C. Jonikas, *Nat. Plants* **8**, 583–595 (2022).
71. AgBioInvestor, *Time and Cost to Develop a New GM Trait* (Porthhead, 2022).
72. S. Strobbe, J. Wesana, D. Van Der Straeten, H. De Steur, *Trends Biotechnol.* **41**, 736–740 (2023).
73. K. Emerick, P. C. Ronald, *Cold Spring Harb. Perspect. Biol.* **11**, a034637 (2019).
74. H. Kimm *et al.*, *Glob. Change Biol.* **27**, 2403–2415 (2021).
75. P. Fu *et al.*, *J. Exp. Bot.* **73**, 3157–3172 (2022).
76. J. Bagley *et al.*, *Global Biogeochem. Cycles* **29**, 194–206 (2015).
77. O. S. O'Sullivan *et al.*, *Glob. Change Biol.* **23**, 209–223 (2017).
78. S. P. Long, A. K. Spence, *Annu. Rev. Plant Biol.* **64**, 701–722 (2013).
79. K. Prado *et al.*, bioRxiv 2023.2006.2023.546155 [Preprint] (2023); https://doi.org/10.1101/2023.06.23.546155.
80. U. M. Ruiz-Vera *et al.*, *Plant Physiol.* **162**, 410–423 (2013).
81. M. H. Siebers *et al.*, *Agric. Ecosyst. Environ.* **240**, 162–170 (2017).

ACKNOWLEDGMENTS

Funding: Funding for this work was provided by the US Department of Energy (DOE) Center for Advanced Bioenergy and Bioproducts Innovation (DOE, Office of Science, Office of Biological and Environmental Research, under award no. DE-SC0018420); the Bill & Melinda Gates Foundation; the Foundation for Food and Agriculture Research and the UK Foreign, Commonwealth & Development Office under grant no. OPP11722157; and Gates Agricultural Innovations grant investment ID 57248. Views expressed in this article are those of the authors and do not necessarily reflect those of the funding agencies acknowledged here.
Competing interests: The authors declare no competing interests.

Submitted 10 March 2025; accepted 25 April 2025

10.1126/science.adv5413

REVIEW

Dispersed components drive temperature sensing and response in plants

Avilash Singh Yadav[1,2], Sridevi Sureshkumar[1], Alok Krishna Sinha[3], Sureshkumar Balasubramanian[1]*

Plants are highly sensitive to temperature, and climate change is predicted to have negative impacts on agricultural productivity. Warming temperatures, coupled with a growing population, present a substantial challenge for food security and motivate research to understand how plants sense and respond to changes in temperature. Here, we synthesize our current understanding of temperature sensing and response in plants. We outline how temperature cues are integrated into preexisting signaling cascades using inherently temperature-sensitive proteins or processes. This dispersed nature of thermo-sensitive proteins and processes makes distinct signaling cascades sensitive to temperature. This model integrates current knowledge and distinguishes thermosensing from other conventional sensing and signaling mechanisms in plants.

Optimizing growth and development in a changing environment requires the development of sensory systems to capture the change and signaling systems to ensure appropriate responses. Light and temperature are two critical, and often inseparable, parameters that influenced life on Earth over evolutionary timescales. Although both have a profound effect on organisms' function, light sensitivity is typically restricted to photoreceptors, whereas temperature sensitivity is seen across a wide range of biomolecules. Temperature constantly influences the physical properties of molecules, including the structure, function, and interactions of proteins and RNA. The rates of all biochemical reactions are also sensitive to temperature. Consequently, many organisms have evolved homeostatic temperature control, facilitated by the neuronal system, as seen in homeothermic animals (*1*, *2*). The mechanisms of temperature control differ between animals, with heterothermic animals relying more on external environment and physical movements compared to homeothermic animals, which rely on their neuronal thermoregulatory systems.

The way plants sense and respond to temperature is fundamentally distinct from that of animals, and recognizing this distinction is essential for understanding thermal responses in plants. Plants neither possess a thermal homeostasis mechanism nor can they move to avoid unfavorable conditions, so their internal temperature is dependent on their surroundings. However, diurnal changes, and unpredictable fluctuations, present additional challenges for plants to adapt and respond to changes in temperature (*3*). As a result, plants are highly sensitive to the environment.

Temperature influences all aspects of plant life (*4*). In seed plants, all developmental stages, from germination to seed set, are affected by temperature, so adverse temperature can substantially affect crop yields (*5*). Even a 1°C increase is predicted to have a substantial negative effect on the yield of major food crops (*6*–*8*). With a growing population, warming temperatures present major challenges for global food security. This has motivated studies to discern the molecular basis of temperature sensing, which could be exploited to engineer climate-smart crops for the future. Here, we review research progress and suggest that temperature-sensitive molecules and/or processes, dispersed in various pathways, allow integration of thermal cues to govern phenotypic responses in plants. For a detailed catalog and mechanisms, we recommend these excellent reviews (*4*, *9*–*17*).

Sensors in a biological context

The origin of the term "sensor" in biology can be traced back to Adler in 1975, who used the term while defining chemosensors (*18*). Adler differentiated sensors from receptors stating that sensors respond in a transient manner for a brief period, upon a change in the concentration of the stimulus, rather than exhibiting sustained activity for the duration of the stimulus. Adler stated that this transient response was a characteristic feature of sensors. The defining feature of a sensor lies in its ability to detect a change in the environment and transmit a short-lived signal to initiate the downstream response pathway, after which the sensor returns to its prior state. In chemotaxis, bacteria detect and move in response to changes in attractant concentration, which is triggered by a transient signaling event initiated by the sensor. Conversely for lactose uptake, membrane-bound receptors continuously detect and transport the nutrient without generating a transient signal or returning to the inactive state (*18*). Therefore, receptors function more like sustained detectors, whereas sensors, per se, are molecules that transmit a signal by sensing the change in the environment in a transient manner to initiate a response (*4*).

In this context, light receptors can be considered as photosensors, where they sense light and transmit that information through a subsequent signaling pathway (*19*). Light sensors (phytochromes, cryptochromes, and phototropins) have chromophore domains that capture photons and transmit the signal to their downstream signaling network (*20*). Similarly, there are mechanosensitive ion channels that can be triggered by touch or wind, resulting in an immediate response via calcium signaling (*21*–*23*). These types of molecules that are rapidly activated by a stimulus and subsequently trigger a transient response and act as the initiating element in a signaling pathway (primum movens) are fit to be considered sensors. The stimulus is essential for this transient effect and downstream signaling rather than being a mere catalyst for a reaction. By contrast, plants also possess receptors that allow them to detect other types of stimuli such as pathogens (*24*). These receptors typically bind to molecules that are either present in pathogens or secreted by them and then initiate a response through a downstream signaling network (*24*). Typically, they remain active as long as the ligands are present, unlike sensors that trigger a downstream response and return to their original state. Hence, we emphasize that temperature-sensitive components should not be considered sensors, as this draws parallels to light or touch sensors triggering a transient response and initiating downstream signaling, and to neurosensory systems, which trigger a coordinated organismal response.

Temperature sensing in plants

Temperature is distinct from other signals (*25*). Thermoreceptors do not bind to a ligand. There are no specific particles such as a photon that could be considered a signal, and what is sensed is the difference in the transfer of kinetic energy of particles caused by their molecular motion, with increases in temperature meaning faster movement of particles and vice versa. In animals, there are transient receptor potential (TRP) channels, which change their activity in response to temperature in a transient manner. This in turn leads to a transient

[1]School of Biological Sciences, Monash University, Clayton Campus, VIC 3800, Australia. [2]Weill Institute for Cell and Molecular Biology and Section of Plant Biology, School of Integrative Plant Sciences, Cornell University, Ithaca, NY, USA. [3]National Institute of Plant Genome Research, Aruna Asaf Ali Marg, New Delhi, 110067, India. ***Corresponding author. Email: mb.suresh@monash.edu**

influx of ions such as Ca^{2+} and generates a signaling cascade triggering a neuronal thermoregulatory response (*26*–*28*). Although there are no direct orthologs of TRP channels in plants, there are mechanosensitive ion channels that exhibit changes in Ca^{2+} flux in response to external cues. However, their role in temperature sensing remains unclear (*29*).

Given that temperature, by definition, is a measure of the average kinetic energy of particles, it influences the physical properties of biomolecules. Both function and activity of biomolecules are inherently dependent on temperature. When these biomolecules are part of signaling cascades, their temperature sensitivity allows them to integrate thermal cues directly into these specific pathways with phenotypic consequences. These phenotypic responses to temperature make it amenable to genetic analysis (*30*–*32*). Indeed, thermal responses of plants are routinely studied using a variety of phenotypes such as seed germination (*33*), hypocotyl elongation (*34*), flowering time (*30*, *31*), heat or cold stress tolerance (*25*), photosynthetic efficiency (*35*), and transpiration rate (*33*, *36*). Several key temperature-sensitive components have been identified over the years, many of which have either been designated as "thermosensors" or have been assigned "sensory" roles in the context of temperature perception (*9*, *10*, *12*, *37*, *38*).

Red and blue light receptors integrate thermal cues in light signaling

Phytochromes are light sensors, of which phytochrome B (phyB) is a well-characterized photoreceptor that senses red light (*19*, *39*, *40*). Phytochromes exist in two interconvertible forms, Pr and Pfr, and conversion between these forms is light dependent. Under red light, phyB converts from the inactive Pr:Pr homodimer form to the active Pfr:Pfr form (Fig. 1A). This activation can be either reversed under far-red light or through a slow process that occurs passively in darkness, referred to as dark reversion. The rate of this passive reversal process increases with temperature (Fig. 1A). This temperature-enhanced reversal process allows phyB to integrate light and temperature signals, a property that has prompted researchers to classify phyB as a thermosensor (*41*, *42*).

A second mechanism has also been attributed to how phyB contributes to temperature perception in plants (Fig. 1A). In addition to the conformational switch, phyB spontaneously undergoes liquid-liquid phase separation (LLPS) to form nuclear condensates, known as photobodies (*43*–*45*), which sequester transcription factors such as PHYTOCHROME INTERACTING FACTOR (PIFs) at lower temperatures (*46*). PIFs play a critical role in driving morphological changes in response to ambient temperature (thermomorphogenesis), with PIF4 being the central hub (*4*, *10*, *19*, *46*). At elevated temperatures, phyB photobodies dissolve, releasing PIFs to activate auxin biosynthesis genes and promote thermomorphogenic responses (*47*). The formation of phyB photobodies is driven by its C-terminal domain, which promotes self-oligomerization, and by the intrinsically disordered region (IDR) in N-terminal extension, which confers thermosensitivity to phyB phase separation. This phase-separation mechanism operates independently of dark reversion, thus providing an additional mode through which phyB integrates temperature cues (*45*). Nevertheless, much of the evidence for phyB phase separation comes from overexpression studies and heterologous systems using fluorescent tags (*45*, *48*). Given that LLPS is sensitive to protein concentration and steric hindrances, the physiological relevance of phyB phase separation in temperature perception should be inferred with caution. In our view, phyB is a temperature-responsive regulator whose activity or state is influenced by temperature, allowing the red-light signaling pathway to gain temperature sensitivity, rather than a classical "sensor" as defined by Adler (*18*).

Because phyB physically interacts with PIF4 to repress its transcriptional activity, a reduction in active phyB levels at elevated temperature derepresses PIF4, which subsequently activates auxin biosynthesis genes to promote hypocotyl elongation (*41*, *42*). However, *phyB* mutants still retain some degree of temperature responsiveness in hypocotyl elongation (*42*). Even the *phyABCDE* quintuple mutant, which exhibits severely impaired hypocotyl elongation in response to elevated temperature, still retains some temperature responsiveness (*41*, *42*). These observations indicate the presence of additional mechanisms of temperature perception.

Evidence for such mechanisms comes from other photoreceptors as well. Phototropins, blue light receptors in *Marchantia*, have been proposed to be thermosensors (*49*). Phototropins contain a light-oxygen-voltage (LOV)–sensing domain that binds noncovalently with

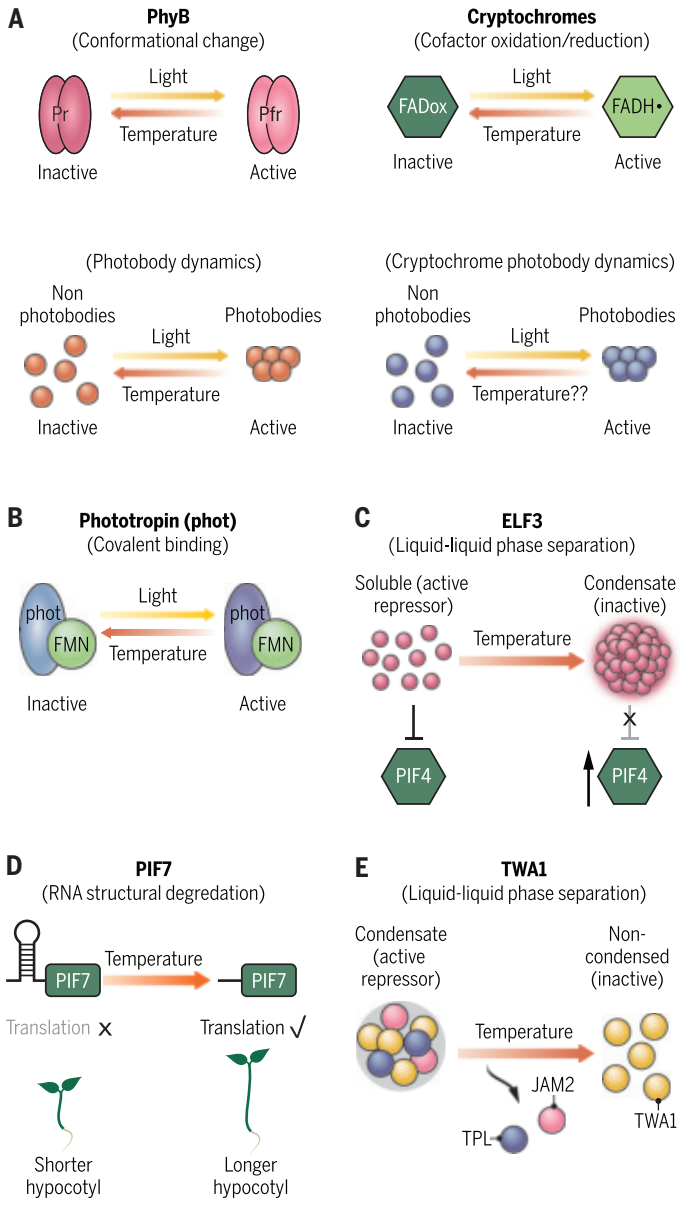


Fig. 1. A summary of designated thermosensors in *Arabidopsis* and their modes of action. (**A** and **B**) The modes of action of photoreceptors reveal that activation to their active state occurs through light and the reversal reaction to inactive state is accelerated by temperature. (**C** and **D**) ELF3 and PIF7 present nodes for integrating thermal cues on to the clock signaling networks. (**E**) TWA1 provides an example of integration of thermal cues onto hormonal signaling. The reactions have been simplified to specifically highlight the influence of temperature. For example, rate of reactions as well as LLPS can also be affected by many other parameters.

CREDIT: ADAPTED BY K. HOLOSKI/*SCIENCE*

a flavin mononucleotide (FMN). Here, the light activates phototropin, which returns to the inactive state in a reaction that is accelerated by elevated temperature (Fig. 1B). Cryptochromes, which are another class of blue-light receptors in plants and animals, also use flavin-based chromophores to sense light by modulating the redox state of the chromophore followed by signaling (*50*, *51*). Temperature is known to affect the rate of change in the redox status (*52*) and cryptochromes also undergo temperature-dependent LLPS, resembling phytochromes (*53*, *54*). However, it is unclear whether temperature influences the cryptochrome photobodies in a way similar to that of phytochromes.

Taken together, the above evidence suggests that the photoreceptors integrate thermal cues in their respective light signaling pathways via conformational changes, redox chemistry, or phase separation. The commonality lies in the fact that all of these processes involve a temperature-dependent change in a rate of a chemical reaction (Fig. 1A), which allows the signaling pathways to acquire temperature sensitivity. Temperature acts as a catalyst for reactions that can also occur in a passive manner, suggesting that these are temperature-sensitive components rather than "thermosensors."

ELF3 and PIF7 integrate thermal cues in clock signaling

The plant circadian clock regulates the expression of multiple genes, including those involved in the perception and response to light and temperature (*55*). One of the critical components of the circadian clock is the evening complex, which comprises the proteins EARLY FLOWERING 3 (ELF3), ELF4, and LUX ARRHYTHMO (LUX), whose levels peak in the evening. ELF3 imparts temperature sensitivity to the evening complex (*56*, *57*). ELF3 represses PIF4 transcriptionally and through their direct physical interaction. Elevated temperatures degrade ELF3 in a ubiquitin-dependent proteasomal response, which relieves its inhibitory activity (*58*). In addition, ELF3 undergoes temperature-dependent phase transition (*59*). At low temperature, ELF3 remains in the soluble (active) state, which allows ELF3 to repress the transcription of its target genes. Conversely, at high temperature, ELF3 forms dense punctate condensates (nuclear speckles) and transitions to an inactive state. This high temperature–mediated inactivation of ELF3 is reminiscent of the thermal deactivation mechanisms exhibited by photoreceptors (Fig. 1C).

A distinct feature of ELF3 is a polyglutamine (polyQ) repeat embedded within a prion-like domain (PrLD) (*59*). The PrLD mediates phase transition, and the length of the polyQ repeat correlates with the thermal sensitivity of ELF3. However, transgenic plants lacking the polyQ tract in ELF3 still exhibit thermosensitivity, indicating that polyQ is not essential for the thermosensitivity of ELF3. Evolutionary analysis suggests that the role of PrLD of ELF3 in thermal adaptation may be lineage or species specific (*60*). Based on the ability of ELF3 to undergo phase separation in response to temperature, the authors concluded that ELF3 is a thermosensor (Fig. 1C) (*59*). In our view, ELF3 behavior reflects a more gradual, thermodynamically driven change in physical state, rather than a transient, initiating component that actively "senses" temperature. Essentially, ELF3 integrates temperature cues into circadian clock network.

Thermal cues can be integrated into the same signaling cascade at multiple nodes (Fig. 1E). For instance, ELF3 physically interacts with PIF7, which promotes hypocotyl elongation in response to shade (low red to far-red light ratio) (*61*). By binding to PIF7, ELF3 inhibits the DNA binding activity of PIF7 to regulate the shade avoidance response in a circadian-dependent manner (*62*). By contrast, PIF7 translation increases at elevated temperature, which enhances the ability of PIF7 to activate key genes required for auxin biosynthesis and subsequent hypocotyl elongation (*63*). The increase in PIF7 translation is attributed to an impact of temperature on RNA structure, which the authors refer to as a thermosensitive "switch" (*63*). Briefly, the RNA structure at the 5′ untranslated region (5′-UTR) of *PIF7* transcript forms a stable RNA hairpin at lower temperature, which inhibits translation. However, elevated temperature melts the hairpin structure to form an open RNA conformation, resulting in an increase in PIF7 translation (Fig. 1E).

Temperature-responsive RNA elements in the 5′-UTRs are a common occurrence in bacterial systems (*64*). The passive response of mRNA structures to temperature is a well-known principle of RNA thermodynamics (*65*). Indeed, Chung and colleagues show that the relaxation of *PIF7* mRNA structure with increasing temperature is a gradual process, rather than a "switch" per se (*63*). Nevertheless, the thermoregulation of *PIF7* mRNA is evidence for a mechanism by which temperature-mediated RNA structural changes can have direct consequences in plants (*63*). In summary, ELF3 and PIF7 represent examples of temperature cues being integrated into at least two distinct nodes of an existing signaling cascade (Fig. 1, C and D).

TWA1 integrates thermal cues into ABA signaling

Abscisic acid (ABA) is a key hormone involved in plant responses to environmental stressors such as drought and heat (*66*, *67*). High temperature induces heat stress in plants leading to the accumulation of ABA, which triggers a downstream signaling cascade resulting in alterations in gene expression and subsequent phenotypic responses to withstand heat stress. The *thermo with aba-response-1* (*twa-1*) mutant is hypersensitive to both ABA and heat stress. The *twa-1* mutant displays severely reduced survival and growth, including increased loss of chlorophyll and shorter root growth, which are key markers for heat-induced cellular damage (*68*).

TWA1 is predicted to encode a protein with an IDR in its N-terminal, and functions as a transcriptional repressor (*68*). Like phyB, TWA1 also undergoes LLPS and forms nuclear condensates in specific nuclear subdomains at low temperature (*68*). In these subdomains, TWA1 interacts with other factors to assemble a temperature-sensitive repressor complex that represses heat stress–responsive genes. However, the nuclear condensates dissolve at high temperature, and subsequently the repressor complex disassociates to release the repressor components from the nuclear subdomain (Fig. 1E) (*68*). This allows the transcriptional activation of heat stress–responsive genes to confer thermotolerance (Fig. 1E).

The authors state that TWA1 functions as a thermosensor based on its ability to undergo temperature-dependent phase separation and regulate gene expression in response to heat (*68*). As with phyB or ELF3, elevated temperature disassembles the TWA1 condensates—although, in this case, loss of condensates alleviates TWA1-mediated repression and activates stress-responsive genes, rather than inactivating the protein itself. Although TWA1 clearly is a temperature-sensitive protein, it does not qualify as a "sensor." In our view, the TWA1 behavior reflects a passive, temperature-dependent property of IDRs rather than an active sensor that undergoes a brief conformational change to produce a transient signal and initiate a cascade. In summary, TWA1 represents an example of a temperature-sensitive component that integrates thermal cues into the ABA signaling cascade (Fig. 1E).

Temperature-sensitive nodes integrate thermal cues into chromatin-level gene regulation

Temperature changes influence gene expression by modulating DNA accessibility via chromatin remodeling (*69*). Most genes that are responsive to external stimuli harbor the histone variant H2A.Z at their +1-nucleosome (*70*). Elevated temperatures lead to eviction of H2A.Z nucleosomes (*71*). Based on the requirement of H2A.Z eviction to facilitate transcription in response to elevated temperature, H2A.Z-nucleosomes were suggested to be critical for thermosensory perception and transcriptomic responses (*71*). However, subsequent studies raised doubts as to whether H2A.Z is a primary sensor of thermal cues. For instance, the HSFA1 clade of heat shock

transcription factors is required for H2A.Z eviction (*72*). Similarly, histone deacetylation is essential for temperature-induced H2A.Z eviction (*73*, *74*). Moreover, post-translational modification of PIF4 by a mitogen-activated protein (MAP) kinase is also required for H2A.Z eviction (*75*).

The histone variant H2A.Z also plays an important role in regulating temperature-dependent immune responses in plants (*76*). Defense responses of plants are generally suppressed at elevated temperatures (*77*). Loss-of-function mutations in *H2A.Z (hta9hta11)* exhibit increased pattern-triggered immunity, regardless of the actual environmental temperature (*76*). Thus, chromatin remodeling involving H2A.Z is indeed a component of thermal response. However, it is a temperature-sensitive node rather than a primary trigger of a signaling cascade. The temperature-sensitive nature of H2A.Z nucleosomal dynamics allows the integration of thermal cues on defense signaling and other transcriptional responses.

Inherent temperature sensitivity underlies thermal responses

A closer look at the temperature-sensitive components described above reveals some recurring patterns in how plants perceive and respond to temperature (Fig. 2). First, in general, temperature passively influences the rate of reactions. Unlike light, temperature does not initiate a switch in the state of photoreceptors but accelerates the kinetics of dark reversion or the dynamics of photobody formation, all of which can occur naturally, independent of the temperature stimulus. Therefore, temperature acts as an influencer rather than an initial sensory input. The temperature-sensitive nature of reaction rates allows integration of thermal cues into signaling cascades.

Second, LLPS appears to be a recurring phenomenon. Several temperature-sensitive proteins such as PHYB, CRY2, ELF3, HSF1A1, and TWA1 as well as other proteins such as FRIGIDA, FCA, and RNA binding proteins in plants undergo LLPS to form dynamic nuclear condensates (*78*–*82*). Being a biophysical process, LLPS is highly sensitive not only to temperature, but also to other environmental parameters such as pH (*83*), ionic strength, concentration of biomolecules such as RNA, post-translational modifications of proteins (*84*), and different stressors (*85*). The condensates are typically enriched for proteins that contain low-complexity regions such as IDRs or PrLDs. They also appear to be enriched for RNA binding proteins. IDRs aid in LLPS in part because their low complexity allows multiple interactions with other proteins and RNA, as well as their potential to undergo post-translational modifications. Temperature, therefore, influences the stability and assembly of the condensates at the biophysical level. Thus, LLPS provides a highly

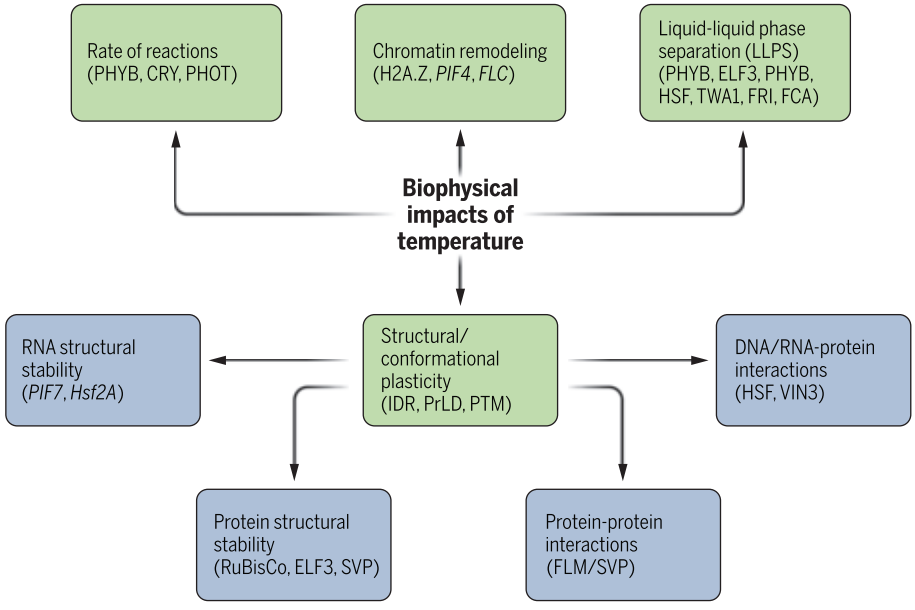


Fig. 2. Temperature perception involves general biophysical properties. Almost all aspects of temperature perception involve generalizable properties, which could in principle be applied to almost all biomolecules. This generalizable thermodynamic property and the absence of discrete "sensory" mechanisms make thermosensing distinct from other sensing and signaling cascades. The structural and conformational plasticity intrinsic to IDRs, PrLDs, and PTMs confer changes in structural stability and interactions that are sensitive to temperature.

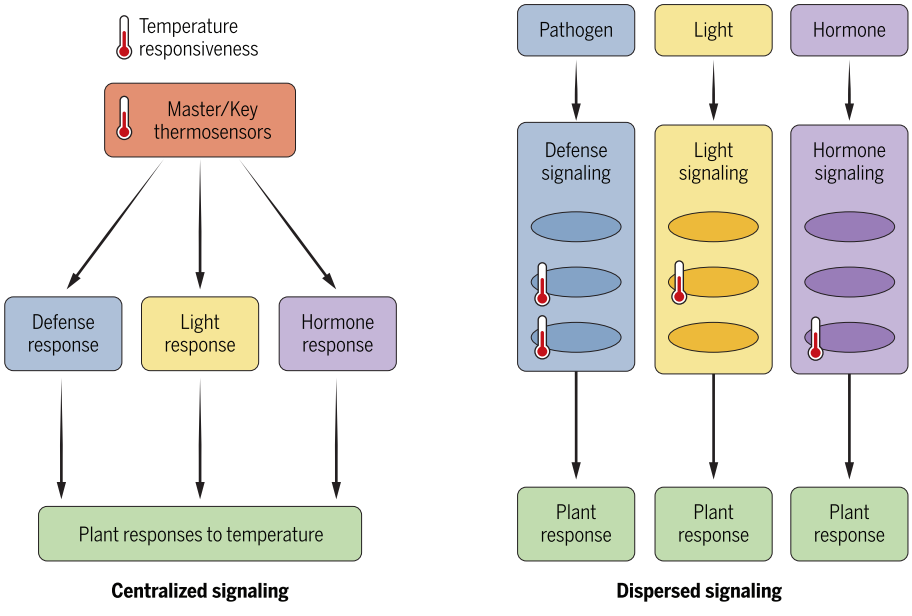


Fig. 3. Dispersed components ensure robust temperature response in plants. We propose that plants sense temperature in a dispersed manner through various pathways, as opposed to dedicated signaling initiated by a few thermosensors. Dispersed signaling occurs owing to the dispersed nature of temperature-sensitive components in varied pathways, which allows thermal cues to be integrated in a multimodal way, ensuring robust organismal response to changing temperatures.

CREDITS: ADAPTED BY K. HOLOSKI/*SCIENCE* (*2*)

temperature-sensitive regulatory layer that enables incorporation of thermal cues into the broader processes or signaling cascades that they are involved in (Fig. 2).

The third aspect of temperature sensitivity comes from the general effect of temperature on the structure and stability of biomolecules. Both protein folding and RNA secondary structure are known to be thermodynamically sensitive. It is well known that in homeothermic animals, proteins are more optimized to function within a narrow range of temperatures, whereas in heterothermic animals, proteins are more resilient to diverse thermal regimes (*86*). Plants, like heterothermic animals, have proteins that are more flexible in their functionality with respect to temperatures, but may have also evolved to acquire new functions (*87*). For example, temperature influences the activity of the carbon fixation enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) to modulate photosynthetic efficiency (*88*). Protein structure is affected by temperature either directly or through post-translational modifications, which can lead to targeted degradation. For example, ELF3 and the flowering time protein, SHORT VEGETATIVE PHASE, undergo ubiquitin-mediated degradation at high temperature (*58*, *89*). At high temperature, a membrane-localized E3 ubiquitin ligase (TT3.1) in rice translocates to endosomes, where it targets a chloroplast precursor protein for degradation (*37*). In the case of TT3.1, the protein might be categorized as a thermosensor, because there is a change in membrane organization and a temperature-dependent immediate response. However, it is still unclear how TT3.1 perceives thermal cues for its translocation. Nevertheless, temperature clearly perturbs the structure of biomolecules, either through conformational changes or through modifications of their interacting partners, which makes biomolecules inherently sensitive to temperature (Fig. 2).

Fourth, like protein structure, RNA secondary structures are influenced by temperature. Bacteria routinely use the inherent temperature-sensitive property of RNAs to assess host temperatures (*64*, *90*). Although there are fewer examples in plants, the case of *PIF7* mRNA presents a clear example of a temperature-sensitive hairpin in the 5′ UTR, which inhibits translation at low temperature but melts at higher temperature to allow translation and downstream phenotypic responses. Similarly, RNA structures influenced by temperature can modulate differential splicing. For example, *Hsf2A* pre-mRNA in tomato undergoes temperature-dependent alternative splicing as a result of changes in the RNA structure (*91*). These examples illustrate that the thermodynamics of RNA structures add a regulatory layer to integrate thermal cues into existing signaling cascades (Fig. 2).

Finally, temperature regulates transcription by acting at the chromatin level, which may be brought about by influencing DNA-protein or protein-protein interactions. Chromatin changes can have substantial impacts on downstream processes (*15*). For example, temperature influences protein-kinase interactions by altering the phosphorylation status of the proteins of interest (*75*, *92*). Similarly, interactions between the polycomb repressive complex and other proteins such as VIN3 are known to be influenced by temperature (*93*). Along the same lines, even the interactions between heat shock transcription factors and their target sites appear to be influenced by temperature (*94*). Collectively, the influence of temperature on the binding affinity of DNA-protein or protein-protein interactions allows thermal cues to be integrated into transcriptional responses (Fig. 2).

Dispersed temperature-sensitive components ensure robust thermal responses

Plants perceive temperature by leveraging the inherent biophysical effects of temperature on molecular components rather than through discrete, specialized "sensory" modules. Temperature-sensitive molecules and processes are dispersed across various signaling pathways (e.g., light signaling, hormone signaling, clock signaling, etc.) and responses (defense responses, stress responses, and developmental responses), essentially conferring thermosensitivity to these pathways themselves (Fig. 3). None of the responses discussed here represent "sensing" in a canonical manner of initiating a discrete signaling cascade. Instead, they reflect the thermodynamic responsiveness that is intrinsic to the molecules and processes involved.

Given the randomness of mutations, we propose that the presence of temperature-sensitive components in a molecular pathway originates as a chance event allowing further diversification in ectothermic organisms. To reflect this randomness, we refer to the mode in which plants perceive temperature as "dispersed" rather than "distributed." Dispersed perception of thermal stimuli thus differs from other sensing mechanisms that operate in plants. This dispersed nature through which temperature influences signaling in general ensures that plants are generally sensitive to temperature in a multimodal way, conferring robustness in thermal response at the organismal level (Fig. 3).

A similar framework has already been proposed (*95*–*97*) in the context of vernalization response, where thermal cues are integrated across timescales and molecular nodes to regulate *FLC* expression. Modeling of these responses, particularly with fluctuating temperatures, has suggested that there are multiple nodes of thermal perception in the system, including the ability to filter noisy thermal inputs over time (*95*–*97*). In addition, reduced growth rate caused by low temperatures could itself be used as a passive measure of temperature over long periods enabling long-term adaptation (*97*). All of these findings suggest that the field would benefit by moving beyond the idea of thermosensors and consider plant thermosensing as an emergent property of multiple temperature-sensitive processes dispersed across signaling networks.

The dispersed perception model is supported by the observation that none of the identified "thermosensors" when mutated leads to plants being "temperature-blind" (Box 1). The mutant phenotypes tend to be specific to the individual pathway in which the proposed "thermosensor" operates. This model is also supported by evolutionary

Box 1. Arguments for dispersed signaling are presented along with some of the questions this model raises.

Arguments for dispersed signaling

- There is no ligand for temperature, which differentiates it from other environmental signals.
- Loss-of-function mutations in named thermosensors do not result in "temperature blind" plants.
- Temperature sensitivity can arise even from single amino acid substitutions.
- None of the designated thermosensors function exclusively in temperature sensing.
- Dispersed components may provide an evolutionary advantage to alter specific pathways.
- A dispersed model allows thermosensitivity to evolve in a species- or lineage-specific manner.

Questions of interest

- How do mutants of a thermosensor behave in assays used for other thermosensitive phenotypes?
- Why do plants have dispersed signaling? Is it related to their poikilothermic and sessile nature?
- If dedicated master-sensors do exist in plants, why has it been difficult to capture them?
- Why would such an adaptive mechanism be exclusive to plants?
- Is there evidence for similar thermosensing mechanisms in other organisms?
- Have the photoreceptors evolved to perceive both light and temperature?

divergence of temperature-sensitive nodes, which suggests that evolution of temperature perception mechanisms could be species specific. For example, PIF7 RNA or the variable length of poly(Q) tract in the PrLD domain of ELF3 (*59*, *63*) are likely evolutionary accidents that vary in a species-specific manner. Our model also demonstrates how distinct signaling cascades that interact with each other could be influenced by the impact of a temperature-sensitive components in another cascade contributing to pleiotropy.

In summary, we propose that dispersed temperature perception, facilitated by the presence of inherently temperature-sensitive components and reactions, ensures robust system-wide phenotypic responses in plants. The model that we proposed immediately suggests that designer crops optimized for varied temperatures can be developed on the basis of several proteins, which may integrate thermal cues on different pathways. By targeting proteins of interest in specific signaling cascades, which may be important in different crops or conditions, one would be able to develop thermo-tolerant varieties that are suited for specific climatic conditions and species. This approach could be coupled with synthetic biology to incorporate temperature sensitivity into proteins of new signaling pathways to create designer crops.

REFERENCES AND NOTES

1. A. Clarke, *Principles of Thermal Ecology: Temperature, Energy and Life* (Oxford Academic, 2017).
2. A. W. Crompton, C. R. Taylor, J. A. Jagger, *Nature* **272**, 333–336 (1978).
3. P. E. Verslues *et al.*, *Plant Cell* **35**, 67–108 (2023).
4. J. J. Casal, S. Balasubramanian, *Annu. Rev. Plant Biol.* **70**, 321–346 (2019).
5. S. Peng *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9971–9975 (2004).
6. E. E. Rezaei *et al.*, *Nat. Rev. Earth Environ.* **4**, 831–846 (2023).
7. C. Zhao *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 9326–9331 (2017).
8. A. J. Challinor *et al.*, *Nat. Clim. Chang.* **4**, 287–291 (2014).
9. S. M. Kerbler, P. A. Wigge, *Annu. Rev. Plant Biol.* **74**, 341–366 (2023).
10. J. J. Casal, G. Murcia, L. Bianchimano, *Annu. Rev. Genet.* **58**, 135–158 (2024).
11. Y. Ding, S. Yang, *Dev. Cell* **57**, 947–958 (2022).
12. Y. Kan, X. R. Mu, J. Gao, H. X. Lin, Y. Lin, *Mol. Plant* **16**, 1612–1634 (2023).
13. M. Noguchi, Y. Kodama, *Plant Cell Physiol.* **63**, 737–743 (2022).
14. X. Li, T. Liang, H. Liu, *Plant Cell* **34**, 955–966 (2022).
15. T. Xu, E. Patitaki, A. Zioutopoulou, E. Kaiserli, *Curr. Opin. Plant Biol.* **83**, 102668 (2025).
16. J. Li, Y. Song, *Plant Sci.* **342**, 112025 (2024).
17. T. Zhu *et al.*, *Front. Plant Sci.* **16**, 1560204 (2025).
18. J. Adler, *Annu. Rev. Biochem.* **44**, 341–356 (1975).
19. M. C. Cheng, P. K. Kathare, I. Paik, E. Huq, *Annu. Rev. Plant Biol.* **72**, 217–244 (2021).
20. C. Kami, S. Lorrain, P. Hornitschek, C. Fankhauser, *Curr. Top. Dev. Biol.* **91**, 29–66 (2010).
21. E. S. Hamilton, A. M. Schlegel, E. S. Haswell, *Annu. Rev. Plant Biol.* **66**, 113–137 (2015).
22. G. B. Monshausen, E. S. Haswell, *J. Exp. Bot.* **64**, 4663–4680 (2013).
23. F. W. Telewski, *Am. J. Bot.* **93**, 1466–1476 (2006).
24. J. D. G. Jones, B. J. Staskawicz, J. L. Dangl, *Cell* **187**, 2095–2116 (2024).
25. J. K. Zhu, *Cell* **167**, 313–324 (2016).
26. K. Venkatachalam, C. Montell, *Annu. Rev. Biochem.* **76**, 387–417 (2007).
27. M. Kashio, M. Tominaga, *Curr. Opin. Neurobiol.* **75**, 102591 (2022).
28. A. Patapoutian, *Chem. Senses* **30** (suppl. 1), i193–i194 (2005).
29. E. Jarratt-Barnham, L. Wang, Y. Ning, J. M. Davies, *Int. J. Mol. Sci.* **22**, 874 (2021).
30. S. Balasubramanian, S. Sureshkumar, J. Lempe, D. Weigel, *PLOS Genet.* **2**, e106 (2006).
31. M. A. Blázquez, J. H. Ahn, D. Weigel, *Nat. Genet.* **33**, 168–171 (2003).
32. K. J. Halliday, M. G. Salter, E. Thingnaes, G. C. Whitelam, *Plant J.* **33**, 875–885 (2003).
33. J. Chory *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 12066–12071 (1996).
34. W. M. Gray, A. Ostin, G. Sandberg, C. P. Romano, M. Estelle, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7197–7202 (1998).
35. S. Hussain *et al.*, *Photosynth. Res.* **150**, 5–19 (2021).
36. M. Quint *et al.*, *Trends Plant Sci.* **28**, 1098–1100 (2023).
37. H. Zhang *et al.*, *Science* **376**, 1293–1300 (2022).
38. J. H. Jung, P. J. Seo, E. Oh, J. Kim, *Trends Plant Sci.* **28**, 924–940 (2023).
39. P. H. Quail *et al.*, *Science* **268**, 675–680 (1995).
40. J. Hughes, A. Winkler, *Annu. Rev. Plant Biol.* **75**, 153–183 (2024).
41. J. H. Jung *et al.*, *Science* **354**, 886–889 (2016).
42. M. Legris *et al.*, *Science* **354**, 897–900 (2016).
43. M. Chen, J. Chory, *Trends Cell Biol.* **21**, 664–671 (2011).
44. Y. Kwon, C. Kim, G. Choi, *New Phytol.* **242**, 909–915 (2024).
45. D. Chen *et al.*, *Mol. Cell* **82**, 3015–3029.e6 (2022).
46. M. A. Koini *et al.*, *Curr. Biol.* **19**, 408–413 (2009).
47. J. Hahm, K. Kim, Y. Qiu, M. Chen, *Nat. Commun.* **11**, 1660 (2020).
48. R. J. A. Kim *et al.*, *Nat. Commun.* **15**, 3519 (2024).
49. Y. Fujii *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 9206–9211 (2017).
50. A. R. Cashmore, J. A. Jarillo, Y. J. Wu, D. Liu, *Science* **284**, 760–765 (1999).
51. J. Ponnu, U. Hoecker, *Front. Plant Sci.* **13**, 844714 (2022).
52. M. C. de Pinto, V. Locato, A. Paradiso, L. De Gara, *Ann. Bot.* **116**, 487–496 (2015).
53. B. Jiang *et al.*, *Nat. Plants* **9**, 2042–2058 (2023).
54. B. Jiang, *New Phytol.* **244**, 2163–2169 (2024).
55. N. Creux, S. Harmer, *Cold Spring Harb. Perspect. Biol.* **11**, a034611 (2019).
56. D. A. Nusinow *et al.*, *Nature* **475**, 398–402 (2011).
57. T. Mizuno *et al.*, *Plant Cell Physiol.* **55**, 958–976 (2014).
58. D. Ezer *et al.*, *Nat. Plants* **3**, 17087 (2017).
59. J. H. Jung *et al.*, *Nature* **585**, 256–260 (2020).
60. Z. Zhu, J. Trenner, C. Delker, M. Quint, *Mol. Biol. Evol.* **41**, msae205 (2024).
61. P. Leivar *et al.*, *Plant Cell* **20**, 337–352 (2008).
62. Y. Jiang *et al.*, *iScience* **22**, 288–298 (2019).
63. B. Y. W. Chung *et al.*, *Nat. Plants* **6**, 522–532 (2020).
64. J. Kortmann, F. Narberhaus, *Nat. Rev. Microbiol.* **10**, 255–265 (2012).
65. A. Becskei, S. Rahaman, *Comput. Struct. Biotechnol. J.* **20**, 4325–4336 (2022).
66. A. S. Raghavendra, V. K. Gonugunta, A. Christmann, E. Grill, *Trends Plant Sci.* **15**, 395–401 (2010).
67. A. Mukherjee, S. Dwivedi, L. Bhagavatula, S. Datta, *Plant Cell Rep.* **42**, 829–841 (2023).
68. L. Bohn *et al.*, *Nature* **629**, 1126–1132 (2024).
69. G. Perrella, A. Zioutopoulou, L. R. Headland, E. Kaiserli, *J. Exp. Bot.* **71**, 5247–5255 (2020).
70. D. Coleman-Derr, D. Zilberman, *PLOS Genet.* **8**, e1002988 (2012).
71. S. V. Kumar, P. A. Wigge, *Cell* **140**, 136–147 (2010).
72. S. Cortijo *et al.*, *Mol. Plant* **10**, 1258–1273 (2017).
73. C. Tasset *et al.*, *PLOS Genet.* **14**, e1007280 (2018).
74. L. C. van der Woude *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 25343–25354 (2019).
75. N. Verma *et al.*, *Plant Cell* **36**, 4535–4556 (2024).
76. R. March-Díaz *et al.*, *Plant J.* **53**, 475–487 (2008).
77. Y. Zhu, W. Qian, J. Hua, *PLOS Pathog.* **6**, e1000844 (2010).
78. C. L. Cuevas-Velazquez, J. R. Dinneny, *Curr. Opin. Plant Biol.* **45** (Pt A), 68–74 (2018).
79. Q. Liu, W. Liu, Y. Niu, T. Wang, J. Dong, *Plant Commun.* **5**, 100663 (2024).
80. P. Zhu, C. Lister, C. Dean, *Nature* **599**, 657–661 (2021).
81. C. Xu *et al.*, *Nat. Commun.* **12**, 1790 (2021).
82. W. Tan *et al.*, *Sci. Adv.* **9**, eadh1738 (2023).
83. G. Li, C. Yuan, X. Yan, *Soft Matter* **21**, 1781–1812 (2025).
84. W. Zhang, Z. Li, X. Wang, T. Sun, *Heliyon* **10**, e34035 (2024).
85. K. Watanabe *et al.*, *Nat. Commun.* **12**, 1353 (2021).
86. P. A. Fields, G. N. Somero, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 11476–11481 (1998).
87. M. Griffith, M. W. Yaish, *Trends Plant Sci.* **9**, 399–405 (2004).
88. W. Yamori, K. Hikosaka, D. A. Way, *Photosynth. Res.* **119**, 101–117 (2014).
89. S. Jin *et al.*, *Plant Commun.* **5**, 100814 (2024).
90. G. G. Weber, J. Kortmann, F. Narberhaus, K. E. Klose, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 14241–14246 (2014).
91. P. Broft, R. Rosenkranz, E. Schleiff, M. Hengesbach, H. Schwalbe, *RNA Biol.* **19**, 266–278 (2022).
92. L. D. Vu *et al.*, *Nat. Commun.* **12**, 2842 (2021).
93. C. Lövkvist *et al.*, *eLife* **10**, e66454 (2021).
94. G. Dundar, V. E. Ramirez, B. Poppenberger, *J. Exp. Bot.* **76**, 1970–1977 (2025).
95. R. L. Antoniou-Kourounioti *et al.*, *Cell Syst.* **7**, 643–655.e9 (2018).
96. J. Hepworth *et al.*, *Nat. Commun.* **9**, 639 (2018).
97. Y. Zhao, R. L. Antoniou-Kourounioti, G. Calder, C. Dean, M. Howard, *Nature* **583**, 825–829 (2020).

ACKNOWLEDGMENTS

We thank Y. Guo (IBCAS, Beijing, China) and R. Ganesan, C. Pradhan, and S. Mukherjee (Monash University, Melbourne, Australia) for comments on the manuscript. We apologize to our colleagues whose contributions could not be cited owing to the limited focus of this Review and space limitations. **Funding:** Sam and Nancy Fleming Research Fellowship (A.S.Y.). Australian Research Council Future Fellowship FT190100403 (S.S.). Sir J. C. Bose Fellowship (JCB/2020/000041), Anusandhan National Research Foundation, Government of India (A.K.S.). Australian Research Council Discovery Project DP190101479 (S.K.B.). **Author contributions:** Conceptualization: S.K.B. Funding acquisition: A.S.Y., S.S., A.K.S., S.K.B. Project administration: S.S., A.K.S., S.K.B. Supervision: S.S., A.K.S., S.K.B. Writing – original draft: A.S.Y., S.K.B. Writing – review & editing: A.S.Y., S.S., A.K.S., S.K.B. **Competing interests:** The authors declare that they have no competing interests.

Submitted 28 February 2025; accepted 22 April 2025

10.1126/science.adv5407

REVIEW

Scaling plant responses to heat: From molecules to the biosphere

Margaret E. K. Evans[1]*, Jia Hu[2]†, Sean T. Michaletz[3,4]

Predicting plant responses to rising temperatures, including acute heat waves and hot droughts of varying intensity and duration, is central to addressing the climate and biodiversity crises. However, plant responses to heat are scale-dependent, complicating cross-scale prediction. We highlight recent progress revealing how and why plant responses to heat change across scales, including scales of biological organization and space versus time. We give examples of scaling up from molecular- and leaf-scale data and processes, which are modified by homeostatic and buffering mechanisms at whole plant and ecosystem scales. We show that scaling down—predicting plant responses to warming from broad-scale spatial patterns—can also be misleading, even in direction. Addressing such scale dependencies is essential to improving the prediction of plant responses to heat.

Over the past 50 years, Earth's climate has shifted dramatically, and biodiversity loss has intensified, with rates expected to accelerate in the coming decades (*1*). Plants play a critical role in the global carbon cycle and are a key component of biodiversity supporting terrestrial ecosystems. Thus, understanding how rising temperature affects plants is necessary for predicting climate and biodiversity change. A key problem in meeting that challenge is scaling: the projection or extrapolation of a pattern or X-Y relationship across space, time, or levels of biological organization (from molecules to ecosystems to Earth's biosphere) (Fig. 1). For example, plant responses to heat are often determined at one scale (such as a leaf) or under controlled conditions (such as unlimited soil moisture) and then extrapolated to larger scales (such as ecosystems or the biosphere) (Fig. 1) and uncontrolled conditions. However, plant responses to heat change in magnitude or even direction across scales or contexts, making cross-scale prediction difficult. In this Review, we highlight emerging insights into how and why plant responses to heat shift across scales, focusing on both bottom-up and top-down approaches to cross-scale prediction.

From first principles, a bottom-up, mechanistic approach to "scaling up" the effects of heat on plants is compelling (*2*) but faces substantial challenges (*3*–*6*). Biochemical kinetics govern reaction rates at the molecular scale, yet these rules are modified by other factors at higher biological, larger spatial, and longer temporal scales. Leaf-scale experiments that quantify plant-atmosphere gas exchange [carbon dioxide (CO_2), water (H_2O)] under controlled temperatures provide a powerful starting point for predicting ecosystem processes such as gross primary productivity. But in whole plants, leaf responses are shaped by processes in other organs, and plants have evolved diverse mechanisms and strategies to maintain homeostasis in the face of heat and drought stress. At the ecosystem level, plants modify microclimate by altering heat fluxes through shading, boundary layer effects, and evapotranspiration (*7*–*14*) and actively influence water distribution and cycling, further complicating efforts to upscale molecular and leaf-scale processes to higher levels of organization or larger spatial scales.

In contrast to a bottom-up, mechanistic approach, space-for-time substitution (*15*, *16*) is a "scaling down" approach that relies on statistical regression (Fig. 1). Patterns inferred across spatial climate gradients are used to predict responses to changing climate across time. This practice is common across plant ecology, forestry, biogeography, macroecology, and Earth system sciences when long-term data are unavailable or less available than spatial data. For example, space-for-time substitution is used to infer a species' climatic tolerances from spatial occurrence data and then predict its response to changing climate (known as species distribution modeling or ecological niche modeling). However, some of the processes governing plant responses to warming (for example, adaptive evolution or colonization that allows species to track climate) are likely to lag well behind climate change (*17*, *18*), making space-for-time predictions potentially misleading.

For these reasons and others, predicting plant responses to heat across biological, spatial, and temporal scales is challenging. We highlight select examples of how and why plant responses shift across scales or contexts. We first consider how smaller-scale (such as leaf level) data fail to predict larger-scale (such as ecosystem level) patterns from a mechanistic perspective. We then consider the pitfalls, caused by lagging processes, of predicting change over time from broad-scale spatial patterns. Last, we discuss strategies to improve cross-scale prediction and reduce uncertainty surrounding plant responses to heat.

Scaling up from biochemical kinetics

From first principles, biochemical reaction kinetics provide a strong foundation for predicting the effects of heat on plants. Acute heat exposure causes injury (reversible or irreversible cellular impairment) and necrosis (irreversible cell death), both of which are largely driven by protein denaturation kinetics (*19*, *20*). Thus, rates of protein unfolding and cellular damage increase exponentially with temperature, and the time required to reach a given damage threshold (for example, proportion of cellular processes or tissue affected) decreases exponentially with increasing temperature (*21*). Less severe heat stress and/or shorter exposure times often injure photosynthetic machinery first, causing inactivation of Rubisco (*21*) and impairment of photosystem II (*23*, *24*). These injuries have traditionally been quantified by using threshold temperatures that cause loss of function, with recent studies improving on this by characterizing responses as time-temperature relationships (combinations of exposure duration and temperature that determine functional loss) (*25*). Photosynthetic heat tolerance can increase within hours of exposure to high temperature, driven by the accumulation of heat-shock proteins and metabolites that stabilize proteins (*26*). At higher temperatures and/or longer exposure times, protein denaturation can exceed protective and repair capacities, resulting in irreversible necrosis. In black spruce (*Picea mariana*) seedlings, for example, 50% foliage loss occurred after 110 min at 44°C, but after only 10 min at 50°C (*27*). Such necrosis thresholds vary widely across species (*28*), but spatial patterns of damage are often predictable: Plant parts with greater exposure to solar radiation, such as south-facing aspects or upper canopy positions, tend to be the most severely affected (*29*). The seasonal timing of heat exposure also matters: When extreme heat coincides with bud burst early in the growing season, it can reduce total leaf area and carbon assimilation for the remainder of the season (*30*).

Limits to the predictive power of biochemical kinetics become apparent at larger spatial and higher biological scales. The monotonic Arrhenius model, developed to describe biochemical reaction

[1]Laboratory of Tree-Ring Research, University of Arizona, Tucson, AZ, USA. [2]School of Natural Resources and the Environment, University of Arizona, Tucson, AZ, USA. [3]Department of Botany, University of British Columbia, Vancouver, BC, Canada. [4]Biodiversity Research Centre, University of British Columbia, Vancouver, BC, Canada. *Corresponding author. Email: mekevans@arizona.edu †Present address: Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, USA.

kinetics, has been applied across biological scales—from molecules to the biosphere—based on the hypothesis that higher-level temperature dependencies are governed by biochemical kinetics (*2*, *6*). Indeed, kinetics underpin the Farquhar–von Caemmerer–Berry model of photosynthesis (*31*) used in nearly all global terrestrial vegetation models. However, organismal performance is not monotonic but instead follows a unimodal temperature response ("thermal performance curve") (*32*), increasing exponentially before peaking at some optimal temperature, T_{opt}, above which performance decreases. Fitting the Arrhenius model to biological response data can therefore mischaracterize temperature dependencies, particularly when fitted near or spanning T_{opt} (*5*, *6*). Thus, Arrhenius estimates may misleadingly suggest weak or no temperature dependence of photosynthesis near its optimum, even though photosynthesis exhibits strong temperature dependence when considered across its full temperature range (*33*). Furthermore, acute temperature responses generally do not account for acclimation. Incorporating acclimation into models can improve agreement between predicted and observed temperature dependence across climate gradients and levels of organization (*4*, *34*, *35*). Although biochemical kinetics provide a powerful framework for predicting acute heat responses at the cellular and tissue scales, they are insufficient to fully explain responses across spatial temperature gradients, which also reflect homeostatic responses, local adaptation, and species turnover.

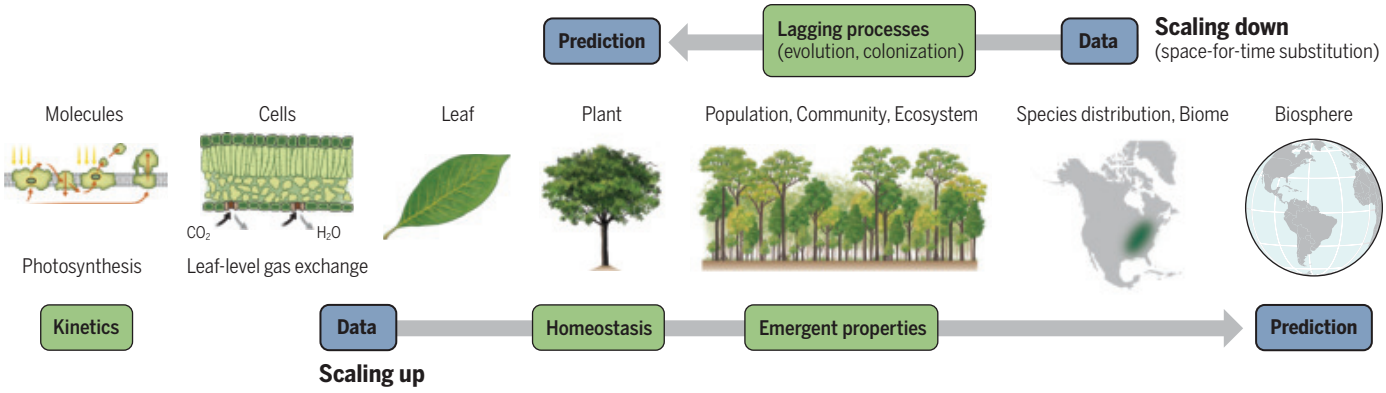
Scaling up from stomatal responses

Leaf-level theory about stomatal responses does not always predict how water use and evaporative cooling vary across contexts or scales, complicating prediction of plant responses to heat. Water loss through stomata (transpiration) plays a major role in regulating leaf temperature and limiting heat injury or necrosis under extreme conditions. Transpiration is also coupled to carbon assimilation because CO_2 enters leaves through stomata as H_2O exits. Stomatal optimization theory hypothesizes that plants regulate stomatal conductance to maximize carbon gain relative to either marginal water loss (*36*) or marginal hydraulic cost (Fig. 2A) (*37*). Stomata are predicted to close at high temperatures, reducing transpiration to either maintain water-use efficiency or minimize hydraulic risk (Fig. 2B). However, emerging evidence shows that stomatal conductance sometimes increases at high temperatures even as photosynthesis declines (*7*, *21*). This suggests that transpiration may be prioritized for leaf cooling to avoid heat stress and necrosis, protecting carbon invested in leaves and enabling continued carbon gain (*6*), rather than optimized for short-term carbon gain (Fig. 2C). This may be particularly pronounced in plants with a stable water supply that enables continued transpiration and evaporative cooling. For example, well-watered poplar trees maintained leaf temperatures 3° to 5°C below air temperature during an extreme heat wave (>48°C) but were unable to do so when water was restricted, causing leaf temperatures to exceed critical thresholds for damage (*24*).

When high temperature does lead to stomatal closure, water loss can shift from stomatal to cuticular pathways (evaporation through the leaf cuticle) (*38*), challenging the view that stomatal responses dominate transpiration under climate extremes. At the canopy scale, total evapotranspiration can remain high even as stomatal conductance or photosynthesis declines (*39*, *40*) because of contributions from cuticular transpiration, soil evaporation, and boundary layer turbulence, resulting in substantial evaporative cooling (*7*). These observations underscore the challenge of scaling stomatal responses from the leaf to whole plants and plant canopies.

Variable responses to heat, aridity, and drought

Predicting plant responses to temperature is further complicated by the great diversity of mechanisms and strategies that plants have evolved to maintain homeostasis. Strong correlations between heat, vapor pressure deficit (VPD), and drought make it difficult to disentangle their individual impacts. Air temperature and humidity together determine VPD, which drives evapotranspiration and hence affects plants both through water supply (available soil water) and demand (transpiration rate). Because saturated vapor pressure increases exponentially with temperature, the evaporative demand of the atmosphere increases nonlinearly with warming. Low soil moisture is often correlated with high VPD, but they can be decoupled during extreme heat waves (*41*). During acute heat waves, leaf necrosis can occur even when soil water is not depleted, indicating that excess heat, not water limitation, is responsible for cell death. In the face of acute heat stress, some plants can adjust their heat tolerance in a matter of hours through biochemical responses (Fig. 2E) (*26*), or buffer against spikes in leaf water potential caused by increases in VPD by using water stored in the stem (*42*, *43*), or mobilize nonstructural carbon for osmoregulation and repairment of embolized tissue (*44*). Over longer timescales, plants more commonly experience sustained periods of concurrent heat and drought, lasting weeks or even years.



CREDIT: ADAPTED BY K. HOLOSKI AND V. PENNEY/*SCIENCE*

Fig. 1. Two approaches to predicting plant responses to heat: Scaling up and scaling down. The first is based on data collected at a lower level of the biological hierarchy, hence a smaller spatial scale and often shorter timeframe, and the second is based on large-scale data collected across spatial gradients of temperature. Barriers to cross-scale prediction modify the pattern observed at the lower scale (for upscaling) or higher scale (for downscaling). In this Review, we highlight how homeostasis and emergent properties complicate upscaling and how lagging processes compromise predictions made with space-for-time substitution.

In response, plants can modify xylem morphology, change leaf area and leaf angle, or decrease the ratio of aboveground to belowground biomass allocation (Fig. 2F). Leaf shedding also can protect plants against hydraulic failure, avoiding extensive damage to carbon-costly organs (*45*, *46*).

Some of the strategies that plants have evolved to maintain homeostasis in response to drought entail trade-offs. For example, during prolonged hot droughts, plants can experience a trade-off between maintaining their thermal safety margin and hydraulic safety margin (*47*). A water transport system (such as a xylem conduit) with a high maximum transpiration rate, that can meet the demands of high evaporative cooling rates, is also more vulnerable to cavitation because embolism (air bubbles that block water flow) is more likely to occur in larger vessels that have larger pit membrane diameters (*48*). In some cases, plants have evolved compensating mechanisms that weaken trade-offs. For example, high leaf mass per area (LMA) is a strong predictor of drought tolerance, and plants with high LMA are generally less efficient at heat loss because of lower transpiration rates, potentially leading to higher mortality risk under heat waves. However, in some plants, high LMA can lead to lower thermal sensitivity (*49*), making thicker leaves actually more resistant to heat stress [but see (*23*)]. In another example, drought-avoidance plants (for example, plants that respond quickly to increasing temperature and VPD by decreasing stomatal conductance) are typically considered less heat or drought tolerant, but some can increase photoprotection by maintaining a higher ratio of carotenoid to chlorophyll content, making them more resistant to heat stress than expected (*50*).

Predicting the impact of heat or drought on plant productivity is further complicated because there can be conditions under which photosynthesis is decoupled from plant growth (*51*). Plant growth, defined as cell division and expansion, is sensitive to changes in turgor pressure, which in turn is sensitive to water potential. In particular, it has been shown that the tension at which cell expansion ceases is substantially less than the tension that induces stomatal closure and limits photosynthetic CO_2 uptake (*52*, *53*). That is, growth can be limited by the environmental conditions that permit cell division and expansion ("sink-limited growth"). By contrast, it is more commonly thought that plant growth is limited by availability of the carbohydrate products of photosynthesis ("source-limited growth"). This source-limited paradigm is at the foundation of most global terrestrial vegetation models (*54*). There is evidence that in cases in which plant growth is source limited (*4*, *55*, *56*), a reduction in photosynthesis in response to heat or drought does lead to a reduction in net primary productivity (*4*). However, under sink limitation, elevated temperature (and drought) may lead to the cessation of growth well before photosynthetic rates cease. This is problematic for scaling because the source-limited paradigm encoded in global terrestrial vegetation models means that they may overestimate

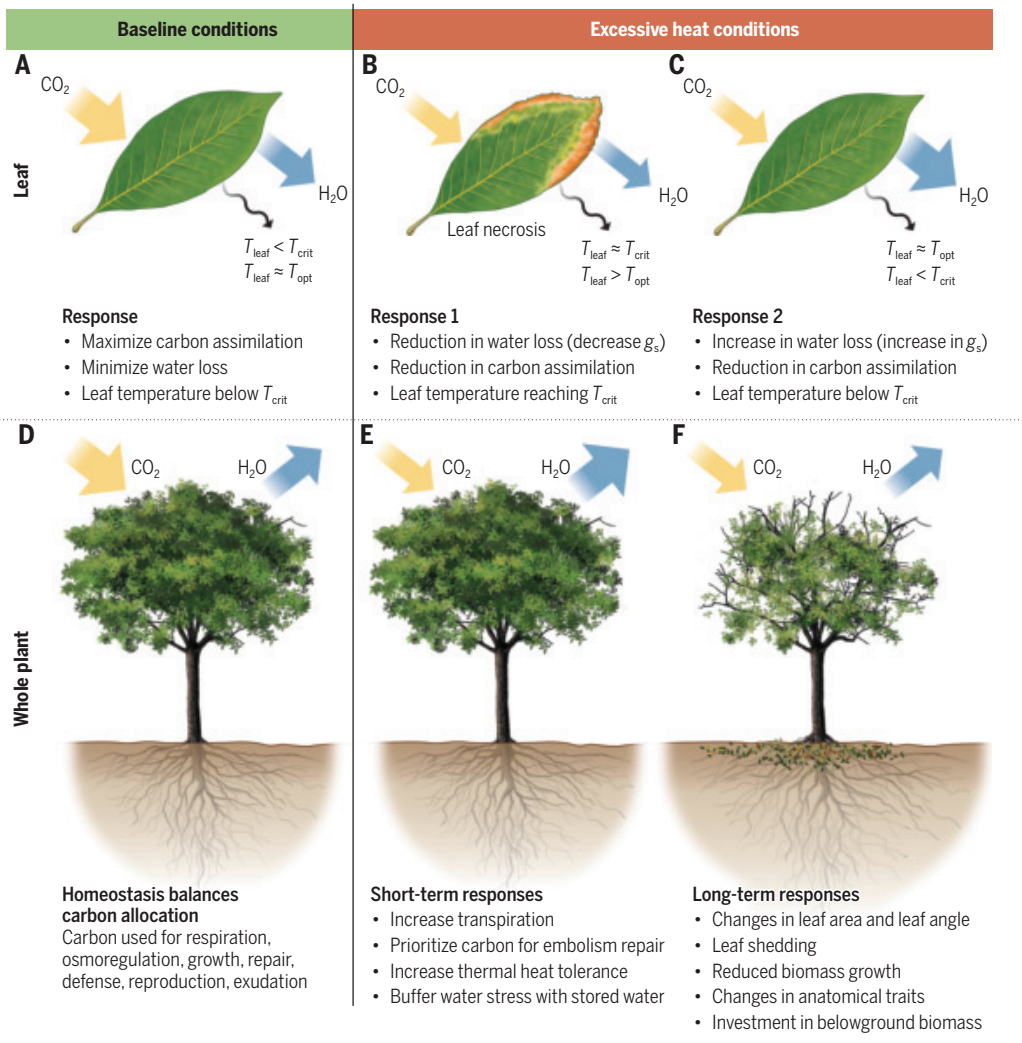


Fig. 2. Leaf and whole-plant responses to excessive heat. (**A**) Plant responses at the leaf scale under baseline conditions are controlled by a balance between carbon gain and water loss, with evaporative cooling acting as a key mechanism to maintain leaf temperature (T_{leaf}) below the optimal temperature (T_{opt}) for biochemical processes. (**B**) Under excessive heat conditions, the expected response is for stomatal conductance (g_s) to decrease, reducing transpiration, to conserve water; this leads to decreases in carbon assimilation and in some cases can bring T_{leaf} close to T_{crit}, which can threaten the physiological function of the leaf and can lead to necrosis. (**C**) Alternatively, in some cases, increased stomatal conductance and transpiration leads to evaporative cooling, which can help T_{leaf} remain below T_{crit} and avoids loss of plant function. (**D**) At the whole-plant scale under baseline conditions, carbon is allocated to support a range of physiological and morphological functions and homeostasis. (**E**) Under acute heat conditions, plant responses include increases in thermal heat tolerance, carbon allocation for maintaining hydraulic function, increases in transpiration rates for cooling, and the use of stored water to buffer against water stress. (**F**) Under longer-term heat or drought conditions, plants can compensate through other morphological changes, including changes in leaf area and leaf angle, leaf shedding, changes in anatomical traits, shifts in the ratio of aboveground versus belowground growth, and reduced growth.

CREDIT: ADAPTED BY K. HOLOSKI/*SCIENCE*

plant productivity under hotter and drier climate futures. Although it is likely that plants operate along a continuum of source and sink limitation (*57*), scaling these processes from organismal to ecosystem scales remains tricky because ecosystem-scale estimates of productivity and carbon sequestration encompass processes such as disturbance and mortality, which are also poorly parameterized in models.

Scaling up from leaves to canopies: Emergent properties

Scaling plant productivity from leaves to landscapes also presents challenges owing to emergent properties of canopy structure and function. For example, leaf temperature, which is a key driver of carbon and water fluxes and thus gross primary productivity, is governed by an energy balance among radiation, convection, and latent (transpiration) heat fluxes. These fluxes are influenced by leaf traits (such as size and stomatal conductance) but also by crown morphology (such as leaf angle and clumping) and canopy complexity (such as porosity and surface roughness). Together, these shape how radiation and airflow penetrate the canopy, influencing vertical temperature profiles and leaf-air temperature relationships. To illustrate how canopy structure affects leaf and air temperatures, we consider radiation regimes and atmospheric coupling across three scenarios: solitary plants, moderately open canopies, and closed canopies.

In solitary plants, solar radiation is relatively uniform with height, whereas longwave radiation from the ground and air temperature both decrease, and wind speed increases, with increasing height. These gradients promote strong atmospheric coupling; hence, leaf temperatures generally decrease with height (*58*). When water is available, leaves with high stomatal conductance may exhibit temperature buffering (limited homeothermy), whereas those with lower conductance or limited water may track air temperature (poikilothermy) or warm faster than air (megathermy) (*10*, *12*, *24*).

In moderately open canopies, such as conifer or mixed temperate forests, crown tapering, spacing, and needleleaf morphology allow solar radiation and air mixing into lower canopy levels. Modeling of conifer canopies suggests that canopy-top leaves exhibit megathermy, transitioning toward poikilothermy lower in the canopy (*59*). However, in a mixed temperate forest, a needleleaf species exhibited poikilothermy in both sunlit and shaded leaves, whereas two co-occurring broadleaf species showed limited homeothermy (*60*). These results suggest that profiles and relationships of leaf and air temperatures in moderately open canopies can vary with plant architecture (for example, conifer versus broadleaf) and canopy structure.

In closed canopies, such as temperate broadleaf or tropical forests, broad, hemispherical, or flat-topped crowns form a continuous upper canopy layer that absorbs much of the incoming solar radiation and restricts atmospheric mixing (the "parasol effect") (*7*, *10*, *12*, *13*). This reduces wind speed and increases boundary layer thickness, further decoupling leaves from air (*13*). In these systems, leaf temperature typically increases with height, with differences of up to 12°C reported from understory to canopy top (Fig. 3) (*12*, *14*, *60*). Canopy-top leaves often exhibit megathermy (Fig. 3D), whereas shaded interior leaves show limited homeothermy or poikilothermy (Fig. 3, E and F). This leaf temperature heterogeneity may complicate bottom-up scaling of gross primary productivity from leaf temperature measurements (Fig. 3C): Thermal cameras, increasingly used to estimate leaf temperature, typically capture canopy-top or peripheral leaves that are sun-exposed and prone to stomatal closure, conditions that promote megathermy (Fig. 3) (*10*, *12*, *14*, *60*). As a result, thermal imaging may overestimate average leaf temperature when primarily sun-exposed leaves are imaged.

Another emergent property of vegetation is its effects on water cycling. Hydraulic redistribution, in which water is moved multidirectionally within the soil profile following water potential gradients,

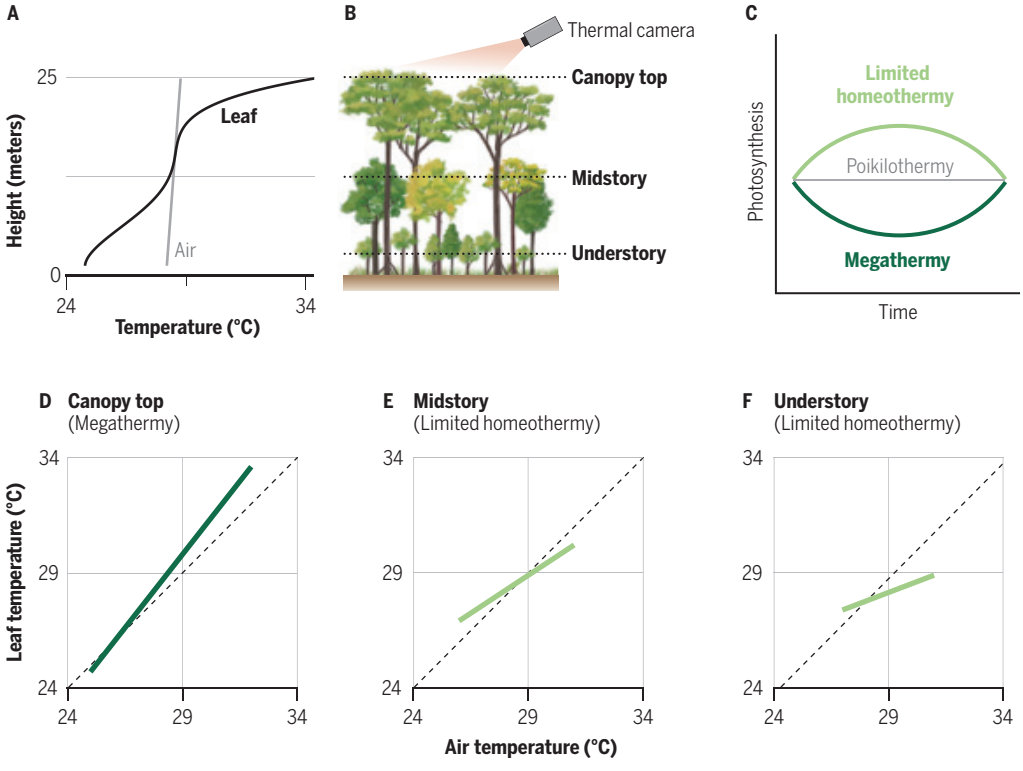


Fig. 3. Microclimate buffering and leaf-air temperature relationships in a tropical forest canopy. (**A**) Typical midday vertical profiles of leaf (black) and air (gray) temperatures. In this example, leaves can be up to 4° to 5°C cooler than air near the ground and up to 7° to 8°C warmer than air at the canopy top. (**B**) Schematic representation of a generalized tropical forest canopy and the field of view of a thermal camera (orange area), illustrating the dominance of canopy-top leaves in thermal imagery captured from above the canopy, as is commonly reported in the literature. (**C**) Hypothesized diurnal photosynthesis for leaves exhibiting different thermal responses: limited homeothermy (light green), poikilothermy (gray), and megathermy (dark green). (**D** to **F**) Leaf versus air temperature relationships at different canopy heights, highlighting contrasting leaf thermal responses, from limited homeothermy to megathermy. In these examples, leaves can be (F) up to 2°C cooler than air in the understory and (D) up to 2°C warmer at the canopy top. Even modest temperature differences can strongly influence plant and ecosystem functioning owing to the exponential temperature sensitivities of biochemical kinetics and VPD, especially when integrated over time. Relationships shown in (A) and (D) to (F) are based on data from different tropical forest canopies reported in (*10*,*12*,*14*); leaf temperatures were measured by using handheld infrared thermometers or fixed infrared sensors mounted within the canopy ~45 cm from foliage, and air temperatures were measured with shielded thermistors.

CREDIT: ADAPTED BY V. PENNEY AND K. HOLOSKI/*SCIENCE*

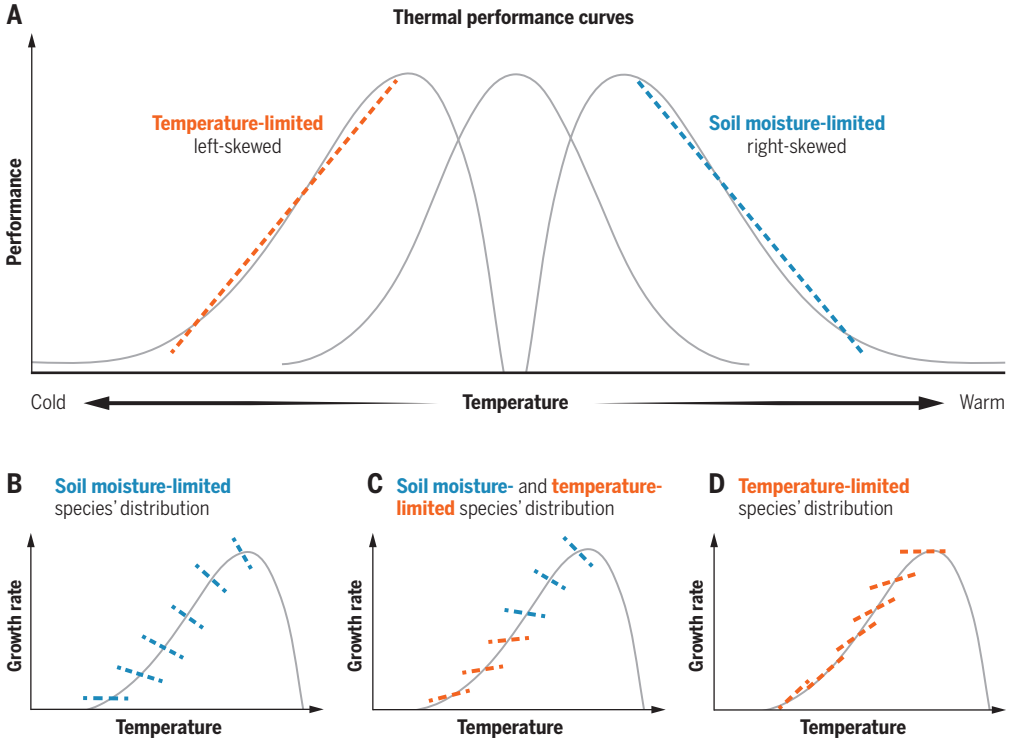
has been observed across climatically diverse ecosystems (*61*). This phenomenon can play a key role in ensuring a reliable supply of water for transpiration during drought and enabling transpirational cooling during short but acute periods of excessive heat. Water cycled through transpiration contributes moisture to the atmosphere and, particularly in tropical or monsoonal regions, can promote regional cloud formation and rainfall (*62*). Evapotranspiration from leaves and soil at the ecosystem scale modifies local to regional microclimate, which can be an influential feedback on stomatal processes, in some cases helping buffer canopies from excessive heat, especially if groundwater is available (*63*). These emergent properties at the whole tree and landscape scale may create ecosystem resiliency by dampening the impacts of excessive heat; however, the degree to which they do largely depends on the severity and length of drought.

Scaling down: Using spatial patterns to predict climate change impacts

The use of spatial climate gradients to predict plant (or ecosystem) responses to changing climate would be valid if the processes responsible for spatial patterns were as rapid as the pace of changing climate. However, there is mounting evidence of contrasting plant responses to temperature variation across space versus time, suggesting that important ecological processes (for example, evolutionary adaptation and colonization of newly climatically suitable habitat) may lag behind the pace of climate change. At the extreme, spatial versus temporal responses are opposite in sign. For example, spatial networks of tree-ring data show that growth of widespread tree species in both eastern and western North America responds positively to spatial temperature variation (trees grow faster at warmer locations) but negatively to time-varying temperature (they grow less in warmer-than-average years) (*64*–*67*). Perret *et al*. (*66*) demonstrated that space-for-time substitution leads to a prediction of increased tree growth with warming, which is contradicted by the observed negative impact of warming on tree growth in the recent past. Similarly, opposite predictions of future productivity, based on the sensitivity of grassland productivity to spatial versus temporal climate variability, were shown across a large area of the intermountain western United States (*68*).

To understand why plant responses to spatial versus temporal temperature variation would be opposite, we first consider thermal performance curves and plant biogeography. Thermal performance curves are unimodal, but not necessarily symmetric (Fig. 4A) (*6*, *32*). Where plant growth is water-limited, plants tend to grow less in a warmer-than-average year, because higher temperatures generally increase VPD and evapotranspiration, intensifying water stress. A negative response to time-varying temperature thus suggests water-limited growth and that the underlying unimodal thermal performance curve is right-skewed (Fig. 4A). Ecosystem ecologists (*69*–*71*) have long established that water is the most common factor limiting terrestrial ecosystem productivity, especially in the temperate zone. At the same time, plant growth rates increase from the poles to the equator or from high to low elevations, largely because of a longer growing season. Plant species distributions in the temperate zone are often large enough to reflect this biogeographical pattern—higher average individual growth rate at the warm edge of a species' distribution and lower average individual growth rate at the cold edge of a species' distribution—while also showing consistently water-limited and hence negatively temperature-sensitive growth (Fig. 4B). For these species, spatial regression of average growth rate with average temperature does not predict how they respond to warming.

Data from plants in boreal or alpine locations, where plant growth can be temperature-limited rather than water-limited, have shown that the response of individual growth to year-to-year variation in temperature can change from positive at the cold edge of a species' distribution to negative at the warm edge (Fig. 4C)



CREDIT: ADAPTED BY V. PENNEY AND K. HOLOSKI/*SCIENCE*

Fig. 4. The biogeography of plant responses to temperature variation. (**A**) Thermal performance curves describe individual plant performance as a function of time-varying temperature. These curves must be unimodal—have a minimum and maximum temperature threshold at which plant performance goes to zero, with maximum performance at some temperature between those bounds. Beyond these constraints, they may be symmetric or skewed. If the curve is skewed, a linear response often suffices to describe variation in individual performance in response to the range of time-varying temperature experienced by plants in the wild, to which they are regularly exposed and adapted (colored, dashed lines). A negative response to time-varying temperature suggests that plant performance is water-limited (red dashed line), whereas a positive response suggests that plant performance is temperature-limited (blue dashed line). (**B** to **D**) Across a plant species' geographic distribution, average growth rate often increases with average temperature, until the species' thermal tolerance limit is reached (solid gray curve). By contrast, the response of individual plant growth to time-varying temperature (dashed lines) may be (B) consistently negative (water limitation), (C) positive in cold parts of the distribution and negative in warm parts of the distribution (temperature and water-limitation, respectively), or (D) consistently negative (temperature limitation).

(*72*–*74*), causing cold-origin populations to benefit from warming and warm-origin populations to decline with warming. Only if plant growth is temperature limited (hence positively sensitive to warming) in every population (Fig. 4D) should we expect space-for-time substitution to be reliable for near-term prediction of plant response to warming. However, even this congruence should be temporary because of the underlying unimodal nature of thermal performance curves (Fig. 4A). Consistent with this, Earth system model simulations predict a widespread shift from energy to water limitation of terrestrial ecosystem productivity with climate change from 1980 to 2100 (*75*), and tree-ring data indicate the geographic area of trees that respond positively to warming has shrunk by 9 million km^2 globally under the relatively mild warming from 1930–1960 to 1960–1990 (*76*). Predicting the impact of anthropogenic warming on plant performance and terrestrial ecosystem productivity on the basis of space-for-time substitution may thus be directionally misleading: Large-scale, biogeographical gradients would suggest a positive response, whereas individual-scale thermal performance curves will lead to a negative response, either immediately (in the water-limited case) or eventually (in the temperature-limited case). Over a longer timescale, plant populations and species are capable of evolving tolerance to warmer temperatures, and communities or ecosystems can adjust to warming through changes in species composition, but evolution and species range shifts are relatively slow processes. Thus, predictions based on spatial patterns are equilibrium predictions, whereas time series data reveal disequilibrium or transient dynamics, and the contrast between the two arises because of slow processes lagging behind the pace of changing climate (*5*, *18*, *67*).

Self-regulating mechanisms at the organism, population, and ecosystem scales may help buffer transient negative impacts of warming. As discussed above, plastic responses to heat or drought stress at intermediate timescales such as leaf shedding or increased biomass allocation to roots may buffer individuals from mortality, and the death of some individuals reduces density-dependent competition for resources among the remaining individuals, potentially increasing their probability of persistence. In the best-case scenario, the intermediate timescales of these additional buffering mechanisms might bridge the (temporal) gap to longer-term evolutionary and species range change, preventing the loss of biodiversity and ecosystem function. Thus, an important knowledge gap to fill is the timescales and limits of these buffering mechanisms, from immediate to intermediate to long-term responses, relative to one another and especially relative to the pace of changing climate. This will help better define which questions can be reliably addressed by using space-for-time substitution and to what extent human intervention (such as assisted evolution or assisted colonization) may be needed to support species and ecosystem responses to climate change.

Conclusions and future directions

These examples demonstrate recent progress in understanding how and why plant responses to heat or rising temperature change across scales, with a few key takeaways. First, individual plants and ecosystems exhibit self-stabilizing feedbacks (homeostasis and emergent properties). Some of these stabilizing mechanisms operate on shorter timescales, such as instantaneous adjustments of transpiration for evaporative cooling, biochemical acclimation of photosynthesis, or canopy-scale buffering of air temperature; others operate on intermediate to longer timescales, from leaf shedding and adjustments of above- to belowground biomass ratio, to adaptive evolution and species range shifts. Each has the potential to modify plant or ecosystem responses to heat, complicating cross-scale prediction.

A second point is that plants have evolved diverse mechanisms or strategies for responding to heat and drought. The tendency has been to focus on a relatively limited set of plant traits or processes (such as stomatal responses or leaf gas exchange), when plants operate in a multidimensional space of traits, trade-offs, and compensatory mechanisms. We recommend considering plant functioning more holistically with respect to plant organs (leaves, stems, and roots); traits; the dynamic and intertwined nature of temperature, water, and carbon status; and the ways that plants modify their environment. A corollary is that this diversity of buffering or self-stabilizing mechanisms likely leads to resilience (*77*), but resilience within limits. Self-stabilizing feedbacks should also generate tipping point or threshold behavior, which creates a great deal of uncertainty in predictions of the future.

Moving forward, given the pitfalls of both bottom-up (mechanistic) and top-down (statistical) approaches to scaling, we suggest that progress can be made by systematically comparing patterns and predictions from data at different scales or by combining complementary sources of data (for example, through hierarchical Bayesian state space models or data assimilation). In addition, empirical or statistical models can be made more mechanistic by adopting causal inference approaches, which entail explicit representation of known or hypothesized cause-effect relationships (*78*–*80*); this should help address the potential for misattribution of causality in spatial regression (space-for-time) approaches, for example. We also suggest hybridizing or blending inferential approaches. Modeling frameworks that are part process-based (include processes known to be important) and part empirical (machine learning or artificial intelligence approaches that excel at describing patterns, without mechanism) offer a promising path forward (*81*).

Last, we recommend exploring additional, currently understudied ways that plant responses to heat and drought may change across scales. The examples highlighted here have been identified largely because of a historical focus on certain scales and processes, as well as the availability of particular types of data and instrumentation used to collect data (such as infrared gas analyzers). This has led to a strong focus on photosynthesis (rather than respiration) and aboveground (not belowground) plant parts and processes. Given their importance, we should also expect soil characteristics (texture, nutrient availability, and organic content), microbial plant associates (such as mycorrhizae), plant pathogens, and natural and anthropogenic disturbances to modify plant responses to heat or play a role in how heat affects plants. Greater attention to these areas is essential for advancing a more complete understanding and better prediction of plant functioning in a changing climate. Pattern and scale has long been considered the central problem in ecology…if not in all of science (*3*). An awareness of the scale-dependence of plant responses to heat is essential to correctly predict and mitigate the climate crisis and the biodiversity crisis, especially given the key role of plants in both.

REFERENCES AND NOTES

1. Core Writing Team, Intergovernmental Panel on Climate Change (IPCC), "2023: Climate Change 2023: Synthesis Report, summary for Policymakers. Contribution of working groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, H. Lee, J. Romero, Eds. (IPCC, 2023); https://doi.org/10.59327/ipcc/ar6-9789291691647.001.
2. J. F. Gillooly, J. H. Brown, G. B. West, V. M. Savage, E. L. Charnov, *Science* **293**, 2248–2251 (2001).
3. S. A. Levin, *Ecology* **73**, 1943–1967 (1992).
4. J. C. Garen, S. T. Michaletz, *Ecol. Lett.* **27**, e70004 (2024).
5. S. Pawar, A. I. Dell, V. M. Savage, J. L. Knies, *Am. Nat.* **187**, E41–E52 (2016).
6. S. T. Michaletz, J. C. Garen, *Ecol. Lett.* **27**, e14381 (2024).
7. B. W. Blonder *et al.*, *New Phytol.* **238**, 2271–2283 (2023).
8. C. L. Kibler *et al.*, *Agric. For. Meteorol.* **339**, 109560 (2023).
9. C. J. Still *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2205682119 (2022).
10. J. C. Garen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2301914120 (2023).
11. P. De Frenne *et al.*, *Glob. Chang. Biol.* **27**, 2279–2297 (2021).
12. B. D. Miller, K. R. Carter, S. C. Reed, T. E. Wood, M. A. Cavaleri, *Agric. For. Meteorol.* **301-302**, 108347 (2021).
13. N. Vinod *et al.*, *New Phytol.* **237**, 22–47 (2023).
14. A. Mau, S. Reed, T. Wood, M. Cavaleri, *Forests* **9**, 47 (2018).
15. C. Damgaard, *Trends Ecol. Evol.* **34**, 416–421 (2019).
16. R. S. L. Lovell, S. Collins, S. H. Martin, A. L. Pigot, A. B. Phillimore, *Biol. Rev. Camb. Philos. Soc.* **98**, 2243–2270 (2023).

17. B. Sandel, C. Merow, J. M. Serra-Diaz, J.-C. Svenning, *J. Ecol.* **113**, 782–794 (2025).
18. P. B. Adler, E. P. White, M. H. Cortez, *Ecography* **43**, 1729–1739 (2020).
19. B. Rosenberg, G. Kemeny, R. C. Switzer, T. C. Hamilton, *Nature* **232**, 471–473 (1971).
20. K. Hüve, I. Bichele, B. Rasulov, U. Niinemets, *Plant Cell Environ.* **34**, 113–126 (2011).
21. R. Teskey *et al.*, *Plant Cell Environ.* **38**, 1699–1712 (2015).
22. S. J. Crafts-Brandner, M. E. Salvucci, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13430–13435 (2000).
23. N. N. Bison, S. T. Michaletz, *New Phytol.* **242**, 1919–1931 (2024).
24. B. C. Posch *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2408583121 (2024).
25. A. M. Cook, E. L. Rezende, K. Petrou, A. Leigh, *Ecol. Lett.* **27**, e14416 (2024).
26. L. Zhu *et al.*, *New Phytol.* **241**, 715–731 (2024).
27. S. J. Colombo, V. R. Timmer, *Tree Physiol.* **11**, 95–104 (1992).
28. A. Guha, J. Han, C. Cummings, D. A. McLennan, J. M. Warren, *Environ. Res. Lett.* **13**, 065008 (2018).
29. C. J. Still *et al.*, *Tree Physiol.* **43**, 203–209 (2023).
30. B. Filewod, S. C. Thomas, *Glob. Chang. Biol.* **20**, 360–371 (2014).
31. G. D. Farquhar, S. von Caemmerer, J. A. Berry, *Planta* **149**, 78–90 (1980).
32. M. J. Angilletta Jr., *Thermal Adaptation: A Theoretical and Empirical Synthesis* (Oxford Univ. Press, 2009).
33. W. Yamori, K. Hikosaka, D. A. Way, *Photosynth. Res.* **119**, 101–117 (2014).
34. D. L. Lombardozzi, G. B. Bonan, N. G. Smith, J. S. Dukes, R. A. Fisher, *Geophys. Res. Lett.* **42**, 8624–8631 (2015).
35. C. M. Zarakas, A. L. S. Swann, C. D. Koven, M. N. Smith, T. C. Taylor, *Glob. Chang. Biol.* **30**, e17449 (2024).
36. B. E. Medlyn *et al.*, *Glob. Change Biol.* **17**, 2134–2144 (2011).
37. J. S. Sperry *et al.*, *New Phytol.* **212**, 577–589 (2016).
38. J. C. Garen, S. T. Michaletz, *New Phytol.* **245**, 1911–1923 (2025).
39. R. M. Marchin, B. E. Medlyn, M. G. Tjoelker, D. S. Ellsworth, *Glob. Chang. Biol.* **29**, 6319–6335 (2023).
40. J. E. Drake *et al.*, *Glob. Chang. Biol.* **24**, 2390–2402 (2018).
41. X. Liu *et al.*, *Glob. Chang. Biol.* **29**, 3072–3084 (2023).
42. D. H. Stout, A. Sala, *Tree Physiol.* **23**, 43–50 (2003).
43. T. Clute, J. Martin, N. Looker, J. Hu, *Oecologia* **197**, 1049–1062 (2021).
44. H. Rennenberg *et al.*, *Plant Biol. (Stuttg.)* **8**, 556–571 (2006).
45. M. T. Tyree, H. Cochard, P. Cruiziat, B. Sinclair, T. Ameglio, *Plant Cell Environ.* **16**, 879–882 (1993).
46. B. T. Wolfe, J. S. Sperry, T. A. Kursar, *New Phytol.* **212**, 1007–1018 (2016).
47. W. M. Hammond *et al.*, *Nat. Commun.* **13**, 1761 (2022).
48. S. Jansen, B. Choat, A. Pletsers, *Am. J. Bot.* **96**, 409–419 (2009).
49. I. K. Münchinger, P. Hajek, B. Akdogan, A. T. Caicoya, N. Kunert, *J. For. Res.* **34**, 63–76 (2023).
50. E. Martínez-Ferri, L. Balaguer, F. Valladares, J. M. Chico, E. Manrique, *Tree Physiol.* **20**, 131–138 (2000).
51. B. Muller *et al.*, *J. Exp. Bot.* **62**, 1715–1729 (2011).
52. T. C. Hsiao, *Annu. Rev. Plant Physiol.* **24**, 519–570 (1973).
53. C. Körner, *J. Ecol.* **91**, 4–17 (2003).
54. S. Fatichi, S. Leuzinger, C. Körner, *New Phytol.* **201**, 1086–1095 (2014).
55. S. Fatichi, C. Pappas, J. Zscheischler, S. Leuzinger, *New Phytol.* **221**, 652–668 (2019).
56. A. Gessler, R. Zweifel, *New Phytol.* **242**, 858–869 (2024).
57. A. T. Trugman, L. D. L. Anderegg, *New Phytol.* **245**, 966–981 (2025).
58. C. Webster, M. Westoby, N. Rutter, T. Jonas, *Remote Sens. Environ.* **209**, 835–847 (2018).
59. K. Ganz, C. J. Still, B. Rastogi, L. M. Moskal, *Agric. For. Meteorol.* **364**, 110456 (2025).
60. K. Yi *et al.*, *J. Geophys. Res. Biogeosci.* **125**, e2020JG005892 (2020).
61. R. B. Neumann, Z. G. Cardon, *New Phytol.* **194**, 337–352 (2012).
62. W. H. Schlesinger, S. Jasechko, *Agric. For. Meteorol.* **189-190**, 115–117 (2014).
63. M. Mu *et al.*, *Earth Syst. Dyn.* **12**, 919–938 (2021).
64. C. D. Canham, L. Murphy, R. Riemann, R. McCullough, E. Burrill, *Ecosphere* **9**, e02368 (2018).
65. S. Klesse *et al.*, *Glob. Chang. Biol.* **26**, 5146–5163 (2020).
66. D. L. Perret, M. E. K. Evans, D. F. Sax, *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2304404120 (2024).
67. M. E. K. Evans *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2315700121 (2024).
68. A. J. Felton *et al.*, *Ecol. Lett.* **25**, 2688–2698 (2022).
69. G. Churkina, S. W. Running, *Ecosystems* **1**, 206–215 (1998).
70. R. R. Nemani *et al.*, *Science* **300**, 1560–1563 (2003).
71. A. Seddon, M. Macias-Fauria, P. Long, D. Benz, K. J. Willis, *Nature* **531**, 229–232 (2016).
72. P. B. Reich *et al.*, *Nat. Clim. Chang.* **5**, 148–152 (2015).
73. J. H. Pedlar, D. W. McKenney, *Sci. Rep.* **7**, 43881 (2017).
74. A. R. Kleinhesselink, P. B. Adler, *Ecology* **99**, 1139–1149 (2018).
75. J. M. C. Denissen *et al.*, *Nat. Clim. Chang.* **12**, 677–684 (2022).
76. F. Babst *et al.*, *Sci. Adv.* **5**, eaat4313 (2019).
77. B. Sakschewski *et al.*, *Nat. Clim. Chang.* **6**, 1032–1036 (2016).
78. J. E. K. Byrnes, L. E. Dee, *Ecol. Lett.* **28**, e70023 (2025).
79. K. Siegel, L. E. Dee, *Ecol. Lett.* **28**, e70053 (2025).
80. J. B. Grace *et al.*, *Ecol. Lett.* **28**, e70029 (2025).
81. C. Shen *et al.*, *Nat. Rev. Earth Environ.* **4**, 552–567 (2023).

ACKNOWLEDGMENTS

We thank C. Still, L. Silva and the CARBS group, J. Needham, and two anonymous reviewers for critical and constructive comments on an earlier version of the paper. **Funding:** S.T.M. was supported by NSERC Discovery. **Author contributions:** M.E.K.E., J.H., and S.T.M. wrote the paper. **Competing interests:** All authors declare that they have no competing interests. **Data and materials availability:** Fig. 3 is based on data from (*9*,*11*,*13*).

Submitted 10 March 2025; accepted 6 May 2025

10.1126/science.adv7597

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

PHOTO: AEL KERMAREC/AFP VIA GETTY IMAGES

PHOTOPHYSICS

A chiral trick by an achiral crystal

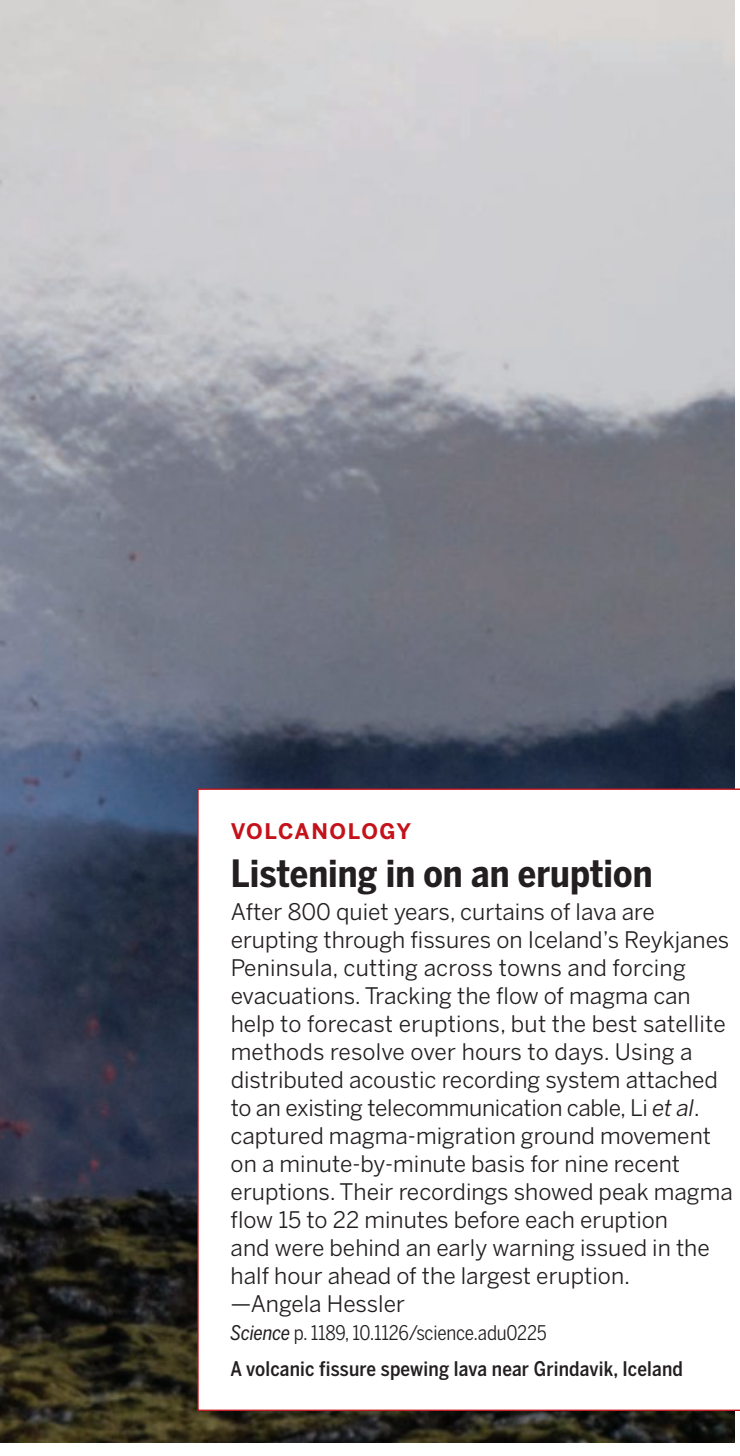
For nearly two centuries, chemists have relied on absorption and rotation of polarized light to distinguish chiral compounds, so much so that the term "optically active" is often used synonymously with the structural absence of mirror or inversion symmetry. Parrish *et al*. now report that a particular centrosymmetric crystalline solid, a lithium cobalt selenium oxide compound, can nonetheless differentially absorb circularly polarized light. They explain the observation as an interference effect between linear dichroism and linear birefringence, and they derive the general symmetry requirements distinguishing the response from conventional optical activity. —Jake S. Yeston

Science p. 1194 10.1126/science.adr5478

VIROLOGY

Revealing the unexpected

The functional annotation of virus genomes is tricky, not least because it can involve culturing viruses, some of which are dangerous to humans. This task is also challenging for computational approaches because it can be difficult to identify virus open reading frames. Weingarten-Gabbay *et al*. devised a mass ribosome profiling method that avoids the hazards of virus culture by using fragments of viral genomes inserted into plasmids that are then transfected into cells. This strategy detects peptides that regulate virus protein expression in ways that modify the course of infection and mediate potent host immune responses, can survey noncanonical translation in many diverse viruses simultaneously, and offers a



VOLCANOLOGY

Listening in on an eruption

After 800 quiet years, curtains of lava are erupting through fissures on Iceland's Reykjanes Peninsula, cutting across towns and forcing evacuations. Tracking the flow of magma can help to forecast eruptions, but the best satellite methods resolve over hours to days. Using a distributed acoustic recording system attached to an existing telecommunication cable, Li *et al*. captured magma-migration ground movement on a minute-by-minute basis for nine recent eruptions. Their recordings showed peak magma flow 15 to 22 minutes before each eruption and were behind an early warning issued in the half hour ahead of the largest eruption. —Angela Hessler

Science p. 1189, 10.1126/science.adu0225

A volcanic fissure spewing lava near Grindavik, Iceland

tool for vaccine development. —Caroline Ash

Science p. 1218, 10.1126/science.ado6670

IMMUNOLOGY

Primed to bind

Humans develop antibodies to dietary antigens upon ingestion. Marini-Rapoport *et al.* investigated the antibody response to one such dietary antigen, the peanut allergen Ara h 2. The authors found that humans frequently carry antibodies that bind to a specific epitope on Ara h 2. These antibodies are convergent because they bound to Ara h 2 in a consistent manner. This structural convergence occurred in these antibodies even when all mutations were removed, supporting the conclusion that the binding interactions are encoded by common antibody genes. These data suggest that humans have a germline antibody repertoire poised to react to peanut, although the implications of this finding for peanut allergy need further investigation. —Courtney Malo

Sci. Transl. Med. (2025) 10.1126/scitranslmed.adw4148

CLIMATE

Balancing act

Climate warms or cools depending on whether the net energy flux from incoming solar radiation is respectively greater or less than that of outgoing long-wave radiation at the top of Earth's atmosphere. Satellite data have shown that an energy imbalance producing warming has strengthened between 2001 and 2023. Myhre *et al*. show that climate models with low climate sensitivity do not reproduce that trend in Earth energy imbalance. Their finding means that increasing concentrations of atmospheric greenhouse gases likely will cause even more warming than most current models predict. —Jesse Smith

Science p. 1210, 10.1126/science.adt0647

MOLECULAR BIOLOGY

G4 resolution through RNA invasion

Genomic DNA typically forms a double-helical structure, but it can also fold into alternative conformations such as G-quadruplex (G4) structures. G4s act as transcriptional regulators, but they can threaten genome integrity if not properly controlled. Using *Xenopus laevis* egg extract and mammalian cell models, Sato *et al*. uncovered an intricate mechanism governing genome-wide G4 landscapes. G4s are recognized as DNA lesions, triggering homology-directed RNA invasion opposite the G4 and forming a "G-loop" structure. Regulated disassembly of this structure leads to G4 resolution and renewal of the RNA-containing strand. This reveals a role for RNA transcripts in restoring genome architecture and maintaining its integrity and function. —Di Jiang

Science p. 1225, 10.1126/science.adr0493

CANCER IMMUNOLOGY

Excommunicating integrin

Macrophages can eliminate cancerous cells through phagocytosis, but tumors are able to evade this process by presenting molecules that prevent engulfment. Zhong *et al.* found that hematological tumors expressing P-selectin glycoprotein ligand 1 (PSGL-1) can evade macrophage-mediated clearance by preventing the integrin signaling that macrophages require for phagocytosis. Genetic deletion of PSGL-1 or a monoclonal antibody blocking human PSGL-1 improved survival in mouse models of leukemia, lymphoma, and multiple myeloma. These findings identify PSGL-1 as a negative regulator of tumor cell phagocytosis and a possible target for cancer immunotherapy. —Claire Olingy

Sci. Immunol. (2025) 10.1126/sciimmunol.adn4302

DEVICE PHYSICS

Doping by donating to a neighbor

Band alignment effects enable high levels of hole doping in a tungsten diselenide bilayer through its transfer of electrons into an adjacent tin disulfide monolayer. Ion implantation is often used to dope in semiconductor films, but this is difficult in few-layer transition metal dichalcogenides. Zhao *et al*. show that tuning of the band offset and charge transfer across the van der Waals interface with an external gate bias can produce a hole density of 1.49×10^{14} per square centimeter, which is about five times the conventional dielectric limit. —Phil Szuromi

Science p. 1183 10.1126/science.adp8444

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

NEUROSCIENCE

Multisensory memories

In everyday life, we easily form speech memories during casual conversations. Remembering what someone said and how they said it requires the brain to form an integrated multisensory representation of speech. Theta oscillations in the neocortex and the hippocampus are important for episodic memory formation. Biau *et al.* showed their subjects movie clips and specifically manipulated the temporal alignment between visual and auditory streams with respect to the dominant theta oscillation. They found that theta synchrony between speech and neuronal activation in neocortex and hippocampus reflected whether audiovisual speech was encoded or not. By contrast, audiovisual asynchrony during encoding affected the accuracy at which speech stimuli were later replayed from memory. —Peter Stern

J. Neurosci. (2025)
10.1523/JNEUROSCI.1797-24.2025

BUTTERFLY GENETICS

Making metallics

One morph of the Mormon fritillary butterfly (*Speyeria mormonia*) bears silver spots on its wings, which is in contrast to the beige spots in the alternate morph. Livraghi *et al*. used genetic data and knockdowns to determine the genetic basis of this trait. They found that silver iridescence was autosomal recessive, and statistical associations identified a haplotype near the pigmentation gene *optix*. RNA interference knockdowns of *optix* produced silver spots, suggesting that this haplotype likely downregulates this transcription factor in silver butterflies. Curiously, the dominant unsilvered haplotype appears to have undergone repeated instances of positive selection and is shared between species. The reason behind this is unclear. —Corinne Simonti

Curr. Biol. (2025)
10.1016/j.cub.2025.03.028

NEUROSCIENCE

Why is jogging an antidepressant?

Physical exercise can reduce depressive symptoms, and several brain regions have been implicated in this effect. However, the neural circuit mechanisms underlying this antidepressant effect are not yet known. Xia *et al*. used a chronic stress model to induce depressive-like behaviors in mice and found that daily treadmill exercise prevented these behaviors. They observed that oxytocin-releasing projections from the paraventricular nucleus to the nucleus accumbens were disrupted by stress and restored by exercise. Experimental inhibition of this circuit blocked the antidepressant effects of exercise, whereas activation prevented stress-induced depressive behaviors. The findings indicate that the oxytocinergic connection between these two brain regions is required for the antidepressant action of exercise and could be a future therapeutic target. —Sarah Lemprière

Proc. Natl. Acad. Sci. U.S.A. (2025)
10.1073/pnas.2503675122

MICROBIOME

Complicated picture for aging mice

Many studies of the gut microbiome are performed in laboratory mice, but we still have an incomplete picture of how the microbiome is influenced by age and host genetics. Studying more than 900 genetically diverse adult mice over 6 months and with five different dietary groups, Litichevskiy *et al.* found that there were microbiome changes with age, but the only community feature associated with age was uniqueness, which could be explained by accumulating stochastic events rather than selection. Diet and host genetics both played a role in the abundance of many microbial genera, and there were correlations between several health parameters and microbiome composition. However, although dietary restriction was associated with prolonged lifespan of mice, the effects were likely not due to microbiome changes. —Michael A. Funk

Nat. Microbiol. (2025)
10.1038/s41564-025-01963-3

NEUROSCIENCE

Targeting interneurons

Interneurons, the inhibitory cells of the brain, are composed of many subclasses and play a major role in all brain functions. Furlanis *et al.* analyzed single-cell genomic data and identified specific enhancers that were able to target single interneuron subtypes when delivered using recombinant adeno-associated viral vectors. Using this new tool in combination with different effectors, the authors were able to target, observe, and manipulate specific interneuron populations in mice in vivo, as well as in brain tissue from nonhuman primates and humans. This toolset can thus be used for understanding the role of specific brain cell types across multiple mammalian species. —Mattia Maroso

Neuron (2025)
10.1016/j.neuron.2025.05.002

MATERIALS SCIENCE

Expanding cementitious precursors

Clinker, a key intermediate product of cement, is formed by an energy-intensive process and contributes to greenhouse gas emissions. Substitute materials, such as coal fly ash and blast furnace slag, are available but have supply constraints. Mahjoubi *et al.* identified potential secondary and natural cementitious precursors using a large language model. They predicted the reactivity metrics of 14,000 materials extracted from 88,000 academic papers and showed that various waste products and certain igneous rock types have sufficiently high reactivity for cement production. Experimental validation of the identified candidate materials could expand the repertoire of sustainable cementitious precursors. —Sumin Jin

Commun. Mater. (2025)
10.1038/s43246-025-00820-4

ORGANIC CHEMISTRY

Palladium goes radical in the dark

Palladium catalysis is central to carbon-carbon coupling in pharmaceutical synthesis. Decades of studies support mechanisms in these reactions that operate through two-electron oxidative addition and reductive elimination steps in the metal coordination sphere. In recent years, photoexcitation has opened a complementary pathway that accesses Pd(I) alkyl radical intermediates. Singh *et al.* report an intriguing example implicating Pd(I) radicals in the dark as well. The room-temperature reaction couples bromoamides with dienes or enynes to produce lactams and appears to proceed by halogen atom transfer and radical-polar crossover. –Jake S. Yeston

Angew. Chem. Int. Ed. (2025)
10.1002/anie.202503446

RESEARCH ARTICLE SUMMARY

BACTERIAL IMMUNITY

Cat1 forms filament networks to degrade NAD^+ during the type III CRISPR-Cas antiviral response

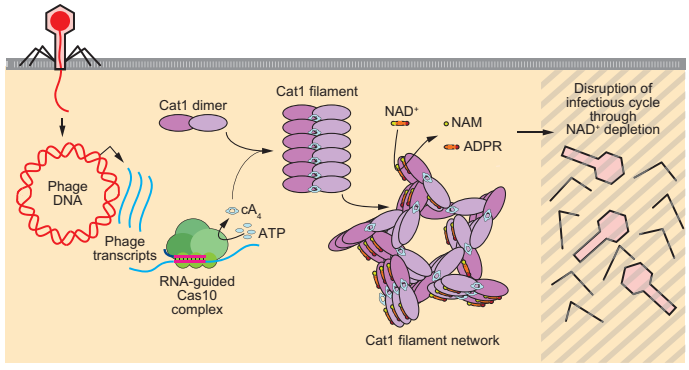
Christian F. Baca†, Puja Majumder*†, James H. Hickling, Dinshaw J. Patel*, Luciano A. Marraffini*

Full article and list of author affiliations: https://doi.org/10.1126/science.adv9045

INTRODUCTION: CRISPR-Cas systems defend bacteria and archaea against viral (phage) infection using RNA-guided complexes that can be classified into different types depending on their molecular mechanism of immunity. In many cases, such as in type II CRISPR-Cas systems, the Cas9 ribonucleoprotein complex directly destroys the phage genome to provide immunity. For type III CRISPR-Cas systems, the Cas10 complex recognizes phage transcripts produced during infection that are complementary to the RNA guide. Target recognition triggers the cyclase activity of the Cas10 complex, which synthesizes cyclic oligoadenylates (cOAs) that act as second messengers to stimulate downstream CRISPR-associated Rossmann-fold (CARF) effectors. These effectors are composed of (i) a CARF domain that dimerizes to generate a cOA binding pocket and (ii) additional domains with enzymatic activities that, upon ligand binding, alter the host to interrupt the phage infectious cycle. Type III CRISPR-Cas systems are associated with an immense repertoire of CARF effectors that display diverse activities such as DNA and/or RNA cleavage, membrane depolarization, and adenosine triphosphate (ATP) deamination. Many other CARF effectors, however, remain uncharacterized.

RATIONALE: We studied an effector containing a CARF domain fused to a Toll/interleukin-1 receptor (TIR) domain, Cat1. Previous work showed that TIR domains in bacteria and plants cleave oxidized nicotinamide adenine dinucleotide (NAD^+) molecules. We expressed Cat1 in staphylococci carrying a type III-A CRISPR-Cas system and evaluated the effects of its activation and antiphage properties in vivo. We also purified the protein to determine cOA binding, enzymatic properties, and cryo–electron microscopy structure.

RESULTS: Upon activation of cOA production, NAD^+ is substantially depleted in staphylococci expressing Cat1, leading to a stop in cell division and an arrest of the growth of the bacterial culture. Cat1's effects on infected hosts interrupt the viral lytic cycle and prevent phage propagation, thus enabling the continuous growth of uninfected cells. In vitro, purified Cat1 binds cyclic tetra-adenylate (cA_4), which activates the enzymatic cleavage of NAD^+ into nicotinamide (NAM) and adenosine diphosphate ribose (ADPR). Alanine substitutions of residues lining both the cA_4 binding pocket and the NAD^+ cleavage site abrogate Cat1-mediated growth arrest. Upon ligand binding, Cat1 dimers stack upon each other to generate long filaments that are maintained by the bound cA_4. Stacked TIR domains, which are radially arranged in the filament, form the NAD^+ cleavage catalytic sites. Furthermore, Cat1 filaments can associate through backbone interactions between the CARF domains of one filament and the TIR domains of the adjacent filament to assemble into large networks formed by trigonal and pentagonal filament bundles. Deletion of the residues involved in this interaction disrupted the formation of the filament network and dramatically reduced NAD^+ cleavage.



Cat1 filament networks provide immunity against phage infection. During the type III CRISPR-Cas response, recognition of a target phage transcript by the guide RNA of the Cas10 complex triggers the synthesis of cyclic tetra-adenylates (cA_4) that bind to CARF effectors such as Cat1. Upon cA_4 binding, Cat1 dimers stack on top of each other to form filamentous structures that further assemble into large networks formed by trigonal and pentagonal filament bundles. The TIR domains within these filaments degrade the essential metabolite NAD^+ to generate an inhospitable host incapable of supporting phage replication.

CONCLUSION: Other prokaryotic defense systems also trigger filamentation of TIR domains to degrade NAD^+ and prevent phage propagation, for example the TIR-SAVED and TIR-STING effectors associated with cyclic oligonucleotide–based antiphage signaling systems (CBASS). Whereas these effectors share a filament architecture in which the cyclic nucleotide–binding domains (SAVED and STING) are arranged on one side of the filament and the TIR catalytic domains are displayed on the other side of the filament, in Cat1 the cyclic nucleotide–binding domains are at the center of the filament with the catalytic TIR domains radially arranged. In addition, Cat1-mediated NAD^+ cleavage presents a novel chemistry (NAD^+ cleavage) not previously associated with the CRISPR-Cas response. The structure of Cat1 is also distinctive. First, in contrast to other CARF effectors, the cA_4 binding pocket is formed by not one but two CARF dimers in which the bottom CARF dimer forms the base of the pocket and the top CARF dimer caps the ligand. Therefore, cA_4 binding facilitates the stacking of CARF dimers into a long filament. Second, filaments assemble into large networks formed by trigonal and pentagonal filament bundles that dramatically enhance NAD^+ cleavage. How this network increases the rate of NAD^+ degradation is currently not known. Our results reveal that the CRISPR-Cas response, once thought to act exclusively on nucleic acids, can be mediated by a diverse range of molecular mechanisms. □

*Corresponding author. Email: majumdp@mskcc.org (P.M.); pateld@mskcc.org (D.J.P.); marraffini@rockefeller.edu (L.A.M.) †These authors contributed equally to this work.
Cite this article as C. F. Baca *et al.*, *Science* **388**, eadv9045 (2025). DOI: 10.1126/science.adv9045

RESEARCH ARTICLE SUMMARY

SPECIATION

Introgression dynamics of sex-linked chromosomal inversions shape the Malawi cichlid radiation

L. M. Blumer†, V. Burskaia†, I. Artiushin†, J. Saha†, *et al.*

Full article and list of author affiliations: https://doi.org/10.1126/science.adr9961

INTRODUCTION: Ecological speciation is responsible for much of the biodiversity on our planet. Despite its fundamental importance, this process, in which new species emerge through evolutionary adaptation to distinct ecological niches, is still not fully understood. Intriguing case studies are adaptive radiations, bursts of ecological speciation that give rise to large numbers of diverse species over timescales that are short compared with the fixation time for new genetic variants. Genome sequencing studies increasingly point towards the importance of hybridization and cross-species gene flow in producing the diversity needed for ecological speciation and adaptive radiation. However, a major conundrum is the role of meiotic recombination in this process. On the one hand, recombination can create new, beneficial combinations of genetic alleles. On the other, it breaks down co-adapted allelic combinations, impeding speciation.

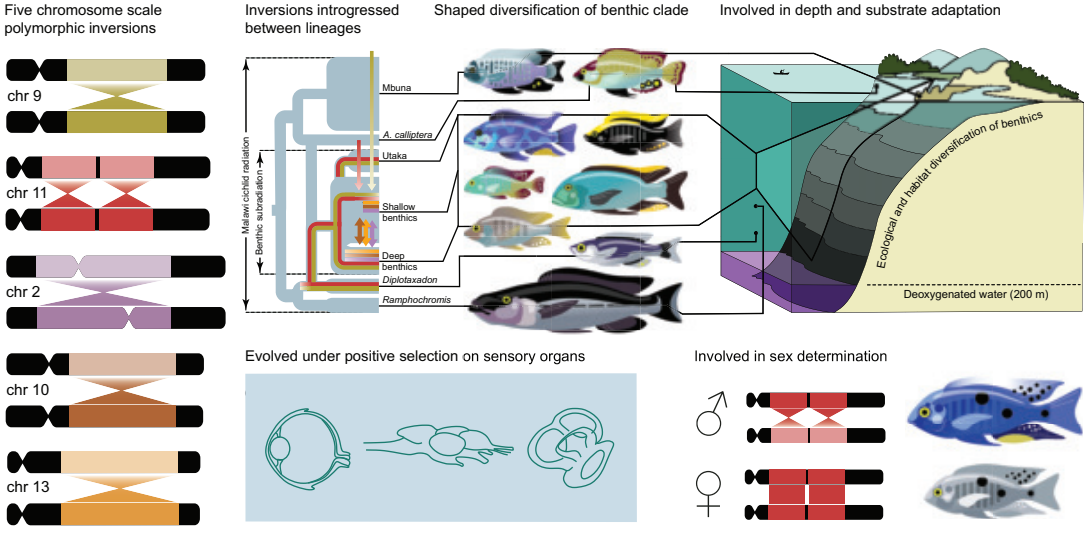
RATIONALE: Chromosomal inversions, stretches of DNA that are flipped in their orientation, provide a potential solution to the conflicting roles of recombination in ecological speciation. This is because inversions show suppressed recombination with the original chromosomal orientation, enabling them to lock together adaptive combinations of alleles in so-called "supergenes." Inversions have been found to be important in ecological adaptation and speciation in many groups of organisms, but so far, there has been little evidence for their significance in adaptive radiations. To address this gap, we systematically investigated the presence and role of inversions across the lake Malawi cichlid fish adaptive radiation, the largest recent vertebrate radiation.

RESULTS: The genomes of 1375 Malawi cichlids from 240 species revealed the presence of multiple chromosomal inversions. The five largest of these segregate within the diverse and species-rich benthic subradiation, with a strong association between inversion states and habitat depth. Phylogenetic tracking of inversion states revealed a hybrid origin of the benthic clade, along with several introgression events transporting inversions and other genetic material between lineages within and outside of the radiation. Inversion haplotypes showed strong signals of adaptive evolution, including being enriched for sensory functions, behavior, and reproduction. For three chromosomes, the re-introgression of haplotypes of the ancestral orientation into benthic lineages coincides with an apparent Y chromosome–like role of this haplotype in the sex determination of some benthic species but not others.

CONCLUSION: The spread of chromosome-scale inversions in Malawi cichlids coincided with the phenotypic and ecological diversification of benthic species across habitats, with evidence for a role of inversion haplotypes in ecological adaptation. The additional transient sex linkage of introgressed inversion-region haplotypes points to an interplay of sex-linked and natural selection in shaping the evolution of inversion haplotypes and the diversification of cichlids. □

Corresponding author: H. Svardal (hannes.svardal@uantwerpen.be); R. Durbin (rd109@cam.ac.uk) †These authors contributed equally to this work. **Cite this article as** L. M. Blumer *et al.*, *Science* **388**, eadr9961 (2025). DOI: 10.1126/science.adr9961

Five large chromosomal inversions contribute to the diversification of Malawi cichlids. Inversions established in the diverse benthic subradiation. Inversion-region haplotypes were exchanged through hybridization of lineages within and outside of the Malawi radiation and contribute to ecological and habitat divergence, sensory adaptation, and sex determination.



RESEARCH ARTICLE SUMMARY

CANCER GENOMICS

Aberrant basal cell clonal dynamics shape early lung carcinogenesis

Sandra Gómez-López *et al.*

Full article and list of author affiliations: https://doi.org/10.1126/science.ads9145

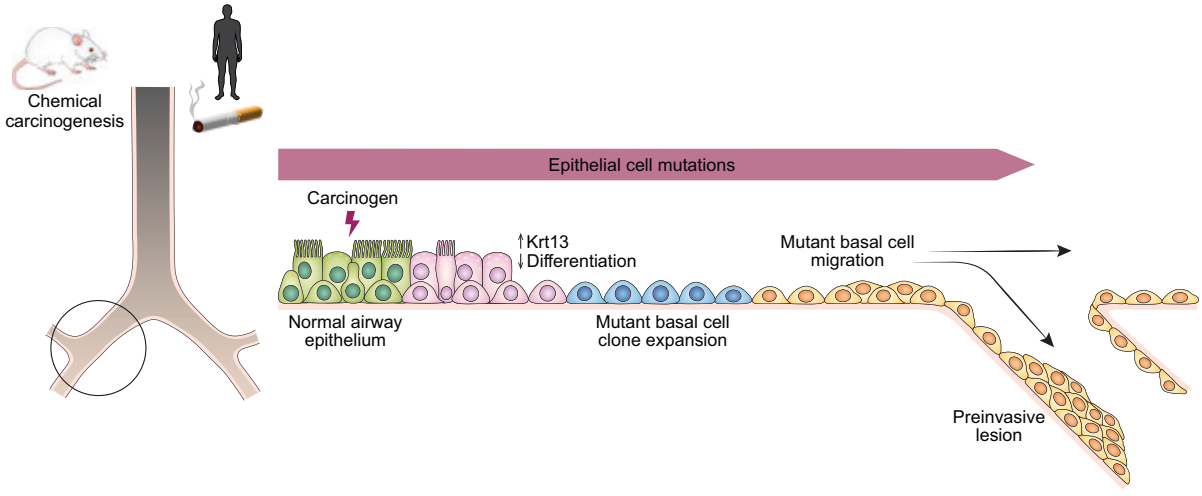
INTRODUCTION: Lung cancer remains the leading cause of cancer-related deaths worldwide. Lung squamous cell carcinoma is the second most common subtype and develops in a stepwise process from increasingly disorganized preinvasive lesions in the bronchial epithelium. The biological mechanisms underlying the transition of the normal epithelium to precancerous states are poorly understood.

RATIONALE: The normal tracheobronchial epithelium is maintained by basal cells, which divide to self-renew and to produce differentiated luminal cells. Basal cells naturally accumulate mutations throughout an individual's lifetime. Exposure to mutagens, such as those found in tobacco cigarette smoke, increases not only the mutational burden in basal cells but also the incidence of cancer-associated mutations. We set out to elucidate the cellular processes that lead to the formation of precancerous lesions by tracking basal cell trajectories in a carcinogen-driven mouse model of lung squamous cell cancer and by delineating cell state changes in the normal airway epithelium of current smokers. To better understand the early evolutionary dynamics of lesion precursors, we investigated clonal relationships between spatially separate lesions in the bronchial tree.

RESULTS: Using genetic-lineage tracing, we have demonstrated that preinvasive lesions originate from basal cells. In mice, we found that carcinogen exposure results in non-neutral competition between basal cell clones, which eventually drives mutant clone dominance in the bronchial tree. This change in clonal dynamics is associated with a shift in basal cell fate, which has also been identified in the airway of human smokers. This leads to disturbed basal cell homeostasis, an increase of cells in transitional state, and decreased luminal differentiation. In mouse and human, we found clonally related lesions across distinct anatomical sites, suggesting aberrant migration of cancer precursor cells.

CONCLUSION: Our work identifies a cell fate shift in airway basal cells as they transition into precancerous states and highlights mutant basal cell clone expansions and aberrant cell migration as key events during the initiation of lung squamous cell carcinogenesis. This provides a conceptual framework for future investigations of the biology of precancerous lesions and may open new avenues for early cancer interception. □

Corresponding author: Sam M. Janes (s.janes@ucl.ac.uk) Cite this article as S. Gómez-López *et al.*, *Science* **388**, eads9145 (2025). DOI: 10.1126/science.ads9145



Disruption of basal cell homeostasis drives early lung carcinogenesis. Carcinogen exposure leads to cell fate changes in airway basal cells. Subpopulations of mutant basal cells may preferentially expand symmetrically, leading to loss of differentiated luminal cells and mutant cell clonal expansions. Mutant basal cell clones can migrate along the bronchial tree, where they eventually produce preinvasive squamous cell lesions.

RESEARCH ARTICLE SUMMARY

ENZYMOLOGY

Evolutionary-scale enzymology enables exploration of a rugged catalytic landscape

Duncan F. Muir, Garrison P. R. Asper, Pascal Notin, Jacob A. Posner, Debora S. Marks, Michael J. Keiser, Margaux M. Pinney*

Full article and list of author affiliations: https://doi.org/10.1126/science.adu1058

INTRODUCTION: Enzymes catalyze the reactions underlying virtually all biological processes, yet our understanding of how sequence variation translates to enzymatic function remains incomplete. Enzyme sequence-function relationships are often viewed as a "landscape," with mutational "walks" tracing paths across peaks and valleys in catalytic performance. Most experimental studies, however, focus on narrowly defined regions of sequence space through mutagenesis. In contrast, genomic databases contain widespread natural sequence variation that could alter catalytic parameters across diverse environments (e.g., temperature). Despite this wealth of natural sequence data, large-scale quantitative measurements of catalytic constants under consistent conditions remain scarce. This lack of data hinders the development of predictive models and leaves the global topology of sequence-catalysis landscapes underexplored.

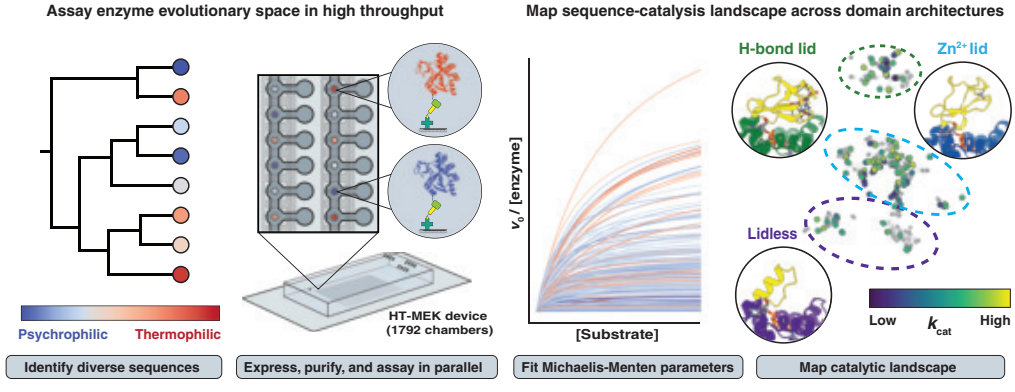
RATIONALE: In this study, we investigated the sequence-catalysis landscape of the model enzyme adenylate kinase (ADK) at the evolutionary scale, sampling broadly across bacterial and archaeal phylogeny. As adaptation to different temperature environments has been suggested to underlie differences in catalytic rates for ADK, we gathered hundreds of orthologs and mutants of ADK spanning the coldest and hottest environments on Earth. Adapting a high-throughput microfluidic platform, HT-MEK, we systematically measured the Michaelis-Menten parameters k_{cat} (catalytic constant), K_M (Michaelis constant), and k_{cat}/K_M (catalytic efficiency) for these ADK sequences under consistent conditions. We further analyzed the organization and traversability of this landscape and quantified how different temperature environments shape ADK function. We then evaluated how well unsupervised deep-learning models—trained solely on protein sequence data—capture these empirical relationships and developed machine-learning models trained on our kinetic dataset to predict the catalytic parameters of these naturally occurring ADKs.

RESULTS: Our high-throughput kinetic data revealed that ADK orthologs vary in k_{cat} by up to three orders of magnitude—even with conserved active sites and similar predicted structures. We address long-standing evolutionary hypotheses demonstrating that thermophilic enzymes are not universally slower than their mesophilic counterparts. We dissect the topology and navigability of this sequence-catalysis landscape, showing that it is rugged, with at least three global neighborhoods organized by distinct domain architectures. Stepwise point mutations and domain swaps show that this landscape remains navigable over long evolutionary timescales through path-dependent mechanisms. Finally, we show that an unsupervised protein language model organizes ADK sequence space by structure but not catalytic activity. k_{cat} prediction models trained on the dataset collected herein outperform prior models trained only on public databases.

CONCLUSION: Combining high-throughput kinetic measurements of natural enzyme sequences with machine learning, we charted an evolutionary-scale sequence-catalysis landscape for a ubiquitous enzyme family. Our data show that high catalytic activity can arise through multiple structural solutions and is not strictly limited by thermal adaptation, challenging assumptions about universal activity–stability trade-offs. Whereas protein language models excel at capturing broad structural features, our results emphasize that experimental annotations—especially for catalytic activity—are critical for building accurate sequence-to-function models. More broadly, coupling high-throughput enzymology with machine and deep learning may uncover insights into enzyme evolution, guide the rational design of novel enzymes, and illuminate the fundamental constraints that shape protein function across the tree of life. □

*Corresponding author. Email: mpinney@berkeley.edu Cite this article as D. F. Muir *et al.*, *Science* **388**, eadu1058 (2025). DOI: 10.1126/science.adu1058

Leveraging evolutionary-scale enzymology to map sequence-catalysis landscapes. Sampling enzyme sequences across phylogeny and diverse environments reveals multiple evolutionary solutions to high catalytic activity arising within different domain architectures. Although protein language models capture sequence-encoded structural organization, the catalytic landscape remains highly rugged.



RESEARCH ARTICLE SUMMARY

MALARIA

A metabolite-based resistance mechanism against malaria

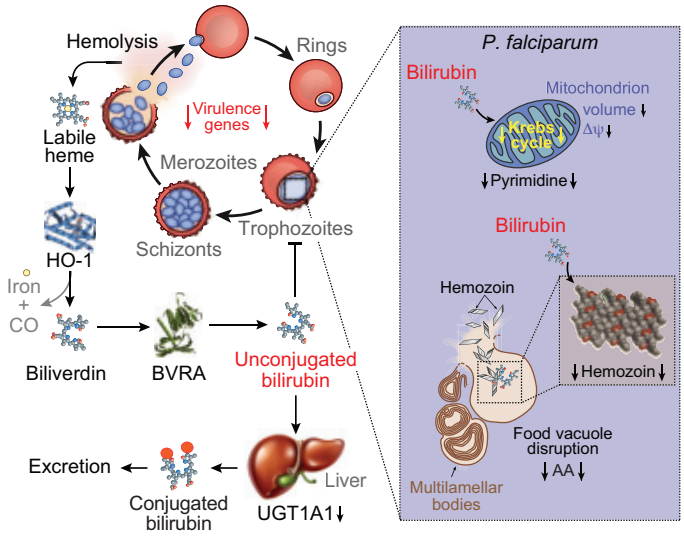
Ana Figueiredo *et al.*

Full article and list of author affiliations: https://doi.org/10.1126/science.adq6741

INTRODUCTION: Jaundice arises when bilirubin, a yellow pigment, accumulates in plasma and gives a yellowish color to the skin and the sclera (the white portion of the eyeball). Bilirubin has long been considered as a "waste product" of heme catabolism. Because of its lipophilic nature, bilirubin excretion requires conjugation to glucuronic acid through a reaction catalyzed in hepatocytes by UDP glucuronosyltransferase family 1 member A1 (UGT1A1). The less toxic water-soluble conjugated bilirubin is then excreted via the bile into the intestinal lumen. Because bilirubin conjugation occurs in the liver, its accumulation in plasma is a reliable biomarker of liver dysfunction. Although accurate, this has led to the widespread perception of jaundice being a maladaptive and eventually pathogenic response. However, several investigators have shown that bilirubin participates in various activities, acting as a lipophilic antioxidant and presumably as a ligand of receptors involved in different metabolic functions.

RATIONALE: Jaundice is a common presentation of malaria, the ancestral infectious disease caused by parasites from the *Plasmodium* genus. These parasites evolved to invade and proliferate inside the red blood cells of their hosts, causing hemolysis and the accumulation of extracellular hemoglobin in plasma. When the prosthetic heme groups of hemoglobin are detached from the globin chains of hemoglobin, there is an accumulation of labile heme in plasma, an independent risk factor for *Plasmodium falciparum* malaria severity. Survival from experimental malaria is contingent on the capacity of the infected host to catabolize heme into biliverdin, the substrate used by biliverdin reductase to produce bilirubin. This raised the hypothesis that bilirubin production by biliverdin reductase participates in a metabolism-based defense strategy against malaria.

RESULTS: Using a highly specific approach to measure bilirubin in plasma, we found a correlation between the levels of circulating unconjugated bilirubin and the onset of symptomatic *P. falciparum* malaria in humans. We established that bilirubin is protective against malaria in an experimental model of malaria in mice, where repressing bilirubin production through genetic loss of function of biliverdin reductase precipitated malaria mortality. This lethal phenotype could be reversed by the administration of bilirubin, verifying that unconjugated bilirubin can be protective against experimental malaria. Repression of bilirubin conjugation by hepatic UGT1A1 was also protective against experimental malaria in mice, further supporting the protective effect of unconjugated bilirubin against malaria. Using several orthogonal approaches in vivo and in vitro, we found that unconjugated bilirubin targets *Plasmodium* inside the red blood cell to repress its proliferation and virulence. Bilirubin targets the parasite's mitochondrion and simultaneously interferes with heme detoxification, disrupting the parasite food vacuole and therefore inhibiting the acquisition of essential amino acids from hemoglobin.



Antimalarial effect of bilirubin. Malaria is associated with hemolysis and labile heme accumulation, which is catabolized into biliverdin by heme oxygenase-1 (HO-1) and converted into bilirubin by biliverdin reductase A (BVRA). Repression of bilirubin conjugation by UGT1A1 increases unconjugated bilirubin to target the *Plasmodium* mitochondrion and inhibit hemozoin crystallization, compromising the parasite's food vacuole and amino acid (AA) acquisition. $\Delta\psi$, mitochondrial membrane potential.

CONCLUSION: The induction of bilirubin production and inhibition of its conjugation in response to *Plasmodium* spp. infection is an evolutionarily conserved resistance mechanism against malaria. Presumably, this metabolism-based defense strategy has a major evolutionary trade-off, namely, the insidious incidence of neonatal jaundice, which can potentially damage neurons in the brain. To what extent this defense strategy can be targeted therapeutically to overcome the enormous burden imposed by malaria on human populations remains to be established. □

Corresponding author: Miguel P. Soares (miguel.soares@gimm.pt) Cite this article as A. Figueiredo *et al.*, *Science* **388**, eadq6741 (2025). DOI: 10.1126/science.adq6741

RESEARCH ARTICLE SUMMARY

ELECTROCHEMISTRY

Acid-humidified CO_2 gas input for stable electrochemical CO_2 reduction reaction

Shaoyun Hao†, Ahmad Elgazzar†, Shou-Kun Zhang, Tae-Ung Wi, Feng-Yang Chen, Yuge Feng, Peng Zhu, Haotian Wang*

Full article and list of author affiliations: https://doi.org/10.1126/science.adr3834

INTRODUCTION: Electrochemical carbon dioxide reduction reaction (CO_2RR) enables efficient carbon utilization by converting CO_2 into valuable products using renewable energy. Among CO_2RR reactor designs, zero-gap membrane electrode assembly (MEA) electrolyzers stand out for their ability to achieve high current densities and energy efficiency. These electrolyzers typically incorporate gas diffusion electrode (GDE) for efficient CO_2 mass transport and an anion exchange membrane (AEM) that separates the electrodes. However, their commercialization is limited by operational instability, primarily because of salt formation in the cathode chamber. This obstructs CO_2 transport, causes flooding, and ultimately leads to electrolyzer failure. Although strategies such as binder engineering, pulsed electrolysis, and membrane modifications show potential, they still require further refinement to enhance stability and practical application.

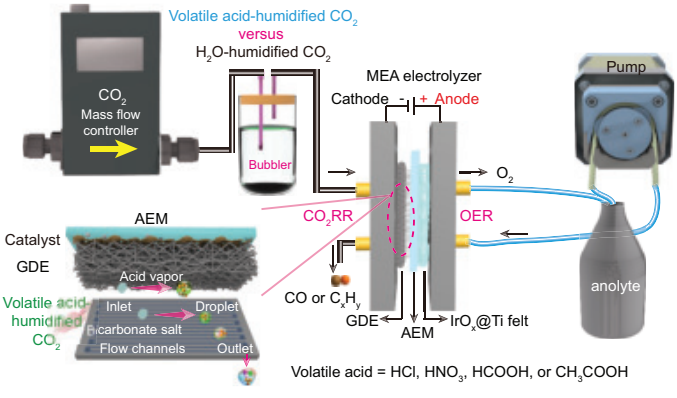
RATIONALE: To address the critical challenge of salt formation in the cathode chamber of CO_2RR MEA electrolyzers, we pursued alternatives to the traditional water humidification method, which still suffers from salt formation that compromises system performance. We hypothesized that introducing trace amounts of volatile acid vapor into the CO_2 gas stream could mitigate salt accumulation and enhance the CO_2RR stability. Unlike water, which facilitates the formation of low-solubility salts like $KHCO_3$, volatile acids are more likely to convert these salts into more soluble potassium salts. If this occurs without disrupting the CO_2RR local environment, this approach could considerably enhance operational stability and serve as a key step toward CO_2RR commercialization.

RESULTS: Our acid-humidification strategy effectively mitigated salt precipitation in CO_2RR MEA electrolyzers, substantially extending the operational stability. By humidifying CO_2 with volatile acids such as HCl, HNO_3, HCOOH, or CH_3COOH, we observed a notable reduction in salt formation compared with conventional water humidification. This salt mitigation occurred while maintaining high CO FE and stable cell voltages using silver nanoparticles (Ag-NP) electrocatalysts, with acid vapors enhancing salt solubility and preventing bicarbonate precipitation. Notably, 0.2 M HCl-humidified CO_2 achieved >1000-hour stability with a 4-cm^2 Ag/GDE at 100 mA cm^{-2}, maintaining CO FEs above 90%. By contrast, water-humidified systems failed within 200 to 500 hours. Microscopy and operando analyses revealed that water-humidified CO_2 caused rapid salt crystallization on the GDE backside, obstructing CO_2 flow. By contrast, HCl vapor dissolved precipitates and slowed crystal growth, preserving unobstructed CO_2 flow channels and extending stability. This approach is applicable beyond Ag-NPs, demonstrating compatibility with various electrocatalysts, including ZnO, Bi_2O_3, and Cu_2O, without compromising activity or stability. We then scaled up the system to a 100-cm^2 electrolyzer at 100 mA cm^{-2} using Ag-NP. In this scaled-up configuration, HCl-humidified CO_2 operated stably for >2000 hours and HNO_3-humidified CO_2 reached 4500 hours, achieving CO production rates of ~3.8 L h^{-1} with >80% CO FEs and minimal cell voltage degradation.

CONCLUSION: This study elucidates the mechanisms behind K^+ crossover through AEM, leading to the formation of (bi)carbonate salt on the GDE backside and within gas flow channels. By introducing a volatile acid-humidified CO_2, we effectively prevented salt precipitation, ensuring stable operation in 100-cm^2 MEA electrolyzers for >4500 hours at 100 mA cm^{-2} while maintaining a CO FE exceeding 80%. Our acid-humidification method outperforms conventional water-humidified systems, providing a straightforward and scalable approach to enhance CO_2RR stability while preserving reaction selectivity and cell voltage, thereby advancing the commercialization prospects of CO_2RR MEA electrolyzers. □

*Corresponding author. Email: htwang@rice.edu †These authors contributed equally to this work. Cite this article as S. Hao *et al.*, *Science* **388**, eadr3834 (2025). DOI: 10.1126/science.adr3834

Schematic of the setup for the acid-humidified CO_2 approach. Setup of CO_2RR MEA electrolyzer using acid-humidified CO_2 versus a conventional H_2O-humidified CO_2. The acid-humidified CO_2 method addresses salt formation in the cathode chamber, thereby enhancing operational stability.



DEVICE PHYSICS

Gate-driven band modulation hyperdoping for high-performance p-type 2D semiconductor transistors

Bei Zhao[1,2]†, Zucheng Zhang[1]†, Junqing Xu[3]†, Dingli Guo[2], Tiancheng Gu[2], Guiming He[2], Ping Lu[1], Kun He[1], Jia Li[1], Zhao Chen[3], Quan Ren[2], Lin Miao[2], Junpeng Lu[2,4], Zhenhua Ni[2,4], Xiangfeng Duan[5], Xidong Duan[1]*

Tailoring carrier density in atomically thin two-dimensional (2D) semiconductors is challenging because of the inherently limited physical space for incorporating charge dopants. Here, we report that interlayer charge-transfer doping in type III van der Waals heterostructures can be greatly modulated by an external gate to realize a hyperdoping effect. Systematic gated-Hall measurements revealed that the modulated carrier density is about five times that of the gate capacitive charge, achieving an ultrahigh 2D hole density of 1.49×10^{14} per square centimeter, far exceeding the maximum possible electrostatic doping limit imposed by typical dielectric breakdown. The highly efficient hole-doping enables high-performance p-type 2D transistors with an ultralow contact resistance of ~0.041 kilohm micrometers and a record-high ON-state current density of ~2.30 milliamperes per micrometer.

The atomically two-dimensional semiconductors (2DSCs) (*1*–*6*) have attracted considerable recent interest for their superior immunity to short-channel effects (*6*–*8*). Tailoring the carrier density in semiconductors is critical for controlling their fundamental electronic properties and the corresponding device performance. For example, substitutional lattice doping (e.g., by ion implantation) is commonly used for controlling the carrier density and minimizing the contact resistance in conventional semiconductors. However, with an atomically thin covalent lattice in 2DSCs, there is simply insufficient physical space to accommodate sufficient lattice dopants without seriously compromising the lattice structures and electronic properties. Additionally, the existing state-of-the-art ion implantation doping approach is generally unsuitable for 2DSCs because the high-energy ion implantation process could seriously damage the atomically thin lattices and degrade their electronic properties.

Charge-transfer doping using selected surface adsorbates (molecules or ions) has also been explored for tailoring the electronic properties of 2DSCs without altering the pristine 2D covalent lattices (*9*–*14*). However, such surface doping processes often involve aggressive chemical treatments, introducing surface impurities or Coulomb scattering centers that can degrade charge transport properties. Moreover, such surface-adsorbate doping usually shows limited stability and is often incompatible with the high-resolution lithography processes needed for fabricating ultrashort-channel devices.

The persistent challenges in tailoring carrier density in 2DSCs have fundamentally limited the capability to achieve sufficiently low contact resistance (R_C) and explore the performance limit of 2D transistors. Minimizing R_C between the metal electrodes and the semiconductor channels is a common challenge in semiconductor devices, particularly in ultrashort channel transistors, where the R_C could largely dictate the total resistance and the ultimate switching speed. Most metal-semiconductor contacts typically show a non-negligible Schottky barrier resulting from a band misalignment or universal Fermi level pinning effect (*15*, *16*). To address this challenge, the semiconductor industry typically uses impurity doping to increase carrier density in the contact region, which shrinks the Schottky barrier width to boost tunneling injection efficiency and reduce R_C. However, this approach remains challenging for the 2DSCs because of difficulty in incorporating impurity dopants in atomically thin 2D lattices. Considerable efforts have been made to reduce R_C using various strategies, including mechanical transferred pinning-free van der Waals (vdW) contacts (*17*), low-energy deposited metal contacts (*18*), semimetallic contacts (*19*, *20*), and electron-doped contact regions (*21*). Some of these strategies have proven successful in greatly reducing R_C for n-type 2DSCs (*18*–*21*); however, they are generally less effective for p-type 2DSCs because of several practical challenges, including the lack of low-melting-point metals with high work functions suitable for low-temperature deposition (*15*, *22*).

We report a gate-tunable band-modulation approach for doping 2DSCs by first forming van der Waals heterostructures (vdWHs) between monolayer SnS_2 (1L-SnS_2) and bilayer WSe_2 (2L-WSe_2) that had a type-III broken band alignment. Electron transfer from the WSe_2 valence band (VB) into the SnS_2 conduction band (CB) induced a hole-doping effect in bilayer WSe_2. An external gate bias applied through a silicon (Si) back gate with a silicon nitride (SiN_x) as the gate dielectric modulates the interlayer charge-transfer doping, enabling a hyperdoping effect that surpasses the maximum electrostatic doping limit imposed by typical dielectric breakdown. Systematic gated-Hall measurements revealed that the modulated carrier density was about five times that of the gate capacitive charge, reaching an ultrahigh 2D hole density of 1.49×10^{14} cm^{-2}. The high hole density enabled a lowest R_C of ~0.041 kΩ μm and a highest ON-state current density (I_{on}) of ~2.30 mA/μm for p-channel 2D transistors.

Modulation doping in SnS_2/WSe_2 vdWHs

The 1L-SnS_2/2L-WSe_2 vdWHs were prepared through a vdW epitaxial growth of 1L-SnS_2 on 2L-WSe_2 with a chemical vapor-deposition process (fig. S2, A and B; see the materials and methods for details). Atomic force microscopy studies revealed that the bottom layer and total thickness were ~1.26 nm (for 2L-WSe_2) and ~1.95 nm (for 1L-SnS_2/2L-WSe_2), respectively, consistent with the expected 1L-SnS_2/2L-WSe_2 vdWHs (Fig. 1A). Energy-dispersive spectroscopy elemental mapping images confirmed a uniform distribution of the W, Se, Sn, and S atoms (Fig. 1B and fig. S2C). The selected area electron diffraction pattern of the 1L-SnS_2/2L-WSe_2 vdWH region showed two sets of identically orientated diffraction spots (Fig. 1C) that matched well with the intrinsic lattice constants of WSe_2 (0.33 nm) and SnS_2 (0.37 nm). The high-angle annular dark-field scanning transmission electron microscopy image of the vdWH revealed apparent moiré structure with a superlattice period of ~3.34 nm (10×10 WSe_2 and 9×9 SnS_2) (Fig. 1D and fig. S2, D and E).

The density functional theory–GW (DFT-GW) calculations (see the materials and methods for details) (*23*) indicated that pure 2L-WSe_2 and pure 1L-SnS_2 have band gaps of 1.68 and 2.43 eV (fig. S3, A to C), respectively, which is consistent with those determined by scanning tunneling spectroscopy studies (1.68 and 2.22 eV, respectively) (fig. S3,

[1]Hunan Key Laboratory of Two-Dimensional Materials, State Key Laboratory of Chemo and Biosensing, College of Chemistry and Chemical Engineering, Hunan University, Changsha, China. [2]School of Physics and Key Laboratory of Quantum Materials and Devices of Ministry of Education, Southeast University, Nanjing, China. [3]Department of Physics, Hefei University of Technology, Hefei, Anhui, China. [4]School of Electronic Science and Engineering, Southeast University, Nanjing, China [5]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA, USA. ***Corresponding author. Email: xidongduan@hnu.edu.cn**
†These authors contributed equally to this work.

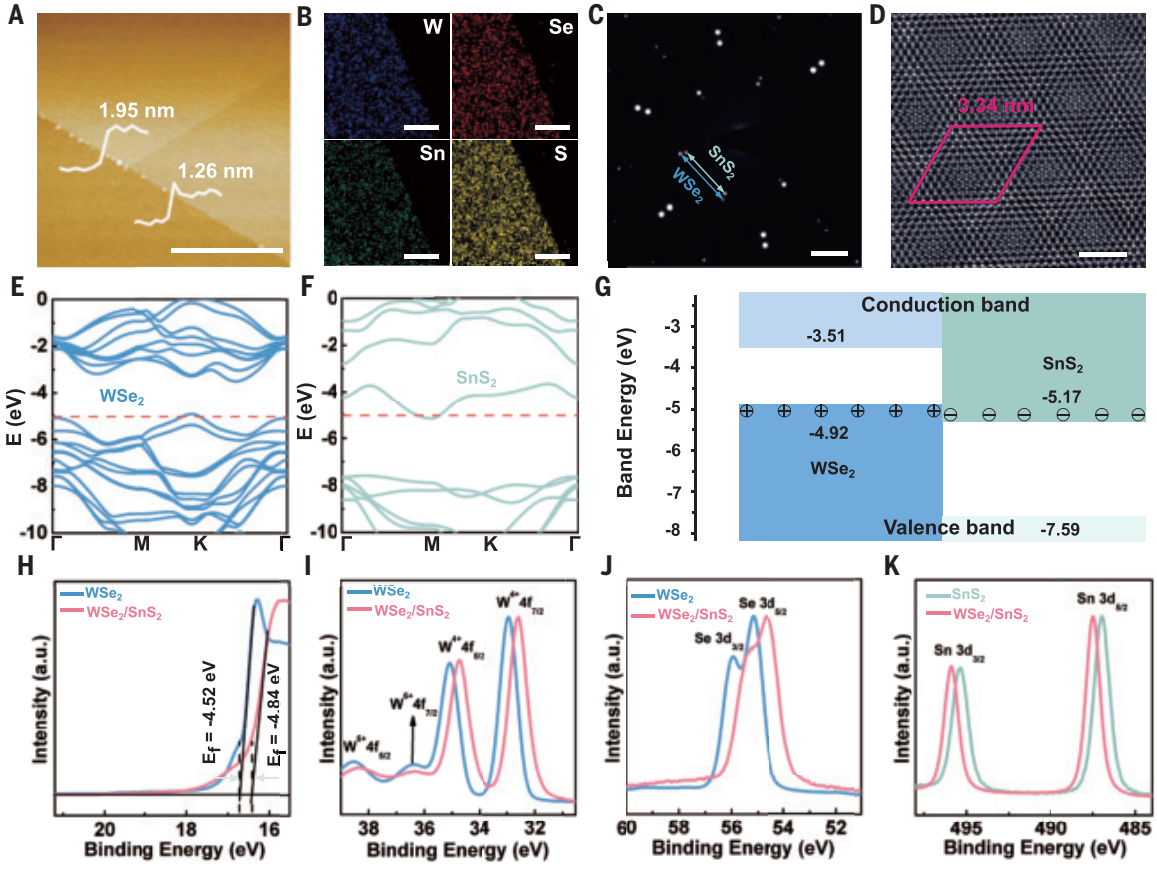


Fig. 1. Modulation doping in 1L-SnS_2/2L-WSe_2 vdWHs. (**A**) Atomic force microscopy image and corresponding height profiles of SnS_2/WSe_2 vdWH. Scale bar, 5 μm. (**B**) Energy-dispersive spectroscopy elemental mapping images of W, Se, Sn, and S for a SnS_2/WSe_2 vdWH. Scale bars, 500 nm. (**C**) Selected area electron diffraction patterns of SnS_2/WSe_2 vdWH. Scale bar, 2 nm^{-1}. (**D**) High-angle annular dark-field scanning transmission electron microscopy image taken on the SnS_2/WSe_2 vdWH region showing the period of the moiré pattern $L = 3.34$ nm. Scale bar, 2 nm. (**E** and **F**) Calculated band structures of 2L-WSe_2 (E) and 1L-SnS_2 (F) in SnS_2/WSe_2 vdWH. (**G**) Relative band alignment of SnS_2/WSe_2 vdWH. (**H**) UPS spectra of WSe_2 (blue line) and SnS_2/WSe_2 vdWH (red line). (**I** and **J**) XPS spectra of W (I) and Se (J) in WSe_2 (blue line) and SnS_2/WSe_2 vdWH (red line). (**K**) XPS spectra of Sn in SnS_2 (green line) and SnS_2/WSe_2 vdWH (red line).

D to I). Moreover, DFT-GW calculations further indicated that the formation of a 1L-SnS_2/2L-WSe_2 vdWH would lead to a slight modification of the band structures of 2L-WSe_2 and 1L-SnS_2 (Fig. 1, E and F) through a dielectric screening effect (see the materials and methods for details) and result in a type-III broken band alignment (2L-WSe_2 heterostructure valance band energy $E^{H}_{VB} = -4.92$ eV and heterostructure conduction band energy 1L-SnS_2 $E^{H}_{CB} = -5.17$ eV) (Fig. 1G). With this broken band alignment, the electron may transfer from the WSe_2 VB to the SnS_2 CB, resulting in p-type doping of WSe_2 and n-type doping of SnS_2. Such an interfacial charge-transfer doping process resembles the modulation doping in conventional semiconductor heterostructures and superlattices and could greatly increase the hole carrier density in WSe_2 without introducing atomic defects in 2D WSe_2 lattices. The band edge positions and the exact band offsets may vary slightly depending on the band structure calculation methods used. Nonetheless, a similar charge-transfer doping effect would persist for a heterostructure with a type III broken band alignment or a type II staggered band alignment with a relatively small interlayer bandgap.

To confirm the hole-doping effect in WSe_2, we conducted ultraviolet and x-ray photoemission spectroscopy (UPS and XPS, respectively) analyses to determine the Fermi level (E_F) in the respective 2DSCs before and after forming SnS_2/WSe_2 vdWHs. The UPS studies revealed a red-shift of 0.32 eV upon forming SnS_2/WSe_2 vdWHs, corresponding to the decrease of E_F from −4.52 eV in WSe_2 to −4.84 eV in SnS_2/WSe_2 (Fig. 1H), consistent with a hole-doping effect (*24*). Furthermore, XPS studies of the core-level binding energies of WSe_2 showed red-shifts from 32.94 to 32.58 eV for W $4f_{7/2}$, from 35.09 to 34.73 eV for W $4f_{5/2}$, and from 55.15 to 54.64 eV for Se $3d_{5/2}$ upon forming SnS_2/WSe_2 vdWH (Fig. 1, I and J) (*9*, *25*), consistent with a downward E_F shift and hole doping of WSe_2. Notably, the W^{6+} peaks (Fig. 1I) were mainly caused by the adsorbed WO_3 nanoparticles at the edges of WSe_2 and SnS_2/WSe_2 vdWH (fig. S4), which is not expected to have a significant effect on the charge doping in the WSe_2 device far away from the edge. Conversely, the core-level binding energies of SnS_2 were blue-shifted in SnS_2/WSe_2 vdWHs (Fig. 1K and fig. S5) compared with those in pristine SnS_2 (from 486.99 to 487.49 eV for Sn $3d_{5/2}$, from 495.34 to 495.89 eV for Sn $3d_{3/2}$, from 226.24 to 226.89 eV for S 2s, from 162.04 to 162.54 eV for S $2p_{3/2}$ and from 163.24 to 163.74 eV for S $2p_{1/2}$) (*9*, *26*), consistent with an upward E_F shift and electron doping in SnS_2.

Gate-tunable ultrahigh 2D carrier density

We next compared the electrical transport behavior of 2L-WSe_2 and the 1L-SnS_2/2L-WSe_2 vdWH Hall bar devices on SiN_x/Si substrate, with Si substrate and 70-nm-thick silicon nitride as the global back-gate and gate dielectrics, respectively, to form field-effect transistors (FETs). The vdWH devices were evaluated in two different geometries (Fig. 2A): The as-grown geometry had a gate stack sequence of SnS_2/WSe_2/SiN_x/Si, and the flipped geometry had a gate stack sequence of WSe_2/SnS_2/SiN_x/Si. Two-terminal current-voltage output characteristics (I_{ds}–V_{ds}) showed that the current of as-grown 1L-SnS_2/2L-WSe_2

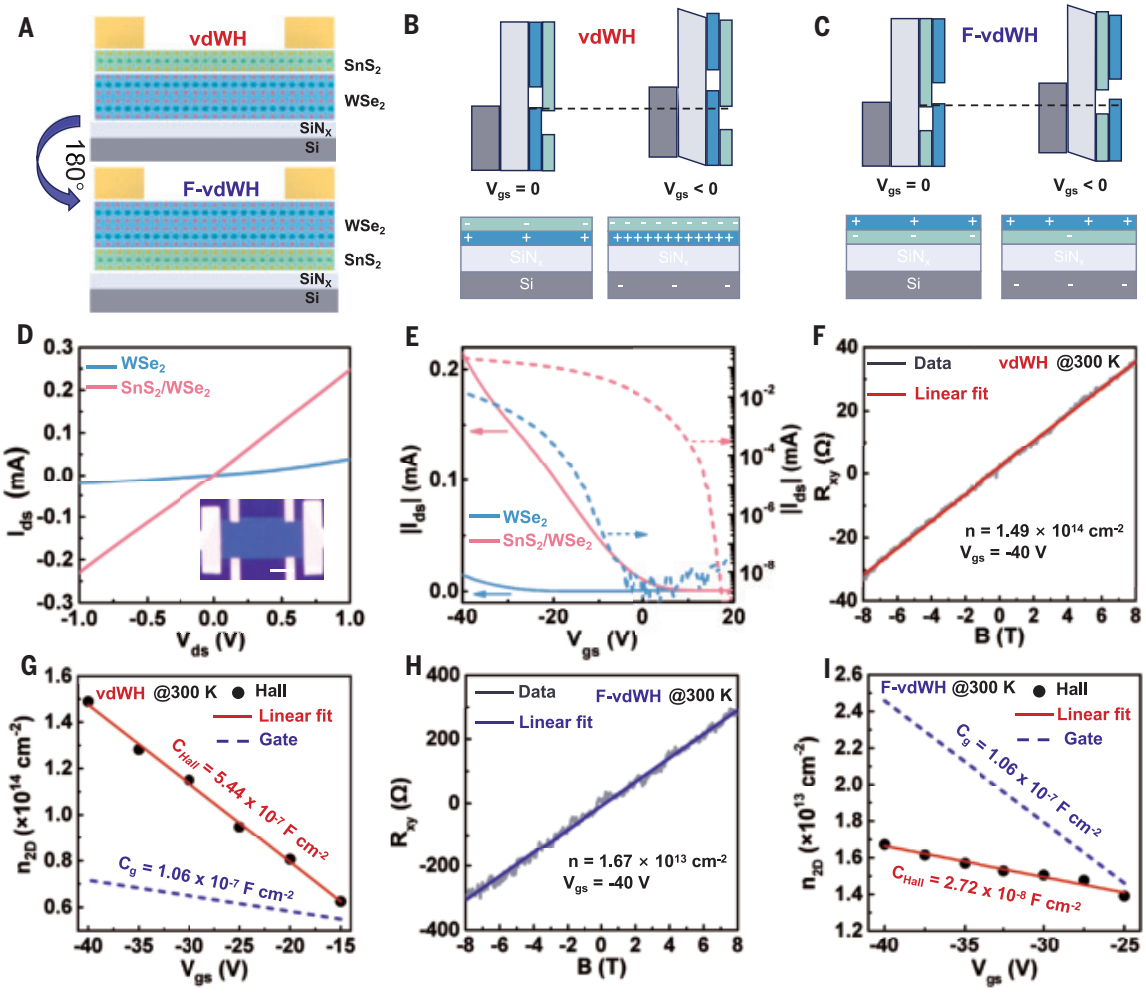


Fig. 2. Gate-tunable band modulation enables ultrahigh 2D carrier density in 1L-SnS_2/2L-WSe_2 vdWHs. (**A**) Schematic of the 1L-SnS_2/2L-WSe_2 vdWH structure and 180° flipped structure. (**B**) Band diagram of the 1L-SnS_2/2L-WSe_2 Hall device showing increased band offset and further enhanced charge-transfer doping at $V_{gs} < 0$. (**C**) Band diagram of the flipped 2L-WSe_2/1L-SnS_2 Hall device showing reduced band offset and suppressed charge-transfer doping at $V_{gs} < 0$. (**D**) Output characteristics of the 2L-WSe_2 (blue) and 1L-SnS_2/2L-WSe_2 vdWH (red) devices at $V_{gs} = -40$ V. Inset is an optical microscopy image of a typical Hall bar device used for carrier density analysis. Scale bar, 5 μm. (**E**) Linear (left) and logarithmic (right) plots of transfer characteristics of the 2L-WSe_2 (blue) and 1L-SnS_2/2L-WSe_2 vdWH (red) devices at $V_{ds} = -1$ V. (**F**) Hall resistance (R_{xy}) of 1L-SnS_2/2L-WSe_2 vdWH as a function of magnetic field (B) at room temperature ($V_{gs} = -40$ V). (**G**) Hole densities of 1L-SnS_2/2L-WSe_2 vdWH determined from Hall measurements as a function of gate voltage. (**H**) Hall resistance (R_{xy}) of the flipped 2L-WSe_2/1L-SnS_2 vdWH as a function of magnetic field (B) at room temperature ($V_{gs} = -40$ V). (**I**) Hole densities of the flipped 2L-WSe_2/1L-SnS_2 vdWH as a function of gate voltage.

vdWH FET (red line) was notably higher than that of 2L-WSe_2 (Fig. 2D).

The transfer characteristics (I_{ds}–V_{gs}) showed that both 2L-WSe_2 and 1L-SnS_2/2L-WSe_2 vdWH FETs exhibited higher conductance toward a more negative gate voltage (Fig. 2E), indicating a p-type transport behavior. Compared with the 2L-WSe_2 FET, the 1L-SnS_2/2L-WSe_2 FET showed a positive shift of the threshold voltage (V_{th}), consistent with the expected additional hole-doping upon forming 1L-SnS_2/2L-WSe_2 vdWHs. Beyond the threshold shift, the transfer curve also showed a much more rapid increase of conductance with increasing gate voltage (i.e., a steeper slope) (Fig. 2E). The steeper slope observed in 1L-SnS_2/2L-WSe_2 FET than that in 2L-WSe_2 FET suggested that an additional charge doping mechanism operated beyond electrostatic doping through the gate capacitor.

To further evaluate carrier density (n_{2D}) in 1L-SnS_2/2L-WSe_2 vdWH, we conducted gated-Hall measurements under different applied gate voltages. The carrier density could be calculated with the equation $n_{2D} = -1/eR_H$, where R_H is the Hall coefficient, defined as the slope of Hall resistance (R_{xy}) versus magnetic field (B). The R_{xy}–B plots for the 1L-SnS_2/2L-WSe_2 vdWH and 2L-WSe_2 revealed a linear relationship with a positive slope (Fig. 2F and fig. S6), indicating hole-dominated transport. The Hall measurements revealed that the hole density in 1L-SnS_2/2L-WSe_2 vdWHs could reach up to 1.49×10^{14} cm^{-2} at $V_{gs} = -40$ V, which is more than two orders magnitude higher than that of 2L-WSe_2 under the same gate voltage (fig. S6K). No apparent signature of electron transport was observed in the 1L-SnS_2/2L-WSe_2 vdWHs, which could be attributed to much lower electron mobility in 1L-SnS_2 (*27*) (fig. S7).

The hole density in 1L-SnS_2/2L-WSe_2 vdWH was readily modulated with the external potential applied to the Si back-gate. As V_{gs} increase negatively from –15 V to –40 V, the hole densities (n_{2D}) determined from the Hall measurements increased from 6.25×10^{13} to 1.49×10^{14} cm^{-2}, which corresponded to an apparent capacitance $C_{Hall} = 5.44 \times 10^{-7}$ F cm^{-2} (Fig. 2G). This value is about five times larger than the gate dielectric capacitance expected for the 70-nm SiN_x gate dielectrics ($C_g = 1.06 \times 10^{-7}$ F cm^{-2}), suggesting the existence of an additional charge-doping mechanism beyond electrostatic capacitive gating. This difference can be understood by considering the device geometry and the gate-modulated band offset within the 1L-SnS_2/2L-WSe_2 type-III heterostructure (Fig. 2B). Because 2L-WSe_2 was closer to the gate and 1L-SnS_2 was

farther away from the gate, the gate-exerted electrostatic field on the 1L-SnS_2 was partly screened by the charge in 2L-WSe_2. Such a screening effect increased with the increasing hole density in 2L-WSe_2 at a more negative V_{gs}. This enhanced screening effect resulted in a smaller upward shift of the 1L-SnS_2 CB than that of 2L-WSe_2 VB and a larger broken band offset, which led to greater electron transfer from 2L-WSe_2 VB to 1L-SnS_2 CB, which manifested as a hyperdoping effect beyond the gate capacitance (*28*, *29*). Conversely, a doping-induced band offset reduction effect (*30*) could partly suppress the gate-induced band modulation due to increasing interfacial electric field at high carrier density, together resulting in a nearly linear doping behavior (Fig. 2G).

For the flipped device geometry with the 1L-SnS_2 closer to the gate (Fig. 2A), the screening effect by the 1L-SnS_2 would produce a smaller upward shift of the 2L-WSe_2 VB than that of the 1L-SnS_2 CB under a negative gate potential (Fig. 2C), leading to a smaller broken band offset and thus less of a charge-transfer doping effect. Indeed, our experimental studies on a flipped 2L-WSe_2/1L-SnS_2 vdWH device on the same SiN_x/Si substrate revealed that the 2L-WSe_2 layer exhibited a much smaller carrier modulation rate and a lower hole density than what would be expected from the gate capacitance (Fig. 2, H and I). These results confirmed that gate-tunable band modulation led to the unusual hyperdoping effect in the 1L-SnS_2/2L-WSe_2 vdWH.

The hole density (1.49×10^{14} cm^{-2}) achieved is much higher than the highest 2D carrier density reported previously in electrostatically gated 2DSCs ($<4.6 \times 10^{13}$ cm^{-2}; see table S2) (*18*–*22*, *31*–*51*) and actually exceeds the breakdown limit of typical SiN_x dielectrics ($<3 \times 10^{13}$ cm^{-2}) (*52*, *53*). This high carrier density beyond dielectric limits further confirmed that the hyperdoping effect in 1L-SnS_2/2L-WSe_2 vdWH originated from interfacial charge transfer that was beyond simple electrostatic gating. With the interfacial charge-transfer doping model, the hole doping in 2L-WSe_2 should be accompanied by a similar electron doping in 1L-SnS_2. However, our studies showed that the transport in the vdWHs was largely dictated by the hole transport in 2L-WSe_2, which may be attributed to much lower electron mobilities in 1L-SnS_2, lower electron injection efficiency through the contacts, or a combination of both effects (*27*).

Ultralow R_C enabled by high carrier density

High carrier density in semiconductors is essential for realizing a low R_C in semiconductor devices. Therefore, we further investigated the R_C of the 1L-SnS_2/2L-WSe_2 vdWH and the bare 2L-WSe_2 FETs with high work function Au ($\Phi_m = 5.3$ eV) thin films as the source and drain electrodes in a back-gate geometry with 70-nm-thick SiN_x gate dielectrics. We first used a transfer-length method (TLM) to determine the R_C in a multiprobe device with varying channel lengths (Fig. 3A) (*17*). This method is more reliable for determining R_C when the current-voltage relationship is linear and the R_C is consistent at each contact, so we focused most of our analyses at more negative gate potential when the current-voltage relationship is linear. Our analyses revealed that the R_C of the 1L-SnS_2/2L-WSe_2 vdWH devices could be modulated from 0.063 to 0.045 kΩ μm as the negative gate voltage increased from −30 to −40 V. The decreasing R_C at more negative gate potential was attributed to the increasing carrier density, the shrinking depletion width, and thus the increasing tunneling injection efficiency. The lowest R_C of 0.045 kΩ μm achieved at $V_{gs} = -40$ V was more than two orders of magnitude lower than that of 2L-WSe_2 (10.63 kΩ μm) under the same gate voltage (fig. S8).

The R_C values extracted from the TLM could be susceptible to considerable uncertainties and thus require alternative method for corroboration (*54*). We have thus further conducted four-probe measurements using Hall bar devices to determine the specific contact resistance of 12 1L-SnS_2/2L-WSe_2 vdWH devices (Fig. 3, B and C). The gate-dependent R_C determined from four-probe measurements (from 0.041 to 0.049 kΩ μm at $V_{gs} = -40$ V) matched well with that obtained from the TLM characterizations, confirming the robustness of our measurements. A comparison with the lowest R_C achieved in 2DSCs reported previously (*18*–*22*, *31*–*51*) revealed that the R_C of ~0.041 kΩ μm achieved in the 1L-SnS_2/2L-WSe_2 vdWH represents the smallest room-temperature R_C reported in 2DSCs, particularly p-channel 2DSCs, to date (Fig. 3D and table S3). These values also compared well with those commonly reported in bulk semiconductors [R_C ~0.1 kΩ μm, defined by the International Roadmap for Devices and Systems (IRDS) (*55*)]. The low R_C achieved in our studies was further supported by the lowest total resistance achieved in the sub–100-nm channel length 1L-SnS_2/2L-WSe_2 vdWH FETs among all reported 2D FETs at the corresponding channel lengths (see discussions below and Fig. 4F).

The R_C in a typical semiconductor device at the low carrier density regime is dominated by thermionic emission and usually shows a notable dependence on the metal work function, whereas that at the high carrier density regime is dominated by tunneling injection and is much less insensitive to the work function of metal. Indeed, the output characteristics of the 1L-SnS_2/2L-WSe_2 vdWH FETs showed a highly comparable current level regardless of the contact metal work functions (5.3 eV for Au, 4.6 eV for Cr, and 4.2 eV for Ti). By contrast, the bare 2L-WSe_2 FETs showed clearly higher hole injection efficiency when higher work function metal was used as the contact (fig. S9).

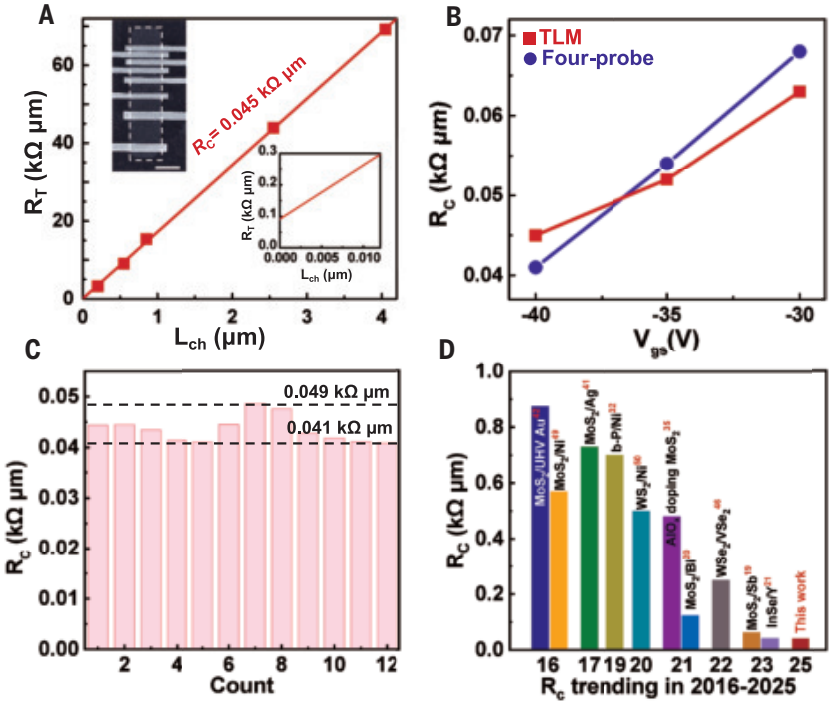


Fig. 3. Ultralow R_C in 1L-SnS_2/2L-WSe_2 vdWH FETs. (**A**) TLM determination of R_C in 1L-SnS_2/2L-WSe_2 vdWH FET. Inset is an SEM image of the multiprobe TLM structure, and the enlarged plot highlights the ultralow R_C (0.045 kΩ μm) achieved in 1L-SnS_2/2L-WSe_2 vdWH device at V_{gs}= −40 V. Scale bar, 5 μm. (**B**) Comparison of the R_C of 1L-SnS_2/2L-WSe_2 vdWH devices determined from TLM and four-probe measurements. (**C**) Statistical bar graph of contact resistance from 12 1L-SnS_2/2L-WSe_2 vdWH devices determined from four-probe measurements. (**D**) Lowest R_C achieved in modulation-doped 1L-SnS_2/2L-WSe_2 vdWH device compared with those of other contacting strategies with $R_C < 1$ kΩ μm.

Ultrahigh current density in p-channel 2D transistors

Ultralow R_C materials could greatly boost transistor performance, particularly in the short-channel devices

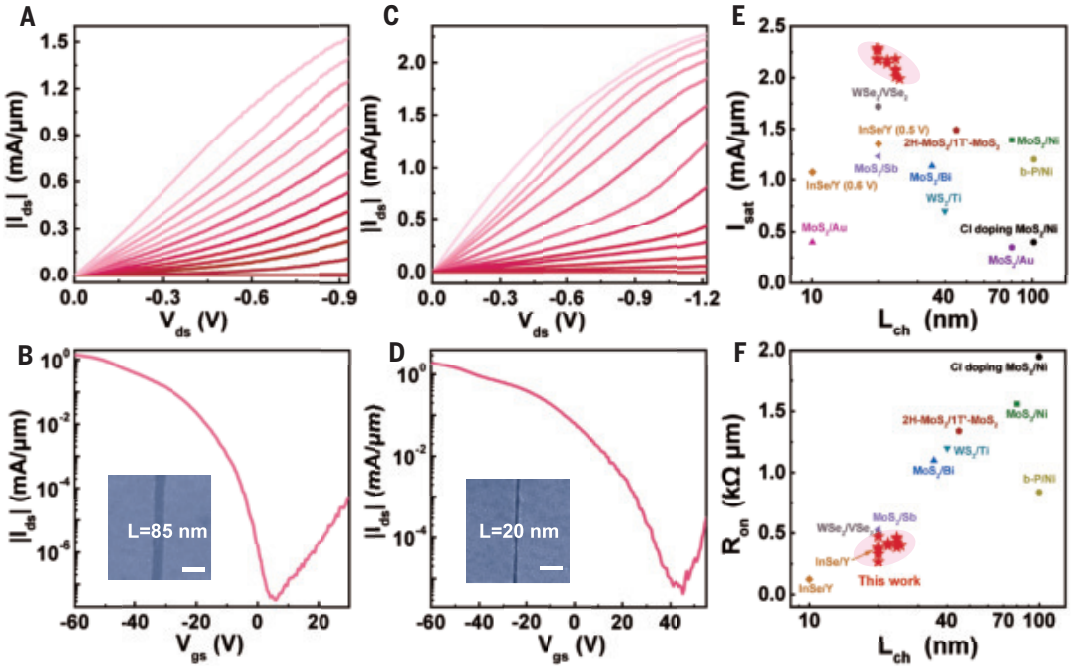


Fig. 4. Ultrahigh I_{sat} and ultralow R_{on} in 1L-SnS_2/2L-WSe_2 vdWH FETs. (**A** and **B**) Output (at V_{gs} from 60 V in −2.5 V steps) and transfer (V_{ds} = −1 V) curves of an 85-nm channel length 1L-SnS_2/2L-WSe_2 vdWH FET. (**C** and **D**) Output (at V_{gs} from 60 V in −2.5 V steps) and transfer (V_{ds} = −1 V) curves of a 20-nm channel length 1L-SnS_2/2L-WSe_2 vdWH FET. Scale bars in the inset SEM images in (B) and (D) correspond to 200 nm. (**E** and **F**) Comparison of I_{sat} and R_{on} of 1L-SnS_2/2L-WSe_2 FETs with other 2D transistors as a function of channel length.

where the R_C dominates the total ON-state resistance (R_{on}). We investigated the output current from the sub–100-nm channel devices made with 1L-SnS_2/2L-WSe_2 vdWHs (Fig. 4, A and C, and figs. S10 and S11). Under a back gate geometry (with 70-nm SiN_x as the gate dielectric), the 85-nm channel length device delivered an I_{on} of 1.25 mA/μm at V_{ds} = 0.7 V (the supply voltage in commercial silicon 10-nm node) and 1.53 mA/μm at V_{ds} = 0.94 V and achieved a current ON/OFF ratio of up to >10^8 (Fig. 4, A and B); the 30-nm channel length device delivered an I_{on} of 1.67 mA/mm at V_{ds} = 0.7 V and 2.10 mA/μm at V_{ds} = 1.0 V (fig. S10C); and the ~20-nm channel length device delivered an I_{on} of 2.00 mA/μm at V_{ds} = 0.7 V and 2.30 mA/μm at V_{ds} = 0.88 V (fig. S11K).

We have further evaluated 12 short-channel-length (~20-nm) devices and found that all of them consistently delivered a high I_{on} of ~2.00 mA/μm or above (fig. S11). A relatively high gate voltage was needed for our back-gated devices because of the thick gate dielectrics used (70-nm SiN_x). By using thinner top-gate dielectrics, the required gate voltage could be greatly reduced while retaining an outstanding performance (*21*, *56*, *57*) of I_{on} ~1.73 mA/μm and a high ON/OFF ratio of >10^7 (fig. S12). In this case, a constant back gate voltage of −45 V was applied to maintain ultralow contact resistance and to ensure superior electrical performance (ON-current). A top gate with ultrathin HfO_2 was used to switch the transistor with a much smaller gate swing.

The maximum I_{on} (saturation current: I_{sat}) achieved in the sub–100-nm 1L-SnS_2/2L-WSe_2 vdWH FETs represent the highest current density ever achieved in 2D transistors at the corresponding channel lengths (Fig. 4E and table S4). In particular, the maximum I_{on} of ~2.30 mA/μm achieved in 20-nm 1L-SnS_2/2L-WSe_2 FET represented the highest value reported among all 2D transistors to date (*3*, *9*, *19*, *20*, *32*, *34-36*, *44-46*, *49*, *50*, *58-83*) and is comparable to the 2024 high-performance current density target set in the 2016 IRDS report (*55*). Additionally, the I_{on} achieved at the standard supply V_{ds} = 0.7 V could reach up to 2.00 mA/μm, which exceeded those of the other 2DSCs reported so far (fig. S13A) (*3*, *9*, *18*, *20*, *21*, *32*, *34–36*, *44–46*, *49*, *50*, *58–63*, *66–69*, *72–75*, *78–80*) and compared well with that of Si FinFETs at the corresponding channel length. With the ultralow R_{on} achieved in our 1L-SnS_2/2L-WSe_2 vdWH FETs, a bias voltage of only 0.27 V was required to achieve the current density of 1.0 mA/μm, which is lower than that of the 2D ballistic transistor (0.40 V) reported recently (*21*).

Additionally, the lowest ON-state resistance (R_{on} derived from the linear regime) achieved in 85-, 30-, and 20-nm 1L-SnS_2/2L-WSe_2 vdWH devices were 0.53, 0.40, and 0.27 kΩ μm, respectively, representing the lowest R_{on} ever achieved for p-channel 2D transistors and all 2D transistors at the corresponding channel lengths (Fig. 4F and fig. S13, B to D). For example, for the ultrashort 20-nm channel 2D transistors reported to date, the R_{on} of our device was only one-third of the previous record for n-type MoS_2 (*19*), one-half of the previous record for p-type WSe_2 (*46*), and three-quarters of the previous record for n-type InSe (*21*) (Fig. 4F and fig. S13B). Similarly, the R_{on} of our devices at other channel lengths was also notably lower than the state-of-art 2D transistors reported to date (Fig. 4F and fig. S13, C and D).

To further evaluate the scalability of our approach, we have prepared continuous films (1.5 × 1.5 cm^2) of the 1L-SnS_2/ 2L-WSe_2 vdWH and fabricated large-area device arrays (fig. S14, A and B). Our studies of 130 FETs showed that all of the devices exhibited excellent ohmic behavior, with an average I_{on} > 1.0 mA/μm at L_{ch} = 0.8 μm and an ON/OFF ratio of 10^{10} (fig. S14, D to F), demonstrating highly reliable contacts achieved in the large-area 1L-SnS_2/2L-WSe_2 vdWH thin films. Moreover, a lightly doped channel with heavily doped contacts is desirable for threshold V_{th} control. Therefore, we used a fast laser carving technique to selectively remove the 1L-SnS_2 and eliminate the charge-transfer doping in the channel region, which led to a negative shift of V_{th} and an OFF state near zero gate without sacrificing the device performance (I_{on}) (fig. S15).

Discussion

We report a gate-tunable band modulation hyperdoping strategy to achieve ultrahigh 2D carrier density in 2DSCs. We showed that the formation of type-III 1L-SnS_2/2L-WSe_2 vdWHs with a broken band alignment allows the electron transfer from the 2L-WSe_2 VB to the 1L-SnS_2 CB, resulting in a hole-doping effect in 2L-WSe_2. By leveraging the strong gate-tunable band modulation in atomically thin 2D vdWHs and efficient

charge transfer across the nearly perfect interface of synthetic SnS_2/WSe_2 vdWHs, we achieved a hyperdoping effect in 2DSCs beyond the maximum electrostatic doping limit imposed by typical dielectric breakdown. This enabled an ultrahigh 2D carrier density (1.49×10^{14} cm^{-2}), allowing us to realize a lowest R_C of ~0.041 kΩ μm, a record-high ON-state current density of ~2.30 mA/μm, and a lowest total resistance of 0.27 kΩ μm among p-channel 2D transistors reported to date. The strategy was general and could be extended to other vdWHs such as 1L-In_2Se_3/2L-WSe_2 and 1L-$SnSe_2$/2L-WSe_2 vdWHs with a partially overlapping type-III band alignment or a slightly staggered type-II band alignment (fig. S16). Regardless of precise exact alignment, an applied negative gate would reduce interlayer bandgap and produce a hypedoping effect. Our study thus defines a versatile band modulation doping approach to tune the carrier density in 2DSCs beyond dielectric limits.

REFERENCES AND NOTES

1. B. Radisavljevic, A. Radenovic, J. Brivio, V. Giacometti, A. Kis, *Nat. Nanotechnol.* **6**, 147–150 (2011).
2. X. D. Yan *et al.*, *Nat. Electron.* **6**, 862–869 (2023).
3. T. Li *et al.*, *Nat. Nanotechnol.* **16**, 1201–1207 (2021).
4. B. Zhao *et al.*, *Nature* **591**, 385–390 (2021).
5. D. Sarkar *et al.*, *Nature* **526**, 91–95 (2015).
6. C. Liu *et al.*, *Nat. Nanotechnol.* **15**, 545–557 (2020).
7. Y. Liu, X. Duan, Y. Huang, X. Duan, *Chem. Soc. Rev.* **47**, 6388–6409 (2018).
8. Y. Liu *et al.*, *Nature* **591**, 43–53 (2021).
9. L. Yang *et al.*, *Nano Lett.* **14**, 6275–6280 (2014).
10. D. Kiriya, M. Tosun, P. Zhao, J. S. Kang, A. Javey, *J. Am. Chem. Soc.* **136**, 7853–7856 (2014).
11. X. Li *et al.*, *Adv. Mater.* **27**, 1547–1552 (2015).
12. S. Zhang *et al.*, *Adv. Mater.* **30**, e1806345 (2018).
13. J. Gao *et al.*, *Adv. Mater.* **28**, 9735–9743 (2016).
14. Y. Zheng, J. Gao, C. Han, W. Chen, *Cell Rep. Phys. Sci.* **2**, 100298 (2021).
15. Y. Wang, M. Chhowalla, *Nat. Rev. Phys.* **4**, 101–112 (2022).
16. D. S. Schulman, A. J. Arnold, S. Das, *Chem. Soc. Rev.* **47**, 3037–3058 (2018).
17. Y. Liu *et al.*, *Nature* **557**, 696–700 (2018).
18. Y. Wang *et al.*, *Nature* **568**, 70–74 (2019).
19. W. Li *et al.*, *Nature* **613**, 274–279 (2023).
20. P. C. Shen, *et al.*, *Nature* **593**, 211–217 (2021).
21. J. Jiang, L. Xu, C. Qiu, L. M. Peng, *Nature* **616**, 470–475 (2023).
22. Y. Wang *et al.*, *Nature* **610**, 61–66 (2022).
23. C. H. Guo, J. Q. Xu, D. Rocca, Y. Ping, *Phys. Rev. B* **102**, 205113 (2020).
24. H. Cho *et al.*, *Nano Lett.* **21**, 3503–3510 (2021).
25. W. T. Kang *et al.*, *Nanoscale* **10**, 11397–11402 (2018).
26. Z. Khan *et al.*, *Electrochim. Acta* **257**, 328–334 (2017).
27. Y. Huang *et al.*, *Appl. Phys. Lett.* **108**, 123502 (2016).
28. S. Maikap *et al.*, *Solid State Electron.* **44**, 1029–1034 (2000).
29. Y. Liang, L. Yang, *Phys. Rev. Lett.* **114**, 063001 (2015).
30. L. Huang *et al.*, *J. Phys. Chem. Lett.* **6**, 2483–2488 (2015).
31. Y. Zhu *et al.*, *Nano Lett.* **18**, 3807–3813 (2018).
32. X. Li *et al.*, *Sci. Adv.* **5**, eaau3194 (2019).
33. J. Wu *et al.*, *Nat. Nanotechnol.* **12**, 530–534 (2017).
34. Y. T. Huang *et al.*, *ACS Appl. Mater. Interfaces* **10**, 33450–33456 (2018).
35. C. J. McClellan, E. Yalon, K. K. H. Smithe, S. V. Suryavanshi, E. Pop, *ACS Nano* **15**, 1587–1596 (2021).
36. A.-S. Chou *et al.*, "High on-current 2D nFET of 390 μA/μm at VDS = 1V using monolayer CVD MoS_2 without intentional doping" in *2020 IEEE Symposium on VLSI Technology, Honolulu, HI, USA, 16–19 June 2020* (IEEE, 2020); https://doi.org/10.1109/VLSITechnology18217.2020.9265040.
37. M. H. D. Guimarães *et al.*, *ACS Nano* **10**, 6392–6399 (2016).
38. S. S. Chee *et al.*, *Adv. Mater.* **31**, e1804422 (2019).
39. X. Cui *et al.*, *Nano Lett.* **17**, 4781–4786 (2017).
40. H. Liu *et al.*, *Nano Lett.* **13**, 2640–2646 (2013).
41. K. K. H. Smithe, S. V. Suryavanshi, M. Muñoz Rojo, A. D. Tedjarati, E. Pop, *ACS Nano* **11**, 8456–8463 (2017).
42. C. D. English, G. Shine, V. E. Dorgan, K. C. Saraswat, E. Pop, *Nano Lett.* **16**, 3824–3830 (2016).
43. S. Das, H. Y. Chen, A. V. Penumatcha, J. Appenzeller, *Nano Lett.* **13**, 100–105 (2013).
44. Y. Jung *et al.*, *Nat. Electron.* **2**, 187–194 (2019).
45. H. C. P. Movva *et al.*, *ACS Nano* **9**, 10402–10410 (2015).
46. R. X. Wu *et al.*, *Nat. Electron.* **5**, 497–504 (2022).
47. J. Wang *et al.*, *Adv. Mater.* **28**, 8302–8308 (2016).
48. H. K. Ng *et al.*, *Nat. Electron.* **5**, 489–496 (2022).
49. Y. Liu *et al.*, *Nano Lett.* **16**, 6337–6342 (2016).
50. C. S. Pang, P. Wu, J. Appenzeller, Z. H. Chen, "Sub-1nm EOT WS2-FET with IDS > 600 μA/μm at VDS=1 V and SS < 70mV/dec at LG=40nm" in *2020 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 12–18 December 2020* (IEEE, 2020); https://doi.org/10.1109/IEDM13553.2020.9372049.
51. G. Kwon *et al.*, *Nat. Electron.* **5**, 241–247 (2022).
52. B. I. Weintrub *et al.*, *Nat. Commun.* **13**, 6601 (2022).
53. Z. Y. Liu *et al.*, *J. Vac. Sci. Technol. B* **34**, 041202 (2016).
54. Z. H. Cheng *et al.*, *Nat. Electron.* **5**, 416–423 (2022).
55. Institute of Electrical and Electronics Engineers, "International Roadmap for Devices and Systems" (IEEE, 2016); https://irds.ieee.org/editions.
56. L. Wang *et al.*, *Nat. Nanotechnol.* **17**, 1206–1213 (2022).
57. J. K. Huang *et al.*, *Nature* **605**, 262–267 (2022).
58. L. M. Yang *et al.*, "High-performance MoS_2 field-effect transistors enabled by chloride doping: Record low contact resistance (0.5 kΩ·μm) and record high drain current (460 μA/μm)" in *2014 Symposium on VLSI Technology (VLSI-Technology: Digest of Technical Papers, Honolulu, HI, USA, 9–12 June 2014*" (IEEE, 2014); https://doi.org/10.1109/VLSIT.2014.6894432.
59. K. K. H. Smithe, C. D. English, S. V. Suryavanshi, E. Pop, *2D Mater.* **4**, 011009 (2017).
60. H. Wang *et al.*, "Large-scale 2D electronics based on single-layer MoS_2 grown by chemical vapor deposition" in *2012 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 10–13 December 2012* (IEEE, 2012); https://doi.org/10.1109/IEDM.2012.6478980.
61. C. D. English, K. K. H. Smithe, R. J. Xu, E. Pop, "Approaching ballistic transport in monolayer MoS_2 transistors with self-aligned 10 nm top gates" in *2016 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 3–7 December 2016* (IEEE, 2016); https://doi.org/10.1109/IEDM.2016.7838355.
62. A. Nourbakhsh *et al.*, *Nano Lett.* **16**, 7798–7806 (2016).
63. S. B. Desai *et al.*, *Science* **354**, 99–102 (2016).
64. Y. Y. Illarionov *et al.*, *Nat. Electron.* **2**, 230–235 (2019).
65. W. S. Li *et al.*, *Nat. Electron.* **2**, 563–571 (2019).
66. J. Li *et al.*, *Nature* **579**, 368–374 (2020).
67. W. Liu *et al.*, *Nano Lett.* **13**, 1983–1990 (2013).
68. H. J. Chuang *et al.*, *Nano Lett.* **16**, 1896–1902 (2016).
69. H. Fang *et al.*, *Nano Lett.* **12**, 3788–3792 (2012).
70. M. W. Iqbal *et al.*, *Sci. Rep.* **5**, 10699 (2015).
71. Y. Cui *et al.*, *Adv. Mater.* **27**, 5230–5234 (2015).
72. C. H. Yeh, W. Cao, A. Pal, K. Parto, K. Banerjee, "Area-selective-CVD technology enabled top-gated and scalable 2D-heterojunction transistors with dynamically tunable Schottky barrier" in *2019 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 7–11 December 2019* (IEEE, 2019); https://doi.org/10.1109/IEDM.2016.7838355.
73. Q. He *et al.*, *Adv. Mater.* **31**, e1901578 (2019).
74. S. X. Yang *et al.*, *Nano Res.* **11**, 554–564 (2018).
75. S. Yang *et al.*, *Nanoscale* **6**, 7226–7231 (2014).
76. M. J. Mleczko *et al.*, *Sci. Adv.* **3**, e1700481 (2017).
77. Y. Zhao *et al.*, *Adv. Mater.* **29**, 1604230 (2017).
78. Y. X. Wang *et al.*, *Nat. Electron.* **1**, 228–236 (2018).
79. M. Z. Zhong *et al.*, *Adv. Funct. Mater.* **28**, 1802581 (2018).
80. S. Yuan *et al.*, *Nano Lett.* **18**, 3172–3179 (2018).
81. J. Guo *et al.*, *Adv. Mater.* **30**, e1705934 (2018).
82. L. Liu *et al.*, *Nature* **605**, 69–75 (2022).
83. M. Badaroglu, "More Moore" in *2021 IEEE International Roadmap for Devices and Systems Outbriefs, Santa Clara, CA, USA, 30 November 2021* (IEEE, 2021); https://doi.org/10.1109/IRDS54852.2021.00010.

ACKNOWLEDGMENTS

Funding: This work was supported by the National Key R&D Program of the Ministry of Science and Technology of China (grant 2022YFA1203801 to X.D.D.), the National Natural Science Foundation of China (grants 62422403, 62205055, and 92464303 to B.Z.; grants 51991340, 51991343, and 52221001 to X.D.D.; and grant 12304214 to J.Q.X.), the Hunan Key R&D Program Project (grant 2022GK2005 to X.D.D.), the Natural Science Foundation of Jiangsu Province (grant BK20220860 to B.Z.), and Fundamental Research Funds for the Central Universities (grant 2242023R40015 to B.Z. and grant JZ2023HGPA0291 to J.Q.X.). This research used the resources of the HPC Platform of Hefei University of Technology. **Author contributions:** X.D.D. designed the research and supervised the project. B.Z. and Z.C.Z. developed and synthesized a series of vdWHs and performed device fabrication, electrical measurements, and analysis. D.L.G., T.C.G., and G.M.H. conducted the growth experiments and transfer of TMDs. P.L. and K.H. performed the electrical measurements. B.Z., Z.C.Z., Q.R., and L.M. performed the STS experiment. J.Q.X. and Z.C. conducted the theoretical calculations and wrote the related discussions. B.Z., Z.C.Z, X.D.D., and X.F.D. co-wrote the manuscript with input from all authors. J.L., P.J.L., and Z.H.N. revised the manuscript. All authors discussed the results and commented on the manuscript. **Competing interests:** Hunan University has filed a Chinese patent application (2025101657875) with X.D.D., Z.C.Z., and B.Z. listed as inventors. The remaining 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.adp8444
Materials and Methods; Supplementary Text; Figs. S1 to S16; Tables S1 to S4; References (*84*–*99*)
Submitted 15 April 2024; resubmitted 2 December 2024; accepted 15 April 2025
10.1126/science.adp8444

VOLCANOLOGY

Minute-scale dynamics of recurrent dike intrusions in Iceland with fiber-optic geodesy

Jiaxuan Li[1,2]*, Ettore Biondi[1], Elías Rafn Heimisson[3], Simone Puel[1,4], Qiushi Zhai[1], Shane Zhang[1], Vala Hjörleifsdóttir[5], Xiaozhuo Wei[1], Elijah Bird[1], Andy Klesh[6], Valey Kamalov[7], Theodór Gunnarsson[8], Halldór Geirsson[3], Zhongwen Zhan[1]

Continuous geodetic measurements near volcanic systems can image magma transport dynamics, yet resolving dike intrusions with high spatiotemporal resolution remains challenging. We introduce fiber-optic geodesy, leveraging low-frequency distributed acoustic sensing (LFDAS) recordings along a telecommunication fiber-optic cable, to track dike intrusions near Grindavík, Iceland, on a minute timescale. LFDAS revealed distinct strain responses from nine intrusive events, six resulting in fissure eruptions. Geodetic inversion of LFDAS strain reveals detailed magmatic intrusions, with inferred dike volume rate peaking systematically 15 to 22 min before the onset of each eruption. Our results demonstrate DAS's potential for a dense strainmeter array, enabling high-resolution, nearly real-time imaging of subsurface quasistatic deformations. In active volcanic regions, LFDAS recordings can offer critical insights into magmatic evolution, eruption forecasting, and hazard assessment.

Monitoring ground deformation in volcanic environments has long been critical for understanding magmatic systems through simple model interpretations (*1*). Classical geodetic techniques, such as electronic distance measurement (EDM), borehole strainmeters, and tilt measurements, remain essential for providing long-term deformation records but have been complemented and partially replaced by satellite-based methods (*2*). Global navigation satellite system (GNSS) and interferometric synthetic aperture radar (InSAR) are the two primary modern geodetic measurements for studying the geometry and dynamics of magma plumbing systems that have appreciably improved the spatial resolution (*2*–*4*). Together, they can reveal magma accumulation and transport over timescales ranging from hours to decades, which is essential for assessing volcanic and seismic hazards (*5*, *6*). For example, dike evolution determined by GNSS data has demonstrated how magma movements can trigger and modulate earthquake swarms (*7*–*10*). A recent study utilizing joint inversions with GNSS and InSAR data on the Reykjanes Peninsula in southwest Iceland revealed an ultrarapid magma flow into the shallower dike accompanied by intense seismicity within a day following 3 years of magma accumulation at depth (*11*). However, capturing the detailed evolution of magmatic intrusions on the minute timescale or in real time remains challenging, as achieving millimeter-level precision with GNSS data typically requires daily averaging (*12*), and it takes days for InSAR to acquire repeating measurements. Thus, the fast-changing dynamics of a magma plumbing system—such as the evolution and emplacement of magma volume, the magma flow rates, and the propagation speed of the magma pulse—remain difficult to resolve. However, such information is critical for understanding the dynamics of magma intrusion and ultimately to forecast whether an intrusion will end as an eruption (*13*).

The emerging distributed acoustic sensing (DAS) technology can convert preexisting telecommunication fiber-optic cables into dense arrays of deformation sensors (*14*–*16*). It has been gaining increasing attention in volcanic regions for its potential to enhance the monitoring and imaging of volcanic systems (*17*–*22*). High-frequency DAS recordings have proven particularly effective in monitoring natural earthquakes and advancing our understanding of earthquake source physics (*23*–*26*). Low-frequency DAS (LFDAS) recordings along fibers in boreholes have been widely used to monitor and characterize fracture growth during hydraulic fracturing (*27*–*29*). Additionally, LFDAS shows great potential for monitoring other geophysical phenomena, such as internal waves and tides (*30*), tsunamis (*31*, *32*), and landslides (*33*).

In this study, we utilized a telecommunication fiber cable as a dense array of strainmeters. We used LFDAS recordings to capture rapid magma intrusion dynamics with minute-scale temporal resolution during the 2023 to 2024 intrusion sequence in Iceland (*11*, *34*). On 10 November 2023, a 15-km-long linear dike formed below the Sundhnúkur crater row in the Svartsengi volcanic system, passing below the town of Grindavík in the Reykjanes Peninsula in southwest Iceland. The diking event was concurrent with the deflation of a magma source at a depth of about 4.1 km to the northwest of Grindavík (*11*). Ten days later, we deployed a DAS interrogator in Keflavík, converting a 100-km-long fiber cable running along the coastline and through Grindavík into a DAS array with about 10,000 recording channels. Following the November dike event, inflation continued in the region, accumulating magma over several months, followed by rapid deflation during dike intrusions. During our recording period from 20 November 2023 to 14 November 2024, nine diking events occurred, six of which resulted in fissure eruptions, and three of which were arrested before reaching the surface. We denote the eruptive events as E1 to E6, whereas the three arrested intrusions preceding events, E4, E5, and E6, are labeled as events E4a, E5a, and E6a, respectviely (Fig. 1 and figs. S1 and S2). The first eruption (E1) started on 18 December 2023, with surface fissures extending from the Sundhnúkur craters to about 1 km north of Stóra-Skógfell (fig. S1A). The second eruption (E2) on 14 January 2024 occurred near Grindavík, destroying several houses there and reaching Mt. Hagafell to the north (fig. S1B). Subsequent eruptions and intrusions, as inferred by the extent of fissures and seismicity, were spatially confined to the areas of the first and second eruptions, except for the last recorded eruption, E6 on 22 August 2024, which extended about 2 km north of Stóra-Skógfell (figs. S1 and S2).

Turning a telecommunication cable into geodetic sensors

We observed distinct strain-rate signals on the DAS array for all nine intrusive events after applying a low-pass filter with a 0.01-Hz cutoff frequency and a spatial median filter to remove common-mode noise [see (*35*) for details]. These signals (smooth blue and red, Fig. 2 and fig. S3) emerged tens of minutes to hours before eruptions and persisted for hours, concurrent with increased seismic activity (fig. S3). Arrested intrusions exhibited weaker strain-rate responses and shorter durations compared with that of eruptive events. Early LFDAS recordings of all intrusive events consistently show a positive strain rate (i.e., extension) between channels 4140 and 4270 and a negative strain rate (i.e., compression) between channels 4270 and 4460 (Figs. 1 and 2), suggesting that the magma fed through a common inlet at the bottom of the dike. For all eruptive events, except for E2, the strain rate along these sections peaked about 15 to 22 min before the eruption (fig. S4A). Additionally, strain-rate polarity flipped shortly after eruptions, whereas no polarity changes were observed during arrested intrusions (Fig. 2). The post-eruption polarity change may indicate a pressure drop in the dike as the fissure formed.

[1]Seismological Laboratory, California Institute of Technology, Pasadena, CA, USA. [2]Department of Earth and Atmospheric Sciences, University of Houston, Houston, TX, USA. [3]Institute of Earth Sciences, University of Iceland, Reykjavík, Iceland. [4]Moody's, Newark, CA, USA. [5]Reykjavik University, Reykjavík, Iceland. [6]NASA Jet Propulsion Laboratory, Pasadena, CA, USA. [7]Valey Kamalov LLC, Gainesville, FL, USA. [8]Google, Mountain View, CA, USA.
*Corresponding author. Email: jxli@uh.edu

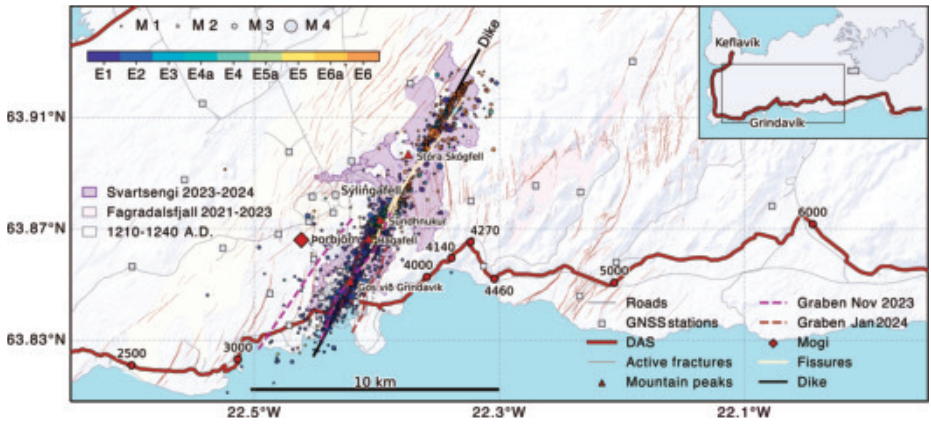


Fig. 1. Location map including seismicity, dike geometry, and fiber geometry. The background map includes roads, the city boundary of Grindavík, GNSS stations, active fractures (*47*), mountain peaks near the dike, and the digital elevation model (*48*). Colored patches represent the lava fields from the 2023 to 2024 eruptive sequence (*36*), the 2021 to 2023 Fagradalsfjall eruptions, and the previous eruption period (1210 to 1240 A.D.). The 100-km DAS array starts from Keflavik (inset) near the airport, going southward along the coast and then going eastward, crossing the town of Grindavík. Select channel numbers between 2500 and 6000 are displayed, with their locations marked by red dots. Earthquake swarms for all nine dike intrusions are color-coded by event, with symbol size scaled by magnitude (M). Lava fissures during all six eruptions are shown. Two sets of grabens formed, one during the November diking event and the other during the January eruption. A 14-km-long linear dike is fitted using the surface lava fissures, and the magma domain is modeled as a Mogi point source.

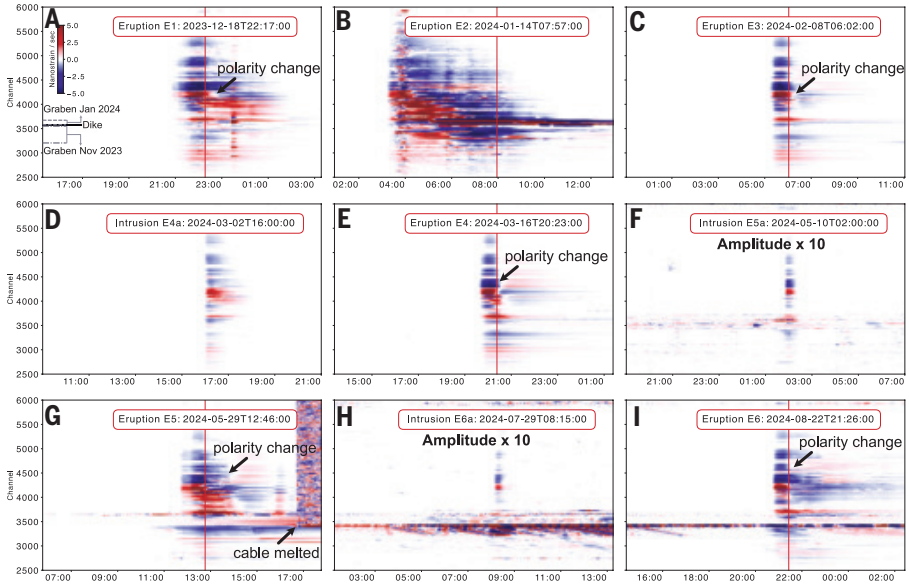


Fig. 2. Denoised strain rate recordings over 12 hours during the nine intrusive events. (**A**) Recording of eruption E1. The intersections of the fiber cable with the dike (black solid line) and grabens (gray dashed lines) formed in November 2023 and January 2024 are marked. The vertical red line represents the eruption time. The black arrow marks the polarity change shortly after the eruption. (**B**, **C**, **E**, **G**, and **I**) Similar to (A) but for recordings of eruptions E2 to E6, respectively. The strong noise in (G) after around 16:40 (UTC) was caused by the destruction of the fiber cable by lava. (**D**, **F**, and **H**) Recordings of arrested intrusions E4a, E5a, and E6a, respectively. The strain rate amplitudes for E5a (F) and E6a (H) are multiplied by a factor of 10 to enhance the visibility of these two weak events.

The strain-rate signals for E2 lasted the longest, with a duration of more than 10 hours, and displayed a progressively localized spatial pattern (Fig. 2B and fig. S3B), suggesting that the magma was approaching the fiber cable. Around 6:00 (UTC) on 14 January 2024 a large strain-rate response emerged, centered around channel 3600. It eventually accumulated to the largest strain recorded on this DAS array with no saturation, with an amplitude of ~1000 microstrain. This signal corresponds to the development of a graben beneath the fiber cable, formed during this dike intrusion and similar to that formed during the November dike intrusion (*11*). By contrast, strain responses corresponding to the distant dike openings mostly had amplitudes of tens of microstrain. During arrested intrusions E5a on 10 May 2024 and E6a on 29 July 2024, strain signals lasted less than an hour and had maximum amplitudes of around 0.1 microstrain.

At about 16:30 (UTC) on 29 May 2024, the fiber cable west of Grindavík was melted by lava during eruption E5, which disabled recordings beyond channel 3437. A week later, a 650-m temporary surface fiber was deployed by the electronic communication company Ljósleidarinn to replace the damaged section, restoring both telecommunication and DAS array recordings. However, this surface cable segment (channel 3389 to 3454) exhibited higher noise levels, as seen in E6a and E6, and was excluded from subsequent analyses.

Dike intrusion dynamics from LFDAS recordings

We integrated strain-rate signals from channel 2500 to 6000 (Fig. 2) to derive strain during intrusive events to further infer the evolution of dike opening and magma domain deflation [see (*35*) for details]. The dike was modeled as a predefined vertical rectangular plane, with its strike determined by fitting a common line to the surface fissures mapped from all eruptions (*36*). We assumed that the dike plane is vertical, as the distribution of earthquakes at the bottom of the dike is approximately aligned with the surface lava fissures (Fig. 1). The magma domain was treated as a point source (Mogi source; red diamond, Fig. 1), located in the same position as that inferred during the November dike event (*11*). The forward strain modeling along the fiber assumes an elastic half-space model and uses analytical solutions for the dike and the point source (*1*, *37*). We further processed strain-rate recordings in fig. S3 to remove spikes and long-period background noise (Fig. 2). Then we integrated the strain rate data and multiplied by a coupling factor of 3.2 to obtain the physical ground deformation strain (fig. S5). We estimated this factor by analyzing DAS

records in comparison with data from a local broadband seismometer (*38*) and nearby GNSS stations [see (*35*) for details]. We extracted the LFDAS strain across channels every 5 min and performed the inversion independently for each time step (fig. S6).

For all nine intrusive events, the models show that the dike openings began between the Sundhnúkur craters and Mt. Stóra-Skógfell (Fig. 3), with earthquake swarms occurring below the opening at a depth of about 4.5 km [see (*35*) for details]. This is similar to the location of initial seismicity starting around 7 a.m. on 10 November 2023 (*11*). This pattern suggests a common magma feeding inlet at the bottom of the dike for all intrusive events, including the November dike event. In eruptive events, E1 to E6, the dike opening breached the surface, aligning with locations of eruptive fissures (Fig. 3). By contrast, for noneruptive event E4a, the dike opening was trapped at a depth of about 3 km (Fig. 3). For the other two noneruptive events, E5a and E6a, the maximum opening occurred at depths of about 2 km (Fig. 3).

We observed a clear spatial migration pattern of dike opening during eruptions E1, E2, E5, and E6 (Fig. 3). Among these, event E2 shows the longest migration. The dike opened at a depth of about 2 to 3 km, like that observed during E4a, before propagating horizontally toward Grindavík and vertically to shallower depths (Fig. 3). The maximum opening eventually reached the surface just north of Grindavík, where the lava erupted. By tracking the location of maximum local opening during intrusion, we estimate the initial migration speed of the magma pocket to be about 0.9 m s^{-1}. The migration then slowed down to about 0.2 m s^{-1}, when the magma pocket reached a depth of about 1 km (figs. S7 and S8). Event E1 started below Mt. Stóra-Skógfell and migrated southwest to Mt. Hagafell (Fig. 3). Both E5 and E6 started with a similar opening pattern to the preceding noneruptive intrusions E5a and E6a but with substantially larger amplitudes (Fig. 3). Whereas event E5 propagated about 2 km southward, E6 extended 2 km north of Mt. Stóra-Skógfell (Fig. 3).

We can determine the volume history of both the dike opening and point source deflation from the inversion (fig. S9). The inferred opening volume of the dike is generally larger than the deflation volume, as also observed during the November dike intrusion (*11*). This discrepancy can be partly explained by the greater compliance of a dike compared with that of an ellipsoidal magma chamber (*39*). Furthermore, large variation in volume ratios (1.4 to 5.9) may reflect melts originating from different sills within the magma domain, each with different volatile content and compressibility (*34*). In the following, we focus on changes in dike volume, as the point source volume is less well resolved owing to the smaller magnitude of deflation-induced strain and potential noise caused by local fault movements. Eruption E2 has the largest dike volume among all intrusive events, with a total volume of about 23.5 million m^3 and a peak volume rate exceeding 3700 m^3 s^{-1} (Fig. 4). Eruptions E1, E3, and E6 have associated dike volumes of approximately 11.1, 7.2, and 9.3 million m^3, respectively. The arrested intrusion E4a and eruption E4 have similar dike volumes and occurred just 2 weeks apart, with a combined dike volume of about 9.4 million m^3. Eruption E5 reached a dike volume of 5.8 million m^3 when the fiber was melted by the lava and data acquisition beyond channel 3389 was seized. Without this damage, the inferred volume would likely have continued to increase. The two smaller intrusive events, E5a and E6a, only result in a dike volume of about 0.1 million m^3. In total, based on available measurements, these nine intrusive events produced a dike volume of about 67 million m^3.

A notable feature of the dike growth history is the similar evolution of volume rates for all eruptive events, except E2 (Fig. 4B and fig. S4B). Dike volume rate first increased rapidly, reaching a peak 15 to 22 min before the eruption, and then gradually declined to zero. Such a repetitive pattern suggests a consistent magma plumbing geometry and relatively stable stress state of the volcanic system, where the magma pressure builds in the magma domain during inflation, driving the magma flow through a conduit into the dike's inlet during deflation, and eventually growing the dike tip to the surface. Sometimes, the dike fails to breach the surface, resulting in arrested intrusions, such as E4a, E5a, and E6a. Eruption E2 is distinguished from other eruptions by its long duration, large dike opening volume, and a two-stage volume increase in which an initially trapped intrusion resumed growth and eventually led to an eruption. This second phase of volume growth may be attributed to the secondary supply of magma from the southwest section of the dike below Grindavík (Fig. 3F) or to the

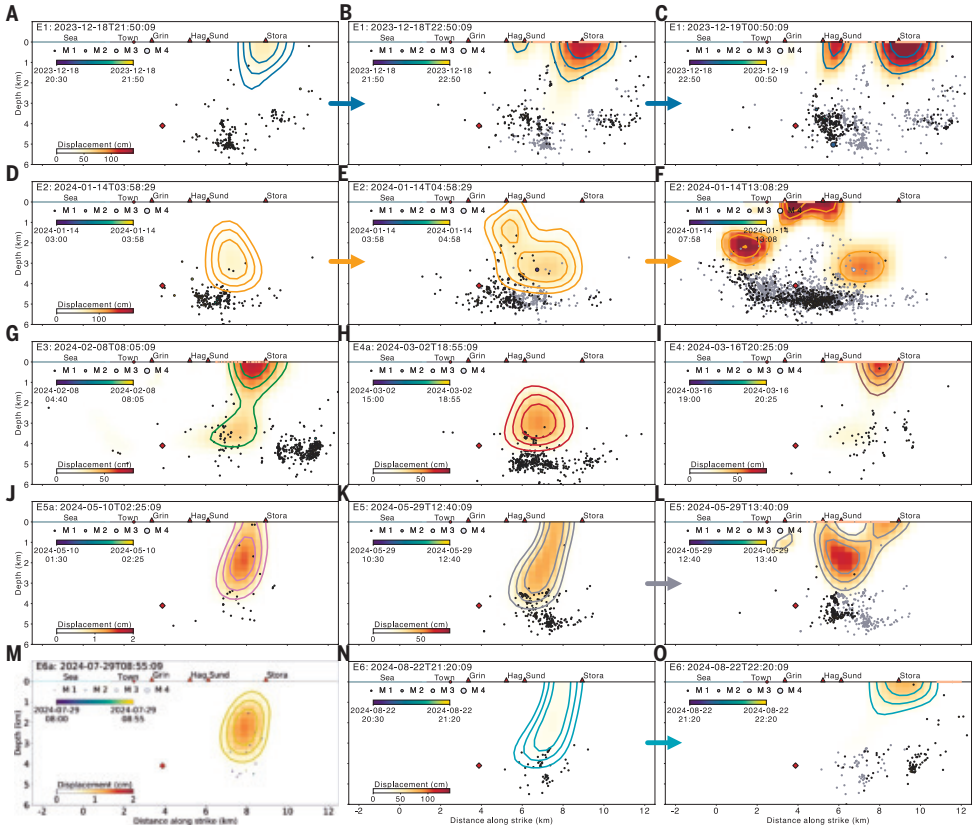


Fig. 3. Distribution and evolution of dike opening and seismicity for nine intrusive events. (**A** to **C**) The temporal evolution of event E1 is illustrated by colored arrows connecting snapshots of dike opening distribution and its associated seismicity [see (*35*) for details]. The contours represent 40, 60, and 80% of the maximum opening and are color-coded by event. The red diamond represents the projection of the Mogi source onto the dike plane. (**D** to **F**) Similar to (A) to (C) but for event E2. (**G** to **I**) Dike opening distribution for events E3, E4a, and E4, respectively. (**J**) Dike opening for event E5a, which is similar to the initial opening distribution during event E5 in (**K** and **L**). (**M**) Dike opening for event E6a, which is similar to the initial opening distribution during event E6 in (**N** and **O**). Note that the maximum amplitude in the displacement color bar is 140 cm for E1 and E6; 180 cm for E2; 80 cm for E3, E4a, E4, and E5; and 2 cm for E5a and E6a.

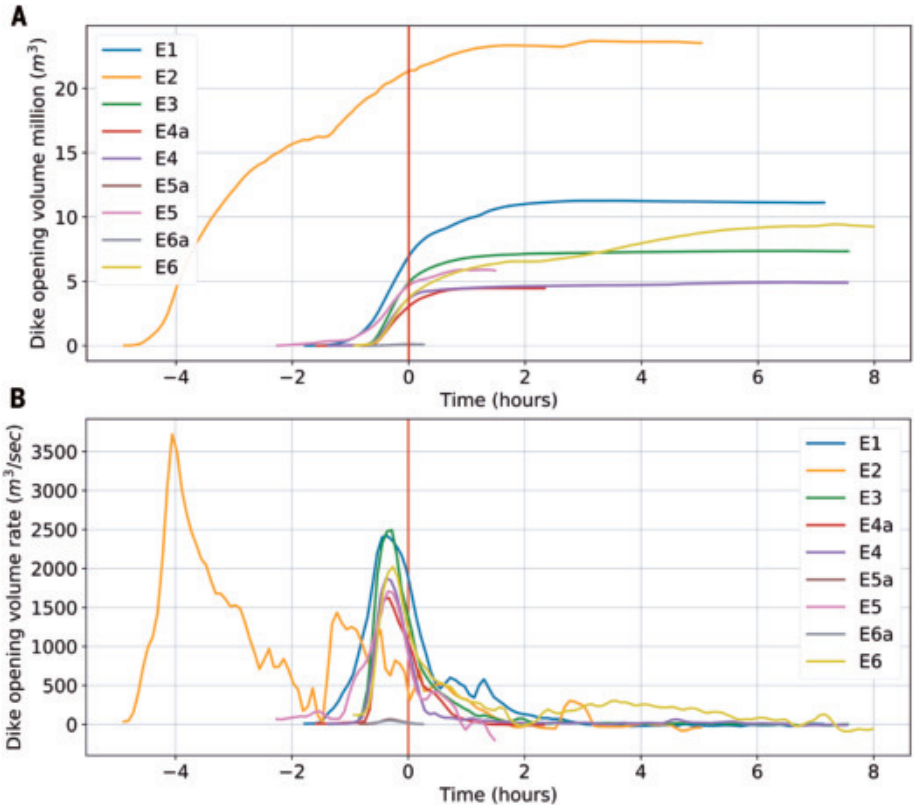


Fig. 4. Temporal evolution of dike volume and dike volume rate. (**A**) Dike volume history for all nine intrusive events. The time series are aligned by the eruption time (vertical red line at time 0). For noneruptive events (E4a, E5a, and E6a), each time series is shifted such that the dike volume rate correlates with that of eruption E4. (**B**) Dike volume rate history. Note that, except for E2, all eruptions reach their peak dike volume rate 15 to 22 min before the eruption onset. The two-stage volume increase of E2 may be caused by the secondary supply of magma or pressure increase due to basalt vesiculation.

increased dike pressure from basalt vesiculation at shallower depths (*13*, *40*). Eruption E6 also shows a two-stage dike volume increase, though the second stage occurred after the eruption began and had a smaller volume increase.

Physical model

We propose that each intrusive event was affected by previous events, among which the November diking event had the most impact. The November dike intrusion did not breach the surface, possibly owing to accumulated tensional tectonic stress and negative buoyancy in the shallow crust (*13*) that favored lateral growth, forming a 15-km-long dike with up to 8 m of inferred opening and a total dike volume about 130 million m^3 (*11*). This is about twice the combined volume of subsequent intrusions by 14 November 2024. This extensive opening at depths ranging from 1.5 to 5 km increased compressive stress at depth but decreased compression in the shallower crust relative to the background stress state (*35*). These stress changes facilitated the dike reaching shallow levels and later eruptions. Eruptions E1 and E2 filled the opening gap in the upper 1.5 km and were spatially separated around Mt. Hagafell. During the early stage of the E2 intrusion (Fig. 3, D to F), its vertical propagation was stalled by the opening boundary of E1, likely because of the negative vertical stress gradient locally induced by E1. Later eruptions and arrested intrusions arranged themselves in the upper crust where there was stress and opening deficit (figs. S10 and S11), suggesting self-organization of diking events and the possibility to estimate the location of the following dike and eruption.

The history of the observed dike volume rate can be modeled by using a simple lumped model with a coupled system of a dike and a crustal magma chamber (*13*). In this model, the dike grows as a semi-ellipsoid, with the magma inlet at its base. Temporal evolution of this system is governed by three dimensionless parameters: α, Ψ, and R [see (*35*) for details]. Parameter α reflects the competition between the combined negative buoyancy in the shallow crust and the vertical tectonic stress gradient against the driving pressure at the dike inlet. Parameter Ψ represents the volume and compressibility ratio between the dike and magma domain. Eruptions are favored by a small α, indicating higher driving pressure and lower resistance, and a small Ψ, indicating a large magma chamber relative to the dike, which makes the pressure within the magma chamber harder to deplete. Parameter R characterizes the hydraulic property of the magma plumbing system and represents the relative rate between magma flow into the dike over dike propagation. A larger R indicates a more conductive conduit and favors eruptions.

We can infer some physical properties of the magma plumbing system, including the initial dike overpressure, the vertical gradient of the tectonic normal stress, and the volume of the magma domain, by using the probable parameter ranges of α, Ψ, and R constrained by the observed dike volume rate patterns (fig. S12). From the observation, the typical time interval between the dike volume rate peak and the eruption is about 20 min (Fig. 4 and fig. S4), and the depth of the dike inlet is about 2.5 km (Fig. 3). Assuming a crustal shear modulus μ of 15 GPa and Poisson's ratio v of 0.27, we can estimate the initial overpressure as a function of magma viscosity (fig. S12C). For example, a magma viscosity of 50 Pa*s would infer an initial overpressure to be about 3.5 MPa. In addition, the set of the probable parameter α indicates a low effective density [see (*35*) for details]. If we assume a magma density of ρ_m of 2610 kg/m^3 and a crust density of ρ_r of 2350 kg/m^3 (*10*, *11*), this implies a positive vertical gradient of normal stress locally induced by the November dike intrusion (fig. S12D). We can also estimate the volume of the magma chamber, which varies with Ψ and the effective compressibility of the magma chamber β_c. For example, with $\Psi = 0.71$, an effective compressibility of 0.14 GPa^{-1} implies a magma domain volume of about 10 km^3. An effective compressibility of 0.07 GPa^{-1} will require the magma domain to be approximately 21 km^3 (fig. S12E). The substantial volume of the chamber, despite uncertainties related to its compressibility, indicates its capacity to store and release large amounts of magma.

Implications for volcano monitoring and fiber-optic geodesy

Our results from the 2023 to 2024 intrusion sequence in Iceland demonstrate how LFDAS recordings along fiber cables can complement current geodetic observation systems, offering minute-scale deformation monitoring previously unattainable with conventional geodetic techniques. Although InSAR can resolve millimeter-level surface deformation over a two-dimensional area, its temporal resolution is limited, often requiring days between measurements. Similarly, GNSS continuously monitors surface displacements at discrete points but requires substantial time averaging to achieve millimeter-level resolution. Classical geodetic techniques, such as borehole strainmeters and tilt meters, require substantial financial investment to provide

continuous but single-point measurements. In addition, EDM is labor intensive and expensive to operate. By contrast, LFDAS recordings provide both high spatial and temporal resolution by converting fiber cables spanning tens of kilometers into dense arrays of thousands of strainmeters spaced just meters apart. Additionally, lower noise levels can be achieved in LFDAS strain recordings through array processing and denoising (Fig. 2). For example, the deformation induced by the arrested intrusions E5a and E6a falls below the noise threshold of both GNSS and InSAR (fig. S13). This approach improves our understanding of rapid magma flow dynamics and potentially provides valuable insights into how the stress induced by dike intrusions interacts with earthquakes (*9*, *10*).

LFDAS requires minimal processing to detect dike intrusion–induced strain, making it particularly attractive for real-time monitoring, early warning, and hazard assessments of volcanic eruptions. Since 27 February 2024, we have operated an early warning system to assist the Icelandic Meteorological Office (IMO) in issuing alerts and evacuation orders. We successfully assisted the IMO in issuing early warnings for eruptions E4 on 16 March 2024 and E5 on 29 May 2024 and in refraining from sending a warning for the arrested intrusion E4a on 2 March 2024. Subsequently, we implemented an automatic detection system that sends an alert when the averaged LFDAS strain rate exceeds a preset threshold of 1 nanostrain per second, determined empirically based on previous events. For the recent eruption E6, this system issued an automatic early warning notification to IMO 26 min before the eruption occurred. With a predefined dike geometry, we expect LFDAS to provide nearly real-time imaging of dike intrusions to assess the location and likelihood of an eruption.

Currently, our inversion of dike openings relies solely on LFDAS recordings along a single fiber cable. However, joint inversions incorporating conventional geodetic measurements, such as GNSS and InSAR, or potentially multiple fiber cables at different azimuths, can greatly enhance the spatial resolution. For example, a joint inversion with LFDAS and GNSS improves the resolution on the dike's northeast section in tests with different opening depths (*35*) (fig. S16). Additionally, GNSS and InSAR are critical to long-term monitoring on timescales of hours to decades, and it remains to be understood whether signals with similar timescales can be extracted from LFDAS. Nevertheless, as the Reykjanes Peninsula has since 2019 entered a new era of volcanic unrest (*41*, *42*), similar studies based on a comprehensive dataset will be needed. Moreover, offshore fiber cables, some of which are near active underwater volcanoes (*21*, *43*, *44*), offer a rare opportunity to monitor areas where conventional geodetic techniques face major challenges. By converting preexisting fiber-optic infrastructure into high-resolution geodetic networks and integrating it with conventional geodetic methods, we anticipate new possibilities for monitoring and understanding volcanic activity as well as other geological processes, such as fault creeping and hillslope deformation (*45*, *46*), heralding the emergence of fiber-optic geodesy.

REFERENCES AND NOTES

1. K. Mogi, *Bull. Earthq. Res. Inst. Univ. Tokyo* **36**, 99–134 (1958).
2. J. Fernández, A. Pepe, M. P. Poland, F. Sigmundsson, *J. Volcanol. Geotherm. Res.* **344**, 1–12 (2017).
3. D. Massonnet, P. Briole, A. Arnaud, *Nature* **375**, 567–570 (1995).
4. M. P. Poland, H. A. Zebker, *Bull. Volcanol.* **84**, 27 (2022).
5. V. Acocella *et al.*, *Nat. Rev. Earth Environ.* **5**, 5–22 (2024).
6. K. R. Anderson *et al.*, *Annu. Rev. Earth Planet. Sci.* **52**, 21–59 (2024).
7. P. Segall, A. L. Llenos, S.-H. Yun, A. M. Bradley, E. M. Syracuse, *J. Geophys. Res. Solid Earth* **118**, 5785–5804 (2013).
8. F. Sigmundsson *et al.*, *Nature* **517**, 191–195 (2015).
9. R. G. Green, T. Greenfield, R. S. White, *Nat. Geosci.* **8**, 629–632 (2015).
10. E. R. Heimisson, P. Segall, *J. Geophys. Res. Solid Earth* **125**, e2019JB018141 (2020).
11. F. Sigmundsson *et al.*, *Science* **383**, 1228–1235 (2024).
12. W. Bertiger *et al.*, *J. Geod.* **84**, 327–337 (2010).
13. P. Segall, P. Cervelli, S. Owen, M. Lisowski, A. Miklius, *J. Geophys. Res.* **106** (B9), 19301–19317 (2001).
14. N. J. Lindsey *et al.*, *Geophys. Res. Lett.* **44**, 11–792 (2017).
15. P. Jousset *et al.*, *Nat. Commun.* **9**, 2509 (2018).
16. Z. Zhan, *Seismol. Res. Lett.* **91**, 1–15 (2019).
17. G. Currenti, P. Jousset, R. Napoli, C. Krawczyk, M. Weber, *Solid Earth* **12**, 993–1003 (2021).
18. P. Jousset *et al.*, *Nat. Commun.* **13**, 1753 (2022).
19. A. Fichtner *et al.*, *The Seismic Record* **2**, 148–155 (2022).
20. E. Biondi, W. Zhu, J. Li, E. F. Williams, Z. Zhan, *Sci. Adv.* **9**, eadi9878 (2023).
21. G. Currenti *et al.*, *Sci. Rep.* **13**, 4641 (2023).
22. C. Caudron *et al.*, *Sci. Rep.* **14**, 3128 (2024).
23. J. Yin *et al.*, *Geophys. Res. Lett.* **50**, e2023GL103045 (2023).
24. W. Zhu *et al.*, *Nat. Commun.* **14**, 8192 (2023).
25. J. Li, W. Zhu, E. Biondi, Z. Zhan, *Nat. Commun.* **14**, 4181 (2023).
26. J. Li, T. Kim, N. Lapusta, E. Biondi, Z. Zhan, *Nature* **620**, 800–806 (2023).
27. G. Jin, B. Roy, *Leading Edge (Tulsa Okla.)* **36**, 975–980 (2017).
28. Y. Liu, G. Jin, K. Wu, G. Moridis, *SPE J.* **26**, 359–371 (2021).
29. Y. Liu, G. Jin, K. Wu, G. Moridis, *SPE J.* **26**, 2703–2715 (2021).
30. E. F. Williams *et al.*, *J. Geophys. Res. Oceans* **128**, e2023JC019980 (2023).
31. H. Xiao, Z. J. Spica, J. Li, Z. Zhan, *Geophys. Res. Lett.* **51**, e2023GL106767 (2024).
32. T. Tonegawa, E. Araki, *Geophys. Res. Lett.* **51**, e2024GL108714 (2024).
33. S. M. Ouellet *et al.*, *Nat. Commun.* **15**, 6239 (2024).
34. S. W. Matthews *et al.*, *Science* **386**, 309–314 (2024).
35. Materials and methods are available as supplementary materials.
36. LMI, Landmælingar Íslands (2024); https://atlas.lmi.is/mapview/?application=DEM.
37. Y. Okada, *Bull. Seismol. Soc. Am.* **82**, 1018–1040 (1992).
38. Q. Zhai *et al.*, *J. Geophys. Res. Solid Earth* **130**, e2024JB030288 (2025).
39. E. Rivalta, P. Segall, *Geophys. Res. Lett.* **35**, 35 (2008).
40. L. Wilson, J. W. Head III, *J. Geophys. Res.* **86** (B4), 2971–3001 (1981).
41. F. Sigmundsson *et al.*, *Nature* **609**, 523–528 (2022).
42. A. Caracciolo *et al.*, *Earth Planet. Sci. Lett.* **621**, 118378 (2023).
43. W. W. Chadwick Jr., S. L. Nooner, D. A. Butterfield, M. D. Lilley, *Nat. Geosci.* **5**, 474–477 (2012).
44. M. A. Clare *et al.*, *Science* **381**, 1085–1092 (2023).
45. R. A. Harris, *Rev. Geophys.* **55**, 169–198 (2017).
46. Y. Xu *et al.*, *Geophys. Res. Lett.* **51**, e2024GL110210 (2024).
47. C. Ducrocq *et al.*, *Bull. Volcanol.* **86**, 14 (2024).
48. LMI, Landmælingar Íslands (data collected by the Icelandic Institute of Natural History and the National Survey of Iceland) (2024); https://atlas.lmi.is/mapview/?application=umbrotasja.
49. J. Li, E. Biondi, E. Bird, Z. Zhan, Iceland eruption DAS data (2024); https://data.caltech.edu/records/tw5s6-nd034.

ACKNOWLEDGMENTS

We thank Ljósleidarinn for providing the fiber access. We thank M. M. Parks for valuable discussions on dike geometry and K. Jónsdóttir, S. Barsotti, E. B. Gestsson, and P. Erlendsson for coordinating the LFDAS volcanic eruption early warning efforts. J.L. thanks W. Zhu for discussions on LFDAS with telecommunication fiber cables in early 2023. The fiber-optic geodesy method is part of a provisional US patent filed by the California Institute of Technology (CIT file no. CIT-9251-P). **Funding:** This work is supported by the Gordon and Betty Moore Foundation, US National Science Foundation (NSF) (EAR-1848166), and NSF Center for Geomechanics and Mitigation of Geohazards (1822214). **Author contributions:** Conceptualization: J.L.; Methodology: J.L., E.Bio., E.R.H., S.P.; Investigation: J.L, E.Bio., E.R.H., Q.Z., S.Z., V.H., X.W., H.G.; Data collection and fieldwork: E.Bir., A.K., V.K., T.G.; Supervision: Z.Z. Writing – original draft: J.L., E.Bio., E.H., Q.Z.; Writing – reviewing & editing: J.L., E.Bio., E.R.H., S.P., Q.Z., S.Z., V.H., X.W., E.Bir., A.K., V.K., T. G., H.G., Z.Z. **Competing interests:** The authors declare that there are no competing interests. **Data and materials availability:** The downsampled LFDAS data before and after denoising for all nine intrusive events in this study are available at CaltechDATA (*49*). **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.adu0225
Materials and Methods; Figs. S1 to S17; Table S1; References (*50*–*58*); Movies S1 to S9

Submitted 21 October 2024; accepted 14 April 2025; published online 24 April 2025

10.1126/science.adu0225

PHOTOPHYSICS

Differential absorption of circularly polarized light by a centrosymmetric crystal

Katherine A. Parrish[1]†, Andrew Salij[2,3]†, Kendall R. Kamp[2]†, Evan Smith[1], M. Iqbal Bakti Utama[4], Anders J. Bergsten[1], Rachel Czerwinski[1], Mackinsey A. Smith[1], Mark C. Hersam[2,4], Kenneth R. Poeppelmeier[2]*, Randall H. Goldsmith[1]*, Roel Tempelaar[2]*

Crystalline solids are governed by universal structure-property relationships derived from their crystal symmetry, leading to paradigmatic rules on what properties they can and cannot exhibit. A long-held structure-property relationship is that centrosymmetric crystals cannot differentially absorb circularly polarized light. In this study, we demonstrate the design, synthesis, and characterization of the centrosymmetric material $Li_2Co_3(SeO_3)_4$, which violates this relationship not by defying symmetry-imposed selection rules but by invoking a photophysical process not previously characterized for crystalline solids. This process originates from an interference between linear dichroism and linear birefringence, referred to as LD-LB, and involves strong chiroptical signals that invert upon sample flipping. In addition to enabling a chiroptical response under centrosymmetry, this process opens up photonic engineering opportunities based on crystalline solids.

Textbook structure-property relationships informed by crystal symmetry are instrumental for guiding the design of solids. For example, the presence or absence of chiral properties of a solid are intricately related to its crystal structure. Perhaps the most widely accepted structure-property relationship in this regard is that crystals featuring centrosymmetry (manifesting inversion symmetry) do not allow for the differential absorption of circularly polarized (CP) light (*1*–*8*), meaning that they should not display circular dichroism (CD). Although natural optical activity (OA) is indeed forbidden without breaking inversion symmetry (*9*), here we show that an alternative photophysical process allows certain centrosymmetric crystals to differentially absorb CP light. The process at play originates from an interference between linear dichroism (LD) and linear birefringence (LB) associated with oblique optical axes, referred to as LD-LB (*10*–*13*). By characterizing this process on the basis of crystal symmetries, we demonstrated its optical selection rules to be quite distinct from those of OA, allowing CP light to be selectively absorbed within certain crystal classes obeying centrosymmetry, which we confirmed experimentally. Notably, whereas OA relies on both electronic and (comparatively weak) magnetic dipole interactions, LD-LB can result solely from electronic dipole contributions, producing a remarkably high associated degree of optical rotation. This combination of factors renders LD-LB–active crystals of particular fundamental and technological relevance.

Since its discovery in 1969 (*11*), LD-LB has been often regarded as an optical artifact obscuring OA measurements of cholesteric liquids (*14*, *15*), and thus it is often referred to as "apparent" CD. After more systematic inquiries made on crystals with embedded chromophores (*16*–*18*), it has since been recognized that LD-LB, rather than a mere artifact, is a chiroptical process complementary to OA (*19*) that could serve as a structural probe (*20*–*22*), while also providing a means of engineering CP absorptive and luminescent materials. In recent years, observations of LD-LB have been reported for organic thin films (*23*–*25*), metal halide perovskites (*20*–*22*), and nanostructures (*26*, *27*). Most attributions to LD-LB were made on the basis of a chiroptical response inversion upon sample flipping—a feature not possible for OA, but a hallmark of LD-LB (*28*). This inversion was even used to make chiral microcavities (*26*, *27*). Notably, although LD-LB has been theoretically related to the electronic structure of molecules (*11*), design principles based on crystal classes have remained lacking.

Here, we identified crystalline $Li_2Co_3(SeO_3)_4$ (LCSO) as a centrosymmetric material that also shows selective and asymmetric response to CP light, following design and synthesis informed by a crystal class characterization of LD-LB. As compared to examples of LD-LB seen in organic and hybrid materials, in this case the chiroptic behavior was observed in macroscopic single crystals and in a manner that is purely attributable to LD-LB as opposed to a mixture of LD-LB and conventional OA. Our discovery thus not only prompts a rethinking of textbook rules but also opens opportunities for the engineering of chiral light-matter interactions—including chiral polaritons (*27*, *29*–*31*) that couple photonic and matter angular momentum—alongside the dielectric, plasmonic, and metamaterials that make up the modern chiroptical functionality toolbox (*32*–*41*).

Crystal class characterization of LD-LB

It is first necessary to develop the crystal class characterization of LD-LB. Broadly speaking, such crystal class characterization follows from application of crystal symmetry transformations on the formal expressions governing the properties of interest. The expression governing OA has been well-characterized and is caused, to first order, by a difference of tensor product elements of magnetic and electronic transition dipoles (*5*). Disambiguating these elements requires a broken centrosymmetry and leads to a crystal class characterization of OA (Fig. 1A). The leading contribution to LD-LB, however, is of second order in the direction of light propagation because the underlying mechanism entails differential absorption and birefringence along axes offset from one another, which in tandem produce an asymmetric absorption of CP light (*11*). This effect relies on the interference between two nondegenerate transition dipoles, which is minimized when

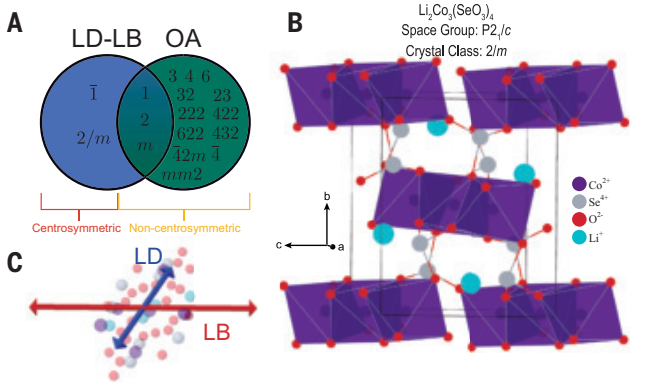


Fig. 1. Structural basis for the occurrence of LD-LB. (**A**) Crystal classes for linear dichroism–linear birefringence (LD-LB) and natural optical activity (OA). (**B**) LCSO structure comprising edge-sharing trimers of CoO_6 octahedra linked through trigonal pyramidal SeO_3 units stacked along the *b* axis. (**C**) Single unit cell alongside axes of maximum LD (blue) and LB (red) in the *ac* plane at a wavelength of 635 nm, as obtained through evaluation of the LCSO dielectric tensor through TD-DFT (supplementary text S9). The Co atoms centering the Co-oxygen octahedra are positioned on inversion centers with one site at the body center of the unit cell. This motif stacking enables light propagation from either side of the crystal to give opposing signals.

[1]Department of Chemistry, University of Wisconsin–Madison, Madison, WI, USA. [2]Department of Chemistry, Northwestern University, Evanston, IL, USA. [3]Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, USA. [4]Department of Materials Science and Engineering, Northwestern University, Evanston, IL, USA. ***Corresponding author. Email: krp@northwestern.edu (K.R.P.); rhg@chem.wisc.edu (R.H.G.); roel.tempelaar@northwestern.edu (R.T.)** †These authors contributed equally to this work.

the interdipole angle approaches 0° or 90°, maximized when the angle amounts to ±45°, and which experiences a signal sign flip when the angle is inverted (*11*). LD-LB gives rise to a form of chirality that reflects through a mirror line, rather than a mirror plane as is commonly found for chiral molecules, and which can be interpreted based on the two-dimensional chiral pattern made by the two transition dipoles when projected into the optical polarization plane (*19*): The angle between these dipoles can be thought of as positive for light traveling in one direction and negative for light traveling in the opposite direction, yielding the aforementioned inverted response upon sample flipping.

Higher-order optical phenomena such as LD-LB are conveniently described by means of Mueller calculus, wherein light polarizations are described by a Stokes vector and a series of light-matter interactions are accounted for by the sequential operation of Mueller matrices onto such a vector. For LD-LB, the CD response takes the form (*42*)

$$\text{LD-LB} \propto d_1\beta_2 - d_2\beta_1 \tag{1}$$

Here, d_1 and β_1 denote the LD and LB contributions, respectively, associated with an arbitrarily chosen reference direction in the plane perpendicular to the light-propagation direction, and d_2 and β_2 are the corresponding contributions associated with a direction rotated by 45° in this plane. Denoting the reference direction x, and the orthogonal direction in the polarization plane y, these contributions can be expressed in terms of the dielectric tensor as

$$d_1 \propto \Im\left(\sqrt{\in_{xx}} - \sqrt{\in_{yy}}\right) \tag{2}$$

$$\beta_1 \propto \Re\left(\sqrt{\in_{xx}} - \sqrt{\in_{yy}}\right) \tag{3}$$

$$d_2 \propto \Im\left(\sqrt{\in_{xx} + \in_{yy} + 2\in_{xy}} - \sqrt{\in_{xx} + \in_{yy} - 2\in_{xy}}\right) \tag{4}$$

$$\beta_2 \propto \Re\left(\sqrt{\in_{xx} + \in_{yy} + 2\in_{xy}} - \sqrt{\in_{xx} + \in_{yy} - 2\in_{xy}}\right) \tag{5}$$

where $\Im$ and $\Re$ denote the imaginary and real parts of complex-valued quantities. For LD-LB to be nonzero, it is required that $\in_{xx} \neq \in_{yy}$ and $\in_{xy} \neq 0$. Examination of dielectric tensor forms specific to the various space groups (supplementary text S2) demonstrates that these properties are conserved only for triclinic and monoclinic crystals. In particular, a crystal displaying LD-LB must be biaxial in order to have a sufficiently nonsymmetric refractive index ellipsoid but should not have three C_2 axes as in rhombohedral crystals.

On the basis of the above principles, we established the crystal class characterization of LD-LB, the result of which is summarized in Fig. 1A. As shown, the crystal classes compatible with LD-LB and OA do partially overlap, with 1, 2, and m supporting both phenomena. Notably, LD-LB permits inversion of Cartesian xyz axes, and as such, the centrosymmetric classes $\bar{1}$ and $2/m$ support LD-LB but not OA. Conversely, many crystal classes support OA but not LD-LB, because OA relies on different broken symmetries of the gyration tensor instead of the dielectric tensor as in LD-LB (*7*).

Chiroptical response of a centrosymmetric solid

The above crystal class characterization predicts the existence of a solid with a centrosymmetric crystal lattice that differentially absorbs CP light. To test this prediction, we designed an appropriate crystalline material, considering additional desirable properties such as the presence of nondegenerate electronic transition dipoles in the visible optical regime. The material we selected is LCSO (Fig. 1B), which was originally identified as a (previously) rare example of alkali-cobalt-selenium compounds with large polyhedra distortions (*43*). LCSO crystallizes in the monoclinic space group $P2_1/c$, which belongs to the centrosymmetric crystal class $2/m$, for which LD-LB is symmetry-allowed but OA is not. To substantiate the proposal that LCSO shows a desirable chiroptical response through LD-LB, we evaluated Eq. 1 on the basis of time-dependent density functional theory (TD-DFT) calculations for this material (supplementary text S4). We then synthesized this material using a procedure modified from its originally reported synthesis (materials and methods, and supplementary text S5), yielding single crystals with thicknesses ranging from 5 to 40 μm and side lengths ranging from 10 to 200 μm. In much of the visible spectrum, LCSO has characteristic LD and LB axes that are offset from one another in the 100 plane, yielding an LD-LB effect (Fig. 1C).

CD measurements were obtained for each crystal by using a home-built 635-nm transmission microscopy setup (materials and methods). Having micron-scale spatial resolution while probing single domains is important because multiple proximal or multidomain crystals may dilute the LD-LB effect. The CD signal in units of ellipticity (mdeg, millidegrees) was obtained as (*44*)

$$\text{CD} = 32980 \times \log_{10}\frac{I_R}{I_L} \tag{6}$$

where I_R and I_L denote the intensity of transmitted right-handed and left-handed CP light, respectively. Shown in Fig. 2, A and B are CD images of a representative LCSO crystal in the forward and backward configurations. The crystal shows a strong CD response with a magnitude on the order of ~0.5×10^4 mdeg. Moreover, an inversion of the CD signal is observed upon flipping the sample, as substantiated by a histogram of the pixel values (Fig. 2C). We have conducted an extended spectrally resolved analysis of LCSO including absorption, LD, CD, and the dissymmetry (g) factor (supplementary text S10). This analysis (fig. S16) shows the response at 635 nm captured in Fig. 2, A to C, to be just slightly detuned from the CD spectral maximum at 625 nm, the CD spectrum to

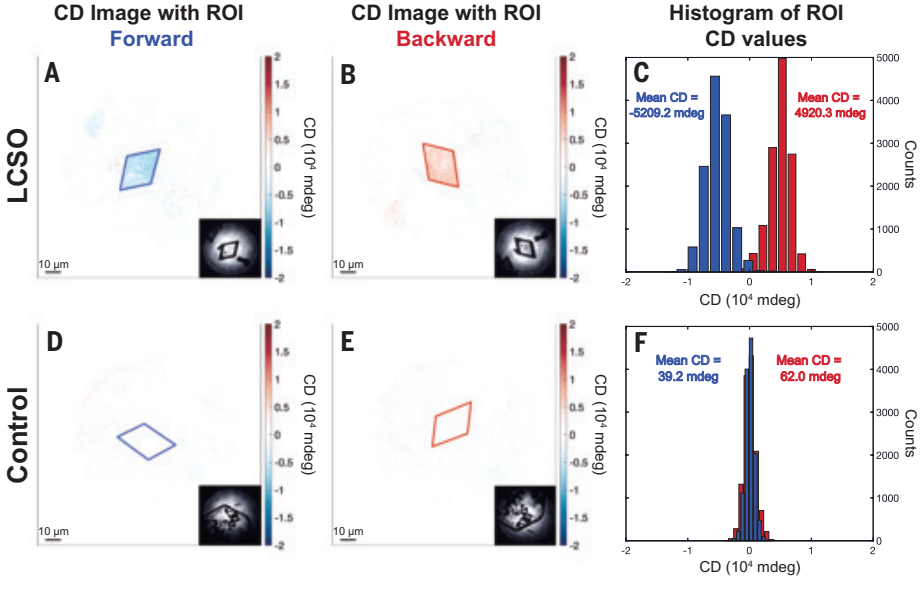


Fig. 2. Circular dichroism microscopy for LCSO and control crystals. (Top) The 12-μm-thick LCSO crystal displays strong CD signals which invert upon sample flipping. (Bottom) The 15-μm-thick control crystal does not display CD. CD images in the forward (**A** and **D**) and backward (**B** and **E**) configurations are generated on the basis of the left-handed (insets) and right-handed CP light transmission images and analyzed across the regions of interest (ROIs), shown as solid colored lines in (A), (B), (D), and (E). The mean CD is calculated for each ROI, and all the pixel values in the ROI are reported in histograms (**C** and **F**).

broadly invert upon sample flipping, and the dissymmetry factor to reach values of ~0.3. These spectral features are the result of a dense manifold of underlying electronic states, with dominant contributions ascribed to *d-d* transitions at the acentric Co2 site (*43*). We have also collected cross-polarized microscopy data confirming the presence of substantial birefringence (fig. S15).

To ensure that the observed chiroptical response for the centrosymmetric LCSO crystal is not an instrumental artifact, we synthesized a control material $Li_6Co(SeO_3)_4$ (*45*), which crystallizes in the centrosymmetric rhombohedral space group $R\bar{3}$ belonging to the crystal class $\bar{3}$, for which neither OA nor LD-LB is symmetry-allowed. Otherwise, this control material behaves similarly to LCSO, including optical absorption in the same wavelength range (fig. S14). CD images of the control material presented in Fig. 2, D to F, show no chiroptical response exceeding the noise level and a concomitant absence of signal inversion upon sample flipping, which is further substantiated by a spectrally resolved analysis (fig. S16). This control experiment, combined with our crystal class characterization of both crystals and detailed spectral calculations (fig. S3), strongly supports the attribution of the LCSO chiroptical response to LD-LB.

Crystal thickness dependence

LD-LB is a form of conversion CD (*46*) in which a sample biases amplitude transfer from one circular polarization to the other, and a detailed theoretical description is necessary to understand the many underlying polarization dependencies of this phenomenon (*47*). In the LB interaction, the two orthogonal $\lambda/4$-out-of-phase linearly polarized components, which comprise CP light, experience a differential phase shift, whereas in the LD interaction there is a differential absorption. This picture suggests rich and diagnostic behavior when crystal thickness is controlled. Our unparalleled access to single-crystal domains exhibiting LD-LB without OA across a range of thicknesses has allowed us to probe this behavior and derive mechanistic insight.

The strong CD response observed for the LCSO crystal arises from the leading contribution to LD-LB being of second order in the context of light-matter interactions. This contribution scales quadratically with crystal thickness. However, higher-order contributions become substantial in thick crystals (supplementary text S3), as has recently been emphasized by Robinson and co-workers (*48*). These contributions represent an accumulation of coherent polarization effects throughout the transmission process, yielding a sinusoidal dependence of the LB magnitude on crystal thickness, whereas the LD magnitude converges to a constant value [as summarized in Table 7 in (*49*)]. These effects combined produce an LD-LB signal manifesting a damped oscillation with respect to the crystal thickness, as shown in Fig. 3.

To experimentally explore this behavior for the LCSO crystals, we determined the thickness of individual samples through optical profilometry (materials and methods) and plotted CD as a function of the obtained thicknesses (Fig. 3A). Remarkably, almost two full cycles of the expected damped oscillation were observed across the range of crystal thicknesses. This behavior is qualitatively different from the expected thickness dependence of OA, for which CD scales linearly with path length in the short path length regime (*49*). The observed oscillation was well reproduced by an empirical fit (Fig. 3A), incorporating the above-described higher-order corrections to the LD-LB mechanism (supplementary text S3). Accompanying scattering-matrix method (SMM) (*50*) simulations parametrized from TD-DFT calculations of LCSO also displayed this oscillation (Fig. 3B). Within the SMM, Maxwell's equations are solved along the axis of optical propagation, with the regions on both sides of the crystal being treated as half-infinite media and with continuous boundary conditions for phase and amplitude of the electric field applied to the perpendicular axes. We note the additional presence of shorter, weaker oscillations for smaller crystal thicknesses due to internal reflections, a phenomenon omitted in the empirical fit based on Mueller calculus of continued

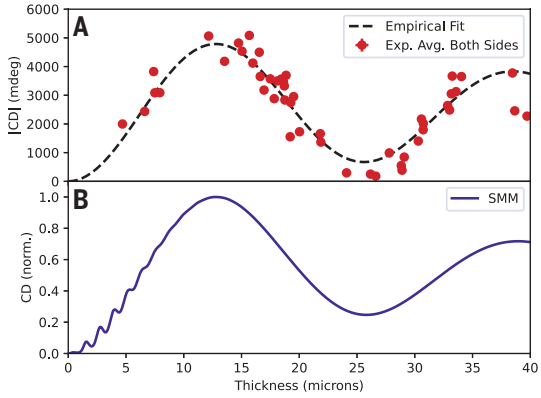


Fig. 3. Experimental and theoretical dependence of LCSO CD signal on crystal thickness probed near 635 nm. (**A**) The experimental CD magnitude is plotted against the crystal thickness as measured by optical profilometry. The absolute values of the experimental CD signals for each crystal in forward and backward transmission configurations are averaged to determine the magnitude irrespective of orientation relative to the direction of light propagation. The CD magnitude is plotted against the crystal thickness as measured by optical profilometry. Empirical fit ($R^2 = 0.76$) (dashed line) to experimental data (red points) is obtained using a thickness-dependent expression of LD-LB (supplementary text S3). (**B**) Plot obtained by solving Maxwell's equations in one dimension through the scattering matrix method (SMM) (*50*). For the dielectric tensor used in the SMM calculations, a dielectric constant of $\epsilon_\infty = 10$ was added to the results from excited-state calculations to mimic the effect of high-frequency contributions.

propagation but necessarily included through Maxwell's equations in the SMM. The magnitude of the SMM simulated signal intensity is within a factor of four of the experiment (fig. S4), which is well within the expected accuracy for TD-DFT in calculating electronic transition energies and transition dipole strengths, both of which contribute sensitively to the LD-LB response (*11*). Taken together, our experiments and simulations conclusively support the notion that the LD-LB mechanism, as distinct from OA, is operational in the LCSO crystals.

Discussion and conclusions

In addition to demonstrating differential absorption of CP light by LD-LB in a centrosymmetric crystal, this work introduces a crystal class characterization of this photophysical phenomenon. In this regard, it is interesting to evaluate the crystal classes and relative properties of earlier materials for which LD-LB has been observed. Reported metal halide perovskites fall in the space groups *C*2 (*20*), *Cc* (*21*), and *P*1 (*22*), which belong to the 2, *m*, and 1 crystal classes, respectively. These crystal classes support both LD-LB and OA, and both phenomena were indeed considered in these previous studies. Among the reported organic thin films, naproxen (*51*) crystallizes in the 2 crystal class in the overlapping part of the Venn diagram (Fig. 1A), whereas the oligothiophene thin films consist of ordered domains distributed in random orientations, for which the crystal structure has not been determined (*10*, *19*). However, the suspected presence of both LD-LB and OA (*10*, *19*) suggests that these structures also fall within the 1, 2, and *m* crystal classes.

In contrast to these earlier LD-LB materials, the properties of LCSO enable chiroptical engineering design based entirely on the LD-LB selection rules. It would be of interest to explore this feature in other single crystals within the $\bar{1}$ or $2/m$ centrosymmetric crystal classes, bearing in mind other material requisites for LD-LB such as the presence of (at least) two oblique and nondegenerate optical transitions within close spectral proximity (*11*). Notably, LCSO may serve as a starting point in screening efforts for such crystals. For example, substitutions to LCSO on the cation (La, Co, and Se) and anion (O) sites may serve as

an engineering strategy for shifting band edges from the visible to higher (ultraviolet and deep ultraviolet) or lower (infrared) spectral regions.

Our work also marks the realization of LD-LB interactions within a single crystal, which offers further advantages in terms of uniformity, synthetic control, characterization, and signal strength compared with previously reported LD-LB materials. For example, the aforementioned organic thin films with randomly oriented domains produce spatially dependent chiroptical signals that vary in strength and sign (*26*, *52*, *53*), whereas LD-LB–active perovskites feature variation of chiroptical signals due to irregularities in the film morphology (*54*). The LCSO crystals, by contrast, present well-defined, uniform, monosignate domains, making them highly appealing for engineering applications that require uniform chiroptical behavior. Notably, the nearly macroscopic LCSO crystals can be manipulated through polydimethylsiloxane stamps (supplementary text S7), and the strength of the LCSO chiroptical response can be tuned by deploying crystals of different thicknesses. The crystals are also phase pure, easily synthesized in a one-pot synthesis, and air-stable and resistant to photodamage (fig. S11)—a major issue in most organic or hybrid materials.

As mentioned above, the LD-LB response operative in this material inverts completely upon sample flipping or, by the same token, an inversion of the optical propagation direction, owing to the absence of OA. Recent theoretical (*31*) and experimental (*26*) work has shown that such inversion can enable enhancement of CP light asymmetries in conventional Fabry-Pérot microcavities, thus producing chiral resonances. The amplification of chiroptical asymmetry is also promising for use in polarization-dependent optical diodes (*55*, *56*), in the directional manipulation of light (*57*), as alternatives to cholesteric liquid crystals (*58*) or chiral dyes for chiral lasing applications, and for enhancing quantum light-matter transduction by means of chiral polaritons, which combine strong coupling with a high degree of chiral asymmetry (*30*–*32*, *59*).

In recent years, refinements of our understanding of chiral structure–function relationships have garnered attention thanks to the emergence of chiral phenomena such as chiral-induced spin selectivity (*60*–*62*), as well as the realization of distinct chiral functionalities (*32*–*41*). Our findings open the library of centrosymmetric solids to chiroptical engineering through the LD-LB mechanism, refining a long-held textbook principle on the interrelationship between crystal symmetry and differential absorption of CP light. As such, our findings help deepen our understanding of solid-state chiral phenomena, which may ultimately unlock technological opportunities.

REFERENCES AND NOTES

1. J. Jerphagnon, D. S. Chemla, *J. Chem. Phys.* **65**, 1522–1529 (1976).
2. L. E. Landau, L. Davidovich, L. Pitaevskii, *Electrodynamics of Continuous Media*, vol. 8 of *Course of Theoretical Physics* (Pergamon Press, ed. 2, 1984).
3. A. M. Glazer, K. Stadnicka, *J. Appl. Cryst.* **19**, 108–122 (1986).
4. W. Kaminsky, *Rep. Prog. Phys.* **63**, 1575–1640 (2000).
5. L. D. Barron, *Molecular Light Scattering and Optical Activity* (Cambridge Univ. Press, ed. 2, 2004).
6. Y. Tokura, N. Nagaosa, *Nat. Commun.* **9**, 3740 (2018).
7. P. S. Halasyamani, K. R. Poeppelmeier, *Chem. Mater.* **10**, 2753–2769 (1998).
8. H. D. Flack, *Helv. Chim. Acta* **86**, 905–921 (2003).
9. K. Claborn, C. Isborn, W. Kaminsky, B. Kahr, *Angew. Chem. Int. Ed.* **47**, 5706–5717 (2008).
10. G. Albano, M. Lissia, G. Pescitelli, L. A. Aronica, L. Di Bari, *Mater. Chem. Front.* **1**, 2047–2056 (2017).
11. A. Salij, R. H. Goldsmith, R. Tempelaar, *J. Am. Chem. Soc.* **143**, 21519–21531 (2021).
12. R. L. Disch, D. I. Sverdlik, *Anal. Chem.* **41**, 82–86 (1969).
13. Y. Shindo, Y. Ohmi, *J. Am. Chem. Soc.* **107**, 91–97 (1985).
14. A. M. Ritcey, D. G. Gray, *Biopolymers* **27**, 1363–1374 (1988).
15. Y. Shindo, M. Nishio, *Biopolymers* **30**, 25–31 (1990).
16. W. Kaminsky, K. Claborn, B. Kahr, *Chem. Soc. Rev.* **33**, 514–525 (2004).
17. J. H. Freudenthal, E. Hollis, B. Kahr, *Chirality* **21**, E20–E27 (2009).
18. K. Claborn *et al.*, *Cryst. Growth Des.* **5**, 2117–2123 (2005).
19. G. Albano, G. Pescitelli, L. Di Bari, *Chem. Rev.* **120**, 10145–10243 (2020).
20. Z. Zhang *et al.*, *J. Am. Chem. Soc.* **144**, 22242–22250 (2022).
21. J. Möbs *et al.*, *J. Am. Chem. Soc.* **145**, 23478–23487 (2023).
22. L. Yan *et al.*, *Adv. Opt. Mater.* **12**, 2302766 (2024).
23. H. Hu *et al.*, *ACS Nano* **15**, 13653–13661 (2021).
24. Y. Yao *et al.*, *ACS Nano* **16**, 20457–20469 (2022).
25. J. Lu *et al.*, *Nature* **630**, 860–865 (2024).
26. T.-L. Chen *et al.*, *Nat. Commun.* **15**, 3072 (2024).
27. Z. Wang *et al.*, *Adv. Mater.* **35**, e2303203 (2023).
28. R. Kuroda, T. Harada, Y. Shindo, *Rev. Sci. Instrum.* **72**, 3802–3810 (2001).
29. S. Sun, B. Gu, S. Mukamel, *Chem. Sci.* **13**, 1037–1048 (2022).
30. D. G. Baranov, C. Schäfer, M. V. Gorkunov, *ACS Photonics* **10**, 2440–2455 (2023).
31. A. H. Salij, R. H. Goldsmith, R. Tempelaar, *Nat. Commun.* **15**, 340 (2024).
32. H. Hübener *et al.*, *Nat. Mater.* **20**, 438–442 (2021).
33. E. Plum, N. I. Zheludev, *Appl. Phys. Lett.* **106**, 221901 (2015).
34. L. Kang *et al.*, *Nano Lett.* **17**, 7102–7109 (2017).
35. W. Liu, V. M. Menon, S. Gao, G. S. Agarwal, *Phys. Rev. B* **100**, 245428 (2019).
36. M. L. Solomon, J. Hu, M. Lawrence, A. García-Etxarri, J. A. Dionne, *ACS Photonics* **6**, 43–49 (2019).
37. X. Zhang, Y. Liu, J. Han, Y. Kivshar, Q. Song, *Science* **377**, 1215–1218 (2022).
38. Y. Zhao *et al.*, *Nat. Nanotechnol.* **12**, 1055–1059 (2017).
39. M. L. Solomon *et al.*, *Acc. Chem. Res.* **53**, 588–598 (2020).
40. J. Feis *et al.*, *Phys. Rev. Lett.* **124**, 033201 (2020).
41. K. Voronin, A. S. Taradin, M. V. Gorkunov, D. G. Baranov, *ACS Photonics* **9**, 2652–2659 (2022).
42. M. J. B. Tunis-Schneider, M. F. Maestre, *J. Mol. Biol.* **52**, 521–541 (1970).
43. M. Wildner, M. Andrut, *Z. Kristallogr. Cryst. Mater.* **214**, 216–222 (1999).
44. A. Rodger, B. Nordén, *Circular Dichroism and Linear Dichroism* (Oxford Univ. Press, 1997).
45. H. Jo, S. Lee, K. Y. Choi, K. M. Ok, *Inorg. Chem.* **57**, 3465–3473 (2018).
46. E. Plum, V. A. Fedotov, N. I. Zheludev, *J. Opt. A, Pure Appl. Opt.* **11**, 074009 (2009).
47. C. S. Brown, A. E. Bak, in *Proceedings Volume 3754, Polarization: Measurement, Analysis, and Remote Sensing II*, D. H. Goldstein, D. B. Chenault, Eds., SPIE's International Symposium on Optical Science, Engineering, and Instrumentation, 18–23 July 1999, Denver, CO (International Society for Optics and Photonics, 1999), pp. 65–74.
48. T. J. Ugras, Y. Yao, R. D. Robinson, *Chirality* **35**, 846–855 (2023).
49. H. P. Jensen, J. A. Schellman, T. Troxell, *Appl. Spectrosc.* **32**, 192–200 (1978).
50. D. Y. K. Ko, J. R. Sambles, *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **5**, 1863–1866 (1988).
51. G. A. Turner, Y. Hwang, J. Rong, C. Strachan, G. J. Simpson, *J. Phys. Chem. B* **127**, 8216–8225 (2023).
52. G. Albano *et al.*, *Aggregate* **3**, e193 (2022).
53. A. Taddeucci, F. Zinna, G. Siligardi, L. Di Bari, *Chem. Biomed. Imaging* **1**, 471–478 (2023).
54. U. Makhija, P. K. Rajput, P. Parthiban, A. Nag, *J. Chem. Phys.* **160**, 021102 (2024).
55. A. H. Gevorgyan, M. Z. Harutyunyan, *Phys. Rev. E* **76**, 031701 (2007).
56. I. Söllner *et al.*, *Nat. Nanotechnol.* **10**, 775–778 (2015).
57. M. Neugebauer, T. Bauer, P. Banzer, G. Leuchs, *Nano Lett.* **14**, 2546–2551 (2014).
58. D. Qu, M. Archimi, A. Camposeo, D. Pisignano, E. Zussman, *ACS Nano* **15**, 8753–8760 (2021).
59. A. Salij, R. Tempelaar, *Phys. Rev. B* **103**, 035431 (2021).
60. K. Ray, S. P. Ananthavel, D. H. Waldeck, R. Naaman, *Science* **283**, 814–816 (1999).
61. R. Naaman, Y. Paltiel, D. H. Waldeck, *Nat. Rev. Chem.* **3**, 250–260 (2019).
62. H. J. Eckvahl *et al.*, *Science* **382**, 197–201 (2023).
63. A. Salij, andrewsalij/LCSO_Repository: Release 1.0, Version v1.0.0, Zenodo (2025); https://doi.org/10.5281/zenodo.15164401.

ACKNOWLEDGMENTS

The authors thank C. Stern for helpful discussions. **Funding:** This research was primarily supported as part of the Center for Molecular Quantum Transduction, an Energy Frontier Research Center funded by the US Department of Energy (DOE), Office of Science, Basic Energy Sciences (BES), under award DE-SC0021314. The authors gratefully acknowledge use of facilities and instrumentation at the UW-Madison Wisconsin Centers for Nanoscale Technology (wcnt.wisc.edu), partially supported by the NSF through the University of Wisconsin Materials Research Science and Engineering Center (DMR-2309000). The authors would also like to thank the IMSERC X-Ray Facility at Northwestern University, which has received support from the Soft and Hybrid Nanotechnology Experimental (SHyNE) Resource (NSF ECCS-1542205), the Northwestern University Materials Research Science and Engineering Center (NSF DMR-2308691), and Northwestern University, and funding from the NSF (DMR-1904701). M.I.B.U. gratefully acknowledges support from the IIN Postdoctoral Fellowship and the Northwestern University International Institute for Nanotechnology. Research reported in this publication was supported in part by the International Institute for Nanotechnology at Northwestern University. **Author contributions:** Conceptualization: K.R.P., R.H.G., and R.T.; Funding acquisition: M.C.H., K.R.P., R.H.G., and R.T.; Investigation: K.A.P., A.S., K.R.K., E.S., M.I.B.U., A.J.B., R.C., M.A.S., M.C.H., K.R.P., R.H.G., and R.T.; Methodology: K.A.P., A.S., K.R.K., E.S., M.I.B.U., M.C.H., K.R.P., R.H.G., and R.T.; Project administration: K.R.P., R.H.G., and R.T.; Supervision: M.C.H., K.R.P., R.H.G., and R.T.; Visualization: K.A.P., A.S., K.R.K., E.S., M.I.B.U., R.H.G., and R.T.; Writing – original draft: K.A.P., A.S., K.R.K., M.I.B.U., K.R.P., R.H.G, R.T.; Writing – review & editing: A.S., K.R.K., E.S., R.H.G., and R.T. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text, the supplementary materials, and Zenodo (*63*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adr5478
Materials and Methods; Supplementary Text; Figs. S1 to S16; Tables S1 to S5; References (*64*–*91*)

Submitted 7 July 2024; resubmitted 4 January 2025; accepted 7 April 2025

10.1126/science.adr5478

TISSUE ENGINEERING

Rapid model-guided design of organ-scale synthetic vasculature for biomanufacturing

Zachary A. Sexton[1,2]†, Dominic Rütsche[1,2]†, Jessica E. Herrmann[3], Andrew R. Hudson[4], Soham Sinha[1], Jianyi Du[1], Daniel J. Shiwarski[4], Anastasiia Masaltseva[1], Fredrik Samdal Solberg[5], Jonathan Pham[5], Jason M. Szafron[2,6], Sean M. Wu[2,7], Adam W. Feinberg[4,8], Mark A. Skylar-Scott[1,2,9,10]*‡, Alison L. Marsden[1,2,6,11]*‡

Our ability to produce human-scale biomanufactured organs is limited by inadequate vascularization and perfusion. For arbitrarily complex geometries, designing and printing vasculature capable of adequate perfusion poses a major hurdle. We introduce a model-driven design platform that demonstrates rapid synthetic vascular model generation alongside multifidelity computational fluid dynamics simulations and three-dimensional bioprinting. Key algorithmic advances accelerate vascular generation 230-fold and enable application to arbitrarily complex shapes. We demonstrate that organ-scale vascular network models can be generated and used to computationally vascularize >200 engineered and anatomic models. Synthetic vascular perfusion improves cell viability in fabricated living-tissue constructs. This platform enables the rapid, scalable vascular model generation and fluid physics analysis for biomanufactured tissues that are necessary for future scale-up and production.

The fabrication of thick, functional tissues and organs depends on the successful generation and integration of perfusable vascular networks (*1*, *2*). Natively, cells of most soft, vascularized tissues lie within 100 μm of a blood vessel; in metabolically demanding tissues, this distance can be less than 20 μm, and vessel density may be greater than 2500 vessels per mm^3 (*3*–*5*). Hierarchical structure allows vascular networks to reach the extensive scales necessary for organ perfusion without sacrificing transport efficiency (*6*, *7*). Thus, the scalability of manufactured tissues depends on recapitulating the tree-like vascular organization observed in vivo.

Recent advances in tissue manufacturing allow fabrication of multimaterial constructs replete with multiple cell types and extracellular matrix compositions at resolutions and speeds necessary for viable three-dimensional (3D) tissues (*8*–*11*). However, a robust method for generating a multiscale and biomimetic perfusable network within these tissues remains elusive. Though it is possible to segment vasculature directly from image data, these data often do not exist at adequate resolution for arbitrary shapes used in bioprinting. To date, only cryomicrotomes and light sheet microscopy techniques have allowed direct segmentation of the microvasculature (*12*, *13*). Furthermore, engineered tissues rarely have direct anatomical analogs from which to extract imaged vascular networks. Modern-day perfusion designs have thus relied on simple lattices, which poorly reproduce the topology and hemodynamics of native vasculature (*14*–*17*). Although acceptable for tissue manufacturing at low cell densities ($<10^6$ cells/ml), these idealized designs fail to satisfy the metabolic demand of clinically relevant cell densities (10^8 cells/ml), thereby precluding the in vitro viability of larger-scale manufactured tissues (*18*).

The cardiovascular computational fluid dynamics (CFD) community has developed methods to generate microvascular tree networks for incorporation into hemodynamic models. Structured trees of fractal topology and morphometric trees have been used to investigate microvascular autoregulation, quantify wall shear stresses, and provide boundary conditions for 3D simulations (*19*, *20*). Although these provide basic hemodynamic information, typically using a Poiseuille flow assumption, they cannot provide the 3D spatial information needed for biofabrication of vascular tree networks. Algorithms to construct vascular tree networks in 3D volumes have therefore been proposed and include global constrained optimization (GCO), Monte Carlo, Lindenmayer systems, fractal trees, growth-based optimization (GBO), and constrained constructive optimization (CCO), among which GBO, CCO, and GCO exploit hemodynamic principles during vascular generation to ensure sufficient perfusion (*21*–*24*). However, prior CCO methods had several notable drawbacks, including long build times (hours to days for complex models), limited handling of nonconvex volumes, and an inability to predict resulting hemodynamics at multiple spatial fidelities. Although some recent work has extended both CCO and GCO for nonconvex perfusion territories, the methods are computationally expensive and CCO remains the only algorithm in this family that is able to produce closed fluid circuits, a requirement for engineered tissues (*25*–*28*). Despite these limitations, CCO has been widely adopted for iterative vascular generation, with applications spanning both patient-specific CFD and perfusion of 3D printed tissues (*29*).

We introduce a synthetic vascular toolkit that generates organ-spanning tree-network models within minutes and provides a complete automated pipeline from model generation to hemodynamic simulation and 3D fabrication. We combined CCO methods with recent techniques for CFD hemodynamic simulation and 3D printing to provide matched flow analysis and vascular geometry in support of tissue fabrication needs. We extended these methods for complex perfusion territories, multifidelity hemodynamic simulation, and closed-loop perfusion networks. We demonstrated our workflow by perfusion of 3D bioprinted vascular networks in a bioreactor and corroborated flow behavior using CFD models. We further characterized and simulated networks within larger engineered designs and organ-scale anatomies to motivate large-scale vascular planning of engineered tissues.

Overview of perfusion topology performance

We simulated and examined transport network metrics such as energetic losses due to resistance, flow distribution homogeneity, intravascular mass transport, and approximate tissue perfusion for network architectures generated within convex and nonconvex geometries (Fig. 1, A to G). This analysis considered thousands of network designs for lattice and randomly seeded synthetic vascular strategies, for which some representative networks are shown (Fig. 1, A and B). In networks constrained by a fixed total blood volume, lattice designs exhibited higher nominal net hydraulic resistance than synthetic tree networks (Fig. 1C), implying that larger pressure gradients are required to achieve equivalent flow rates. Furthermore, lattice networks with the same total transport volume and inlet flow rates showed greater variability in flow distributions. The lowest quartile of flow rates in simulated lattice networks was typically two to three orders of magnitude lower than the worst performing vessels in synthetic architectures (Fig. 1D).

[1]Department of Bioengineering, Stanford University, Stanford, CA, USA. [2]Cardiovascular Institute, Stanford University, Stanford, CA, USA. [3]School of Medicine, Stanford University, Stanford, CA, USA. [4]Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA. [5]Department of Mechanical Engineering, Stanford University, Stanford, CA, USA. [6]Department of Pediatrics, Stanford University, Stanford, CA, USA. [7]Division of Cardiovascular Medicine, Department of Medicine, Stanford University, Stanford, CA, USA. [8]Department of Materials Science and Engineering, Carnegie Mellon University, Pittsburgh, PA, USA. [9]Basic Science and Engineering Initiative, Children's Heart Center, Stanford University, Stanford, CA, USA. [10]Chan Zuckerberg Biohub, San Francisco, CA, USA. [11]Institute of Computational and Mathematical Engineering, Stanford University, Stanford, CA, USA. ***Corresponding author. Email: skyscott@stanford.edu (M.A.S.-S.); amarsden@stanford.edu (A.L.M.)** †These authors contributed equally to this work. ‡These authors contributed equally to this work.

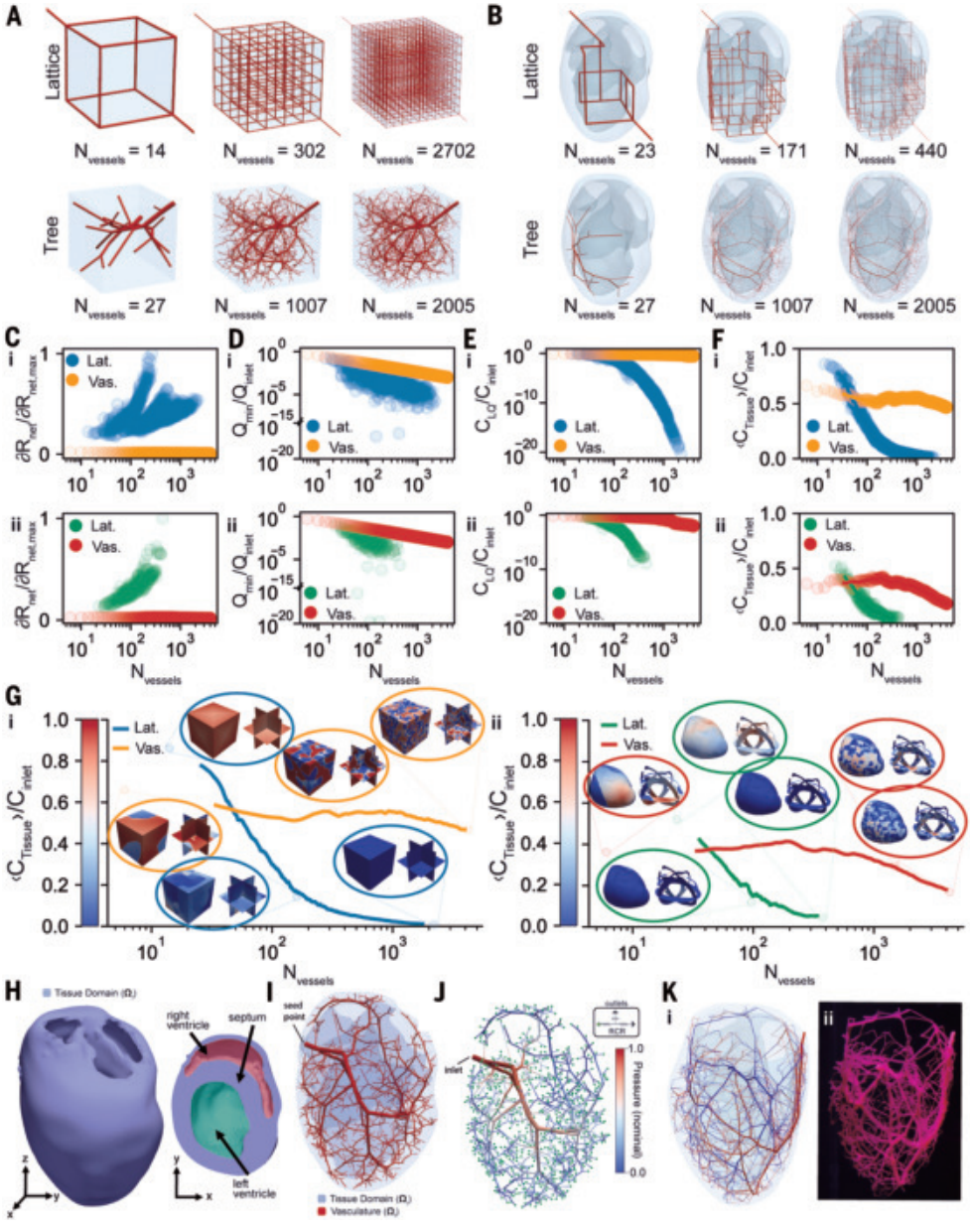


Fig. 1. Perfusion performance and capabilities of the synthetic vascular toolkit. (**A** and **B**) Representative lattice and synthetic vascular networks at a fixed total blood volume in convex (cube) and nonconvex (biventricle) domains. (**C**) Net resistance per unit length (∂R_{net}) for networks in (i) convex and (ii) nonconvex geometries, normalized by the maximum value ($\partial R_{net,\,max}$) and plotted against the number of vessels ($N_{vessels}$). Lat., lattice; Vas., synthetic vasculature. (**D**) Minimum flow ranges [(Q_{min}/Q_{inlet}), 1.0] versus network size in (i) convex and (ii) nonconvex geometries. (**E**) Normalized lower quartiles of solute concentrations (C_{LQ}/C_{inlet}) within vessels for (i) convex and (ii) nonconvex network geometries. (**F**) Average bulk solute concentrations in tissue ($\langle C_{tissue} \rangle / C_{inlet}$) approximated by a linear depletion model, plotted for (i) convex and (ii) nonconvex tissues over increasing network sizes (solid lines indicate moving averages). (**G**) Color-mapped tissue concentrations for networks of three different sizes in (i) convex and (ii) nonconvex geometries, with average bulk concentrations and vessel counts indicated. (**H**) A biventricle tissue domain (Ω^t) featuring challenging thin-walled regions (ventricular septum, right ventricular wall). (**I**) Open-loop synthetic vasculature tree with 1000 outlets defining the fluid domain (Ω^f) for blood or media flow. The root (seed) location is indicated. (**J**) Matched CFD simulation generated from the 1000-outlet tree, with an inlet flow waveform and outlet impedances modeling capillary and venous resistance. (**K**) Connections between synthetic arterial and venous trees form a closed-loop network suitable for embedded 3D printing.

When we constrained all networks to an equivalent Péclet number at the inlet, we could evaluate intravascular solute distributions under quasi-steady flow using analytical solutions to advection-diffusion-permeation equations [supplementary materials; (*30*)]. These simulations revealed substantial differences in solute concentrations between the lowest-flow vessels in lattice networks and their synthetic vascular counterparts (Fig. 1E), suggesting that the regions of nutrient and oxygen depletion may develop in networks with prolonged residence times. Extending the analysis into the extravascular domain, solute depletion patterns provided insight into areas of tissue at risk of nutrient and oxygen deprivation. With increasing network scale, lattice architectures consistently exhibited larger regions of poor approximate perfusion than synthetic vascular analogs (Fig. 1F). Visualization of the approximated tissue perfusion is shown in Fig. 1G for selected networks. In nonconvex geometries, these disparities were further accentuated, underscoring the relatively poor performance of lattice networks across all evaluated transport metrics.

In application, our synthetic vascular algorithm accommodated arbitrarily complex tissue geometries, Ω^t, including cavities and thin walls (Fig. 1H). A highly nonconvex biventricle shape illustrates this capability, where septum and right ventricle walls pose challenges for previous methods. By using implicit geometry reconstructions, we efficiently embedded a vascular tree, Ω^f, throughout these regions (Fig. 1I). Because actual hemodynamic conditions often exceed the simplified Poiseuille assumptions used during vascular generation, multifidelity simulations can be automatically coupled for post hoc CFD analyses (Fig. 1J). As expected, normalized pressure decreased with bifurcation depth. In contrast to synthetic vasculature, lattice architectures showed greater flow heterogeneity and higher pressures (fig. S1). Finally, watertight vascular networks could be exported as common mesh file formats (e.g., .stl, .3mf) or tool path files (e.g., .gcode) for 3D biofabrication. Figure 1K shows a closed-loop arterial-venous system for a biventricle tissue volume subsequently committed to printing.

Performance of acceleration techniques during synthetic vascular generation

Beyond data-structure considerations (fig. S2), we introduce four key features for synthetic vascular tree generation: (i) partial binding optimization, (ii) partial implicit volumes, (iii) collision avoidance, and (iv) closed-loop vasculature. New vessels were added by bifurcating from existing trees, and each bifurcation was optimized using the cost function $C = w(x) \sum l_i^u r_i^\lambda, \forall i \in \mathcal{T}$, which combines vessel radii (r_i) and lengths (l_i) scaled by power-law constants λ and u, respectively, and weighted by a position-dependent penalty [$w(x)$].

During optimization, hydraulic changes propagate back to the tree root and then forward again, making data traversal the main computational bottleneck. To mitigate this, we introduced partial binding (closure) (Fig. 2A), which caches intermediate structures to avoid re-evaluating unchanged branches. Our cost function arises solely from fluid physics and allometric scaling laws, without incorporating local feedback mechanisms found in structural adaptation models; such models could provide more precise flow-matching local metabolic requirements than fluid mechanics alone (*31*).

Previously, bifurcation optimization required two full tree traversals per evaluation $\mathcal{O}(N^2)$ for trees of size N, causing multiday compute times beyond a few hundred vessels (Fig. 2B). To address this inefficiency, recent methods approximated the cost function to achieve constant-time optimization (*32*), which works well for deep bifurcations but diverges from the true optimum near the root (depth ≤20). By contrast, partial binding scales at worst as $\mathcal{O}[\log(N)]$, evaluates the exact cost function, and depends only on optimization tolerances. It achieves median alignment accuracy of ≥60% across all bifurcation depths and ≥95% for deep bifurcations, as measured against a brute-force finite-difference solution (Fig. 2C). Of the numerical optimizers we tested, a quasi-Newton [limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS-B)] approach offers an effective balance of speed and accuracy (fig. S3).

Most realistic tissue geometries are nonconvex and costly to parse. To enable rapid interrogation during vascular generation, we used a partial implicit volume approach that extends recent work in surface reconstruction methods (materials and methods and fig. S4, A and B). This approach defines an implicit function from sampled data as a set of quadratic optimization subproblems but avoids the $\mathcal{O}(D^3N^3)$ overhead and potential ill-conditioning for N data points of D dimensions (fig. S4, C and D). We tested this partial implicit volume scheme on 10 engineered geometries and >200 anatomic brain regions (*33*) and observed efficient, accurate reconstructions for vascular generation (fig. S5 and table S4). Leveraging this implicit geometry and previous acceleration techniques, we observed that generating 8000 terminal trees in a brain gyrus geometry takes a median of 15 min, compared with 59.96 hours reported elsewhere (*34*). We also successfully created 1000-outlet vascular trees in a cube, annulus, heart, and brain gyrus (Fig. 2E and movie S1). Even sharp edges were reasonably captured and can be further improved by adapting the underlying implicit functions (fig. S6). Crucially, these implicit volumes offer cheaper point-enclosure queries

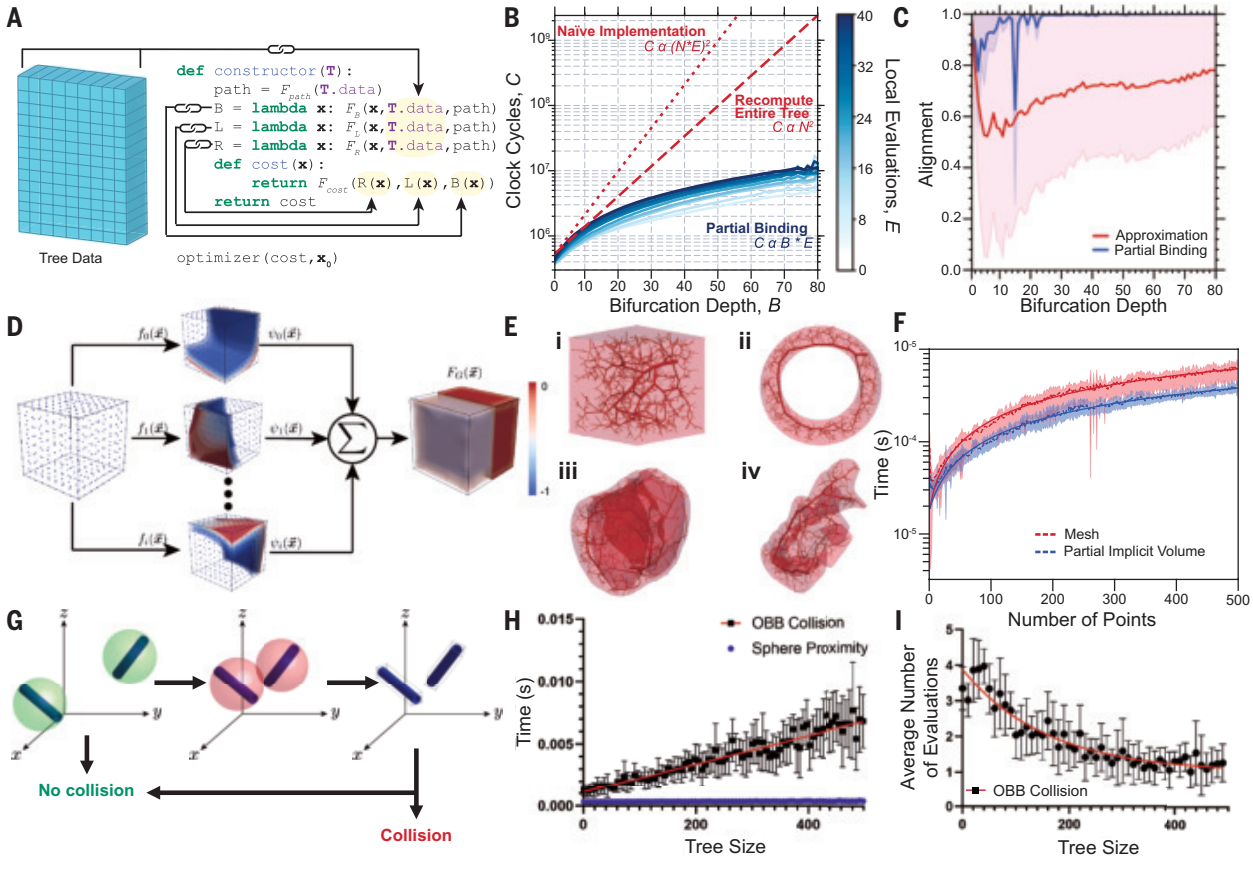


Fig. 2. Performance of synthetic vascular acceleration techniques. (**A**) Conceptual schematic of partial binding for bifurcation optimization. A "constructor" builds functions for bifurcation ratios (F_β), hydraulic length updates (F_{L*}), and hydraulic resistance updates (F_{R*}) based on locally bound tree configurations (dependent on bifurcation position x). The resulting cost function (F_{cost}) is returned for optimization. (**B**) Theoretical scaling complexities (red) for previously reported bifurcation optimization schemes versus partial-binding methods (blue). Plots reference bifurcation depth in full, balanced trees (shading: ±2 SD, $N = 100$). (**C**) Alignment accuracy versus a high-resolution finite-difference solution for partial binding (blue) and a prior approximation (red). Solid lines show median alignment (shading spans minimum-maximum alignment; $N = 100$). All trees are generated to 8000 outlets within a cube perfusion volume. (**D**) Schematic of partial implicit volumes. Surface point data are decomposed into local patches, then blended during implicit reconstruction. (**E**) Partial implicit volumes generate vascular trees (1000 outlets) across (i) a cube, (ii) an annulus, (iii) a biventricle heart, and (iv) a brain gyrus. (**F**) Scaling complexity for mesh-based versus implicit evaluations (shading = ±2 SD, $N = 25$). (**G**) Collision avoidance schematic. Sphere proximity detects potential collisions for precise OBB checks. (**H**) Sphere proximity alone is cheap but less accurate (error bars ±2 SD, $N = 100$). (**I**) Used as a filter, heuristic sphere proximity eliminates most OBB evaluations as tree complexity grows (error bars ±2 SD, $N = 100$).

than mesh-based methods, especially as mesh sizes grow (Fig. 2F and fig. S7).

Collision checks can also be a bottleneck. We accelerated these checks by first screening potential collisions with cheap, unorientable bounding surfaces [using kd-tree and Hierarchical Navigable Small World (HNSW) algorithms] and then applying more precise oriented bounding box (OBB) tests only as needed (Fig. 2G) (*35*). Even as trees grew, most collisions resolved with minimal OBB checks (Fig. 2, H and I), allowing efficient open-loop vascular generation across >200 engineering and anatomic domains (figs. S8 and S9 and table S5).

Finally, closed-loop vasculature was formed through multiobjective optimization (supplementary materials). Terminal segments were connected using cubic Bézier curves that balance vessel tortuosity, collisions, boundaries, and blood volume. Unlike many methods restricted to convex shapes, ours easily handles nonconvex shapes such as an annulus or a biventricle heart, making it suitable for anatomically realistic perfusion designs. Connections among multiple interpenetrating trees further enabled complex vascular mixing networks with minimal additional scripting. The present serial pairwise approach is constrained to maintain uniform flow through terminal vessels; however, this method can be expensive for large systems and may not produce the most efficient solute delivery compared with alternative methods (*36*).

Automatic multifidelity hemodynamics modeling

To facilitate simulations, we automated converting discrete vascular trees into watertight meshes (Fig. 3A) using open-source SimVascular pipelines. Standard steps—segmentation, 3D mesh generation, boundary-condition application, and finite-element hemodynamic simulation—were integrated into a largely automated workflow (*37*, *38*). For each synthetic tree, we performed contour checking, smoothing, meshing, boundary-condition assignment (supplementary materials), and blood flow simulation on high-performance computing clusters. The resulting finite-element simulations yielded high-fidelity pressure, velocity, and wall shear-stress distributions, which are demonstrated here under steady flow; time-varying conditions can also be imposed using idealized pulsatile waveforms (*39*).

Depending on the caliber and number of vessels, high-fidelity 3D finite-element simulations can be prohibitively expensive or unnecessary for large synthetic vascular trees, especially when many vessels operate in low Reynolds number (Re) or near-Stokes regimes (Re ≤50) well described by simplified, low-fidelity simulations (*40*). 1D approaches integrate the Navier-Stokes equations over the vessel cross section, whereas 0D models use circuit analogies to form inexpensive differential-algebraic equations. Both methods efficiently capture flow and pressure waveforms in low-Re vasculature. We automated the parameter setup for 0D and 1D simulations by extracting centerlines and cross sections from the original 3D models (Fig. 3B). Using the same cube, annulus, heart, and brain gyrus geometries (Fig. 2E), we computed bulk flow quantities such as average pressure and volumetric flow rate over a cardiac cycle (Fig. 3, C and D). Unlike the simplified assumptions used during vascular generation, these models incorporated vessel resistance and compliance to capture time-varying pressure and pulse propagation (movie S2). In many cases, wall shear stress is also well approximated under idealized pipe-flow assumptions, allowing for convenient analysis, given its central role in much of vascular remodeling.

3D biofabrication with on-demand synthetic vasculature designs

In a large-scale demonstration of our rapid, scalable approach, we computationally generated synthetic vasculature with 10^4 to 10^6 terminal vessels in nonconvex biventricular and annular geometries (Fig. 4, A and B). At 1 million outlets, vessel diameters ranged from

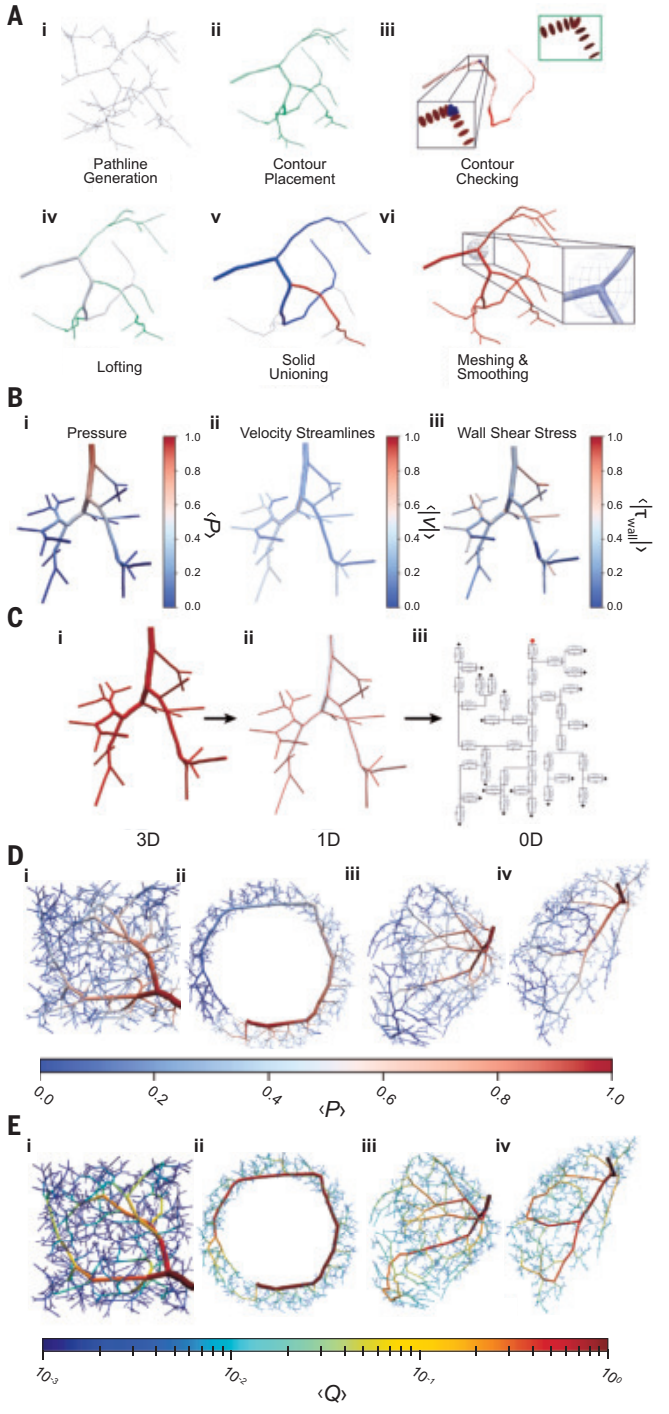


Fig. 3. Automatic multifidelity hemodynamics modeling. (**A**) Major steps for automatic 3D watertight model generation from discrete synthetic vessels. (**B**) Normalized pressure ⟨*P*⟩, velocity along streamlines ⟨|*v*|⟩, and wall shear-stress magnitude $\langle|\tau_{wall}|\rangle$ obtained from 3D finite-element method simulation with a steady inflow profile. (**C**) Illustration of reduced-order 1D and 0D models extracted from original 3D model files. Average normalized pressure ⟨*P*⟩. (**D**) and flow rates ⟨*Q*⟩ (**E**) obtained from 0D steady-flow simulations for vascular trees with 1000 terminals within (i) cube, (ii) annulus, (iii) biventricle, and (iv) brain gyrus tissue domains.

1.2 mm to 5 μm, matching morphometric data for newborn coronary vessels (*41*). Each million-vessel tree took about 5 hours to generate (table S6), and mean distances from tissue spaces to the nearest vessel measured ~180 to 190 μm (fig. S10). Incorporation of angiogenic methods may help to decrease this average distance through the inclusion

of a capillary network (*42*). Finally, to illustrate tissue-engineering utility, we printed 500-branch biventricular and annular trees into a hydrogel support bath using embedded 3D printing (Fig. 4, C and D, and movie S3) (*43*).

We next used freeform reversible embedding of suspended hydrogels (FRESH) to fabricate a four-vessel planar network, confirming watertightness and perfusion. Above 37°C, the support melted to evacuate vascular channels (Fig. 5A). Optical coherence tomography showed minimal (~100 μm) deviation from the digital design (Fig. 5B). After support removal, we used dye perfusion to verify lumen integrity (Fig. 5C and movie S4). Printing resolution was limited by a 150-μm needle diameter and 60-μm layer height (~400-μm minimal feature). A corresponding steady 3D CFD simulation [volumetric flow rate (Q_{inlet}) = 0.25 ml/min] characterized pressure and wall shear-stress distributions (Fig. 5D).

To validate the 3D CFD-predicted flow distribution, we fabricated a silicone elastomer model (Fig. 5E) using an acellular sacrificial writing process (supplementary materials). We confirmed perfusability using epifluorescence microscopy with aqueous fluorescein isothiocyanate–dextran (FITC-Dextran). We then performed particle image velocimetry (PIV) on a water-glycerol mixture (10-μm particles) delivered at 50 μl/min, capturing flow fields at four locations (fig. S11). The measured flow rate ratios ($N = 3$) between each branch and the inlet were nearly uniform among the three main branches, closely matching the CFD results (Fig. 5F). We attributed measurement discrepancies mainly to limited fabrication accuracy, particularly at branching points causing nonuniform head loss. Additionally, residual sacrificial fluid remaining after multiple flushes may disrupt flow owing to its higher viscosity and interfacial tension.

To illustrate the utility of our model-driven pipeline for tissue fabrication and flow simulation, we created a nonplanar, single-inlet, single-outlet vascular network (Ω^f; fluid domain) within a cubic tissue domain (Ω^t; tissue domain) (Fig. 5G). We performed fabrication on a scaled model by extruding a granular ink into a granular support bath, replicating the elastic modulus of soft tissue (supplementary materials). A key feature of this synthetic vasculature is uniform hydraulic resistance across all branches, ensuring even flow distribution and commensurate improvements in perfusion and viability. Before printing, we simulated pulsatile flow using digital twins of the proposed network to confirm hemodynamic integrity. We assumed fluid density and viscosity similar to whole blood ($\rho = 1.06$ g/cm^3 and $\nu = 0.04$ cP, respectively) and imposed a sinusoidal inlet flow with a mean inflow of (Q_{inlet} = 0.25 ml/min) and an amplitude of 0.1 ml/min. Flow and pressure waveforms at vessel midpoints (Fig. 5H, i and ii) closely matched theoretical profiles, with minimal deviations of 0.02 μl/min (flow) and 0.03 mmHg (pressure) (Fig. 5I). These simulations assumed a rigid container, where the soft matter is effectively incompressible. Under nonrigid conditions, flow distributions varied more because of fluid-structure interaction (supplementary materials and fig. S12). Such simulations informed our understanding of viscous power losses and pressure dissipation in synthetic vascular networks. The physical network was visualized with red dye in the coembedded granular gel (Fig. 5J).

Lastly, we leveraged nested embedded 3D bioprinting to fabricate a perfused annulus in two sequential steps (*44*, *45*). Microcomputed tomography (μCT) of the printed model revealed minimal surface deviations compared against the digital model (Fig. 5M). For cellular prints, an annulus with a radius of 10 mm and a height of 18 mm was first printed into a suspension bath using a cell-laden bioink containing human embryonic kidney (HEK) 293 cells at a density of 10 million cells/ml (tissue domain). Next, an integrated

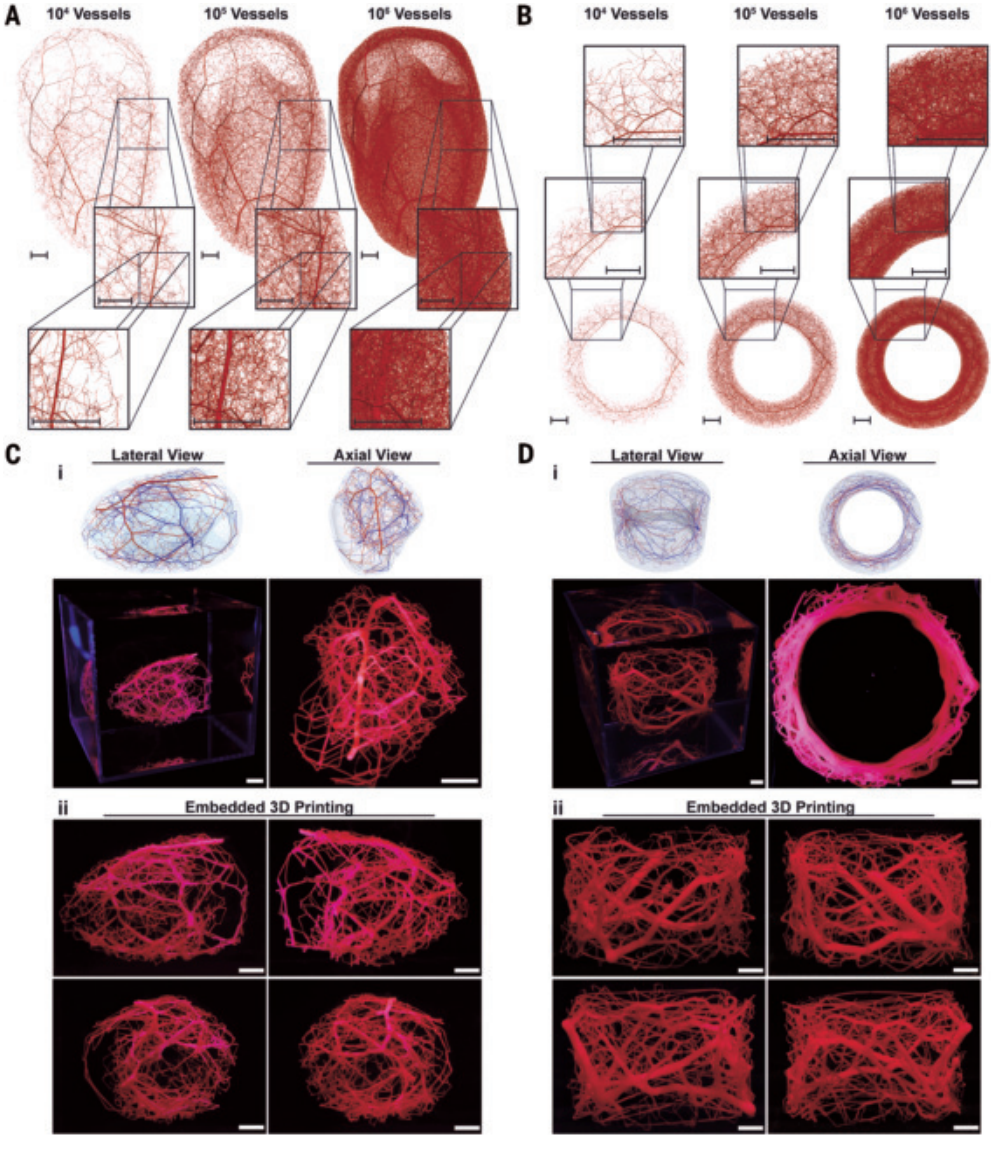


Fig. 4. Scalable vessel densities in algorithmic generation and bioprinting of vascular models. (**A** and **B**) Synthetic vascular trees scaled across three terminal vessel densities (10^4, 10^5, 10^6) terminal vessels shown for (A) biventricular and (B) annulus geometries (scale bars are 1 cm). (**C** and **D**) (i) Tissue (gray) and vascular domains (red and blue) depicted as lateral and axial views alongside corresponding models with 500 interconnected unevacuated branches, printed in Carbopol ink containing red pigment. (ii) Orthogonal views of printed (C) biventricular and (D) annulus geometries (scale bars are 1 cm).

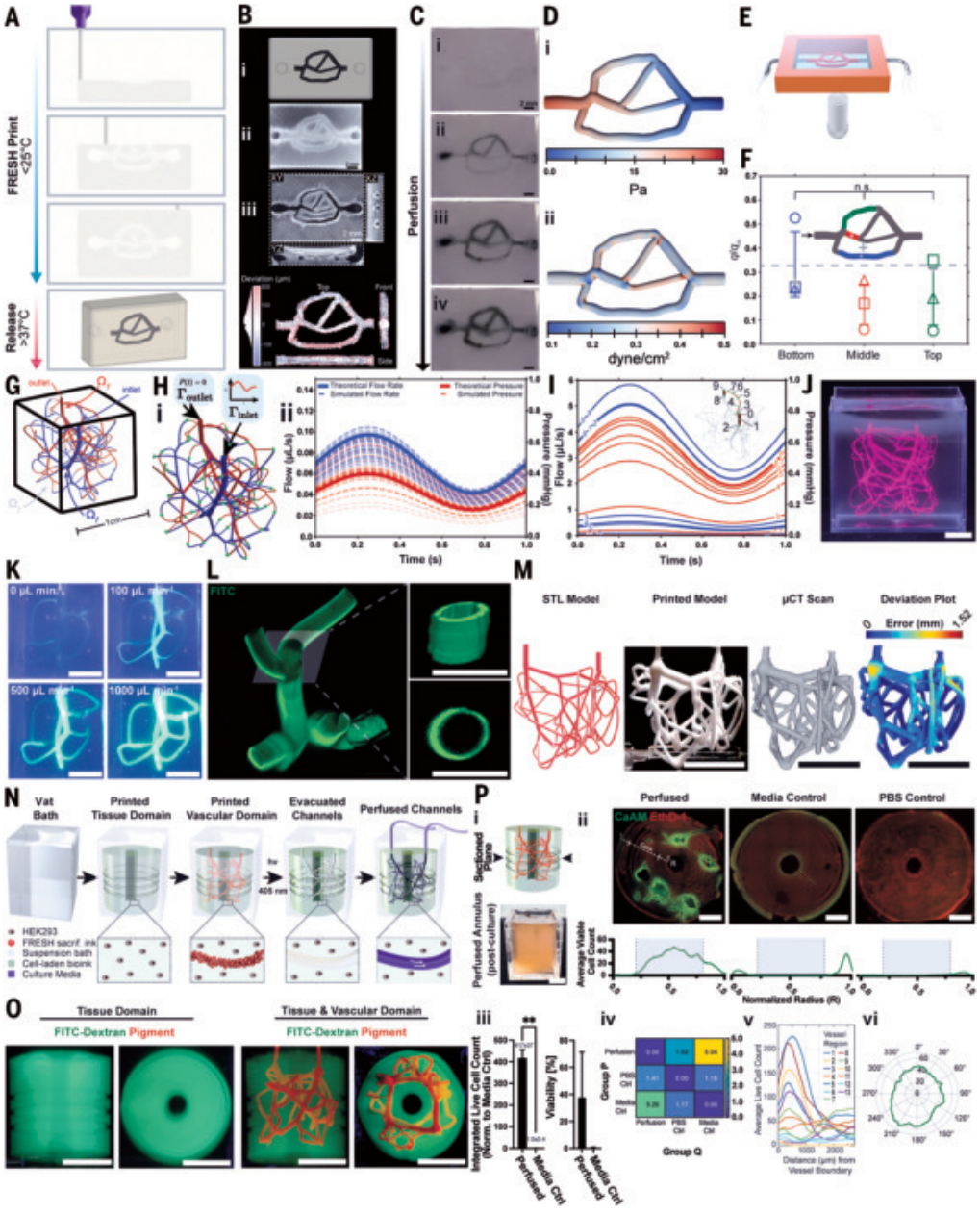


Fig. 5. Perfusion and viability of bioprinted vascular models. (**A**) Two-step FRESH printing and (**B**) OCT gauging of printed construct. (**C**) Dye perfusion demonstrating watertight lumen. (**D**) CFD simulation (Q_{inlet} = 0.25 ml/min) showing (i) pressure and (ii) wall shear stress. (**E**) Schematic of PIV validation setup. (**F**) Normalized flow rates (q/q_{in}) from PIV compared across three vessels using unpaired *t* tests with Welch's correction (N = 3). Averaged (dashed) and individual ("+") CFD simulation data are shown. Error bars indicate ± 1 SD. n.s., not significant. (**G**) 3D vascular model schematic (boundary conditions: inlet Γ_{inlet}, outlet Γ_{outlet}; domains: tissue Ω^t, fluid Ω^f). (**H**) (i) Pulsatile flow simulation boundary conditions (mean Q_{inlet} = 0.25, amplitude = 0.1 ml/min). (ii) Simulated (dashed) versus theoretical waveforms (solid) are shown. (**I**) Flow and pressure along cross sections of a vessel. The inset shows the selected vessel, with numbers indicating cross section slice indices. (**J**) Printed, unperfused network (scale bar is 1 cm). (**K**) FITC-Dextran perfusion at varying flow rates (scale bars are 1 cm). (**L**) Volumetric model and circularity after ink evacuation (scale bars are 1 mm). (**M**) µCT reconstruction and deviation plot (scale bars are 1 cm). (**N**) Nested bioprinting schematic of perfused annulus. (**O**) Printed annulus matrices: (left) annulus alone and (right) combined vascular and tissue (scale bars are 1 cm). (**P**) Vascularized annulus viability: (i) sectioning plane (top) and cultured annulus after 7 days (bottom; scale bar is 1 cm); (ii) LIVE/DEAD (CaAM/EthD-1) staining (top), with normalized radius (R) and core region ∈ [0.2 to 0.8] indicated for average radial cell viability (bottom) (N = 3 tissue constructs per condition, scale bars are 1 mm); (iii) comparison of integrated viable cell count [perfused: (417 ± 37)-fold higher] and viability (perfused: 37.8 ± 33.6%; media control: 0.4 ± 0.4%) within the core region normalized to media control using unpaired *t* tests with Welch's correction (N = 3; ** indicates $0.05 < p < 0.01$); (iv) Kullback-Leibler divergence comparing observed and approximating distributions; (v) spatial distribution of viable cells near channels; and (vi) polar histogram of viability near the perfused channel (N = 3).

perfusable vascular domain consisting of 25 interconnected vessel segments was printed into the annulus using a sacrificial bioink based on gelatin microparticles (vascular domain). After photo–cross-linking of the entire construct, channels were evacuated and perfused for 7 days (Fig. 5, N and O). The suspension bath allowed for photo–cross-linking but did not contain any cell culture media or nutrients [phosphate-buffered saline (PBS)]. Consequently, media was supplied exclusively through the vascular domain. As controls, annuli without internal vascular domains were printed and cultured for 7 days either in cell culture media or PBS. Transverse cross sections (Fig. 5P, i) revealed that perfusion maintained high viability in the annular core ("perfused"; Fig. 5P, ii), whereas non-perfused tissue in a media bath showed a steep viability gradient at the inner and outer interfaces. Submerging the avascular annulus in PBS yielded negligible cell viability. Integrated over the core area, perfused tissues had a 417-fold higher viable cell count than the media control. Similarly, viability in the core region was higher in the perfused sample (37.8%) than in the media control (0.4%) (Fig. 5P, iii). Kullback-Leibler divergence highlighted the large relative entropy between the perfusion and media-control distributions (Fig. 5P, iv). Further, most live cells were found within ~1000 µm of a perfused channel boundary (Fig. 5P, v, and fig. S15), with a largely uniform radial distribution around vessels (Fig. 5P, vi).

Discussion and conclusions

The future of engineered organ-scale tissues depends on a robust software infrastructure capable of designing and validating vascular networks through physics-based, functional targets. Generating custom vasculature across multiple scales has posed a substantial barrier to progress in the bioprinting community. In this work, we present a comprehensive model-driven pipeline for efficient vascular generation in arbitrary geometries capable of creating watertight models suitable for biofabrication and digital twins for advanced multifidelity hemodynamic simulations.

Although we further present a method for biofabricating perfusable vascular trees, large and complex trees with smaller-diameter vessels remain slow to build, and print inaccuracies can lead to discontinuities or blockages that inhibit perfusion.

Long-term, engineered tissues aspire to replace damaged or diseased organs in patient-specific interventions. Verifying such biomanufactured constructs will require simulation-optimized designs before resource-intensive fabrication. Similar to surgical planning, numerical models will become increasingly necessary to generate rational designs and predict the functional efficacy of engineered tissue before clinical translation (*46*). Preliminary applications of these synthetic vascular methods have integrated clinical imaging into personalized myocardial blood flow simulations (*47*). With the presented model-driven pipeline, we created complex synthetic vascular models on commonplace compute hardware within minutes. Such efforts expand the potential for more extensive hemodynamic analysis and design optimization in future synthetic vascular structures for wide-ranging applications in biofabrication and patient-specific modeling.

REFERENCES AND NOTES

1. J. Rouwkema, N. C. Rivron, C. A. van Blitterswijk, *Trends Biotechnol.* **26**, 434–441 (2008).
2. J. D. Kakisis, C. D. Liapis, C. Breuer, B. E. Sumpio, *J. Vasc. Surg.* **41**, 349–354 (2005).
3. K. E. Schlageter, P. Molnar, G. D. Lapin, D. R. Groothuis, *Microvasc. Res.* **58**, 312–328 (1999).
4. G. Helmlinger, F. Yuan, M. Dellian, R. K. Jain, *Nat. Med.* **3**, 177–182 (1997).
5. M. E. Stoker, A. M. Gerdes, J. F. May, *Anat. Rec.* **202**, 187–191 (1982).
6. H. Ronellenfitsch, E. Katifori, *Phys. Rev. Lett.* **123**, 248101 (2019).
7. F. Corson, *Phys. Rev. Lett.* **104**, 048703 (2010).
8. M. A. Skylar-Scott *et al.*, *Nat. Biomed. Eng.* **6**, 449–462 (2022).
9. B. Grigoryan *et al.*, *Science* **364**, 458–464 (2019).
10. M. A. Skylar-Scott, J. Mueller, C. W. Visser, J. A. Lewis, *Nature* **575**, 330–335 (2019).
11. W. Wu, A. DeConinck, J. A. Lewis, *Adv. Mater.* **23**, H178–H183 (2011).
12. S. Anbazhakan *et al.*, *Nat. Cardiovasc. Res.* **1**, 775–790 (2022).
13. A. Goyal *et al.*, *IEEE Trans. Med. Imaging* **32**, 56–72 (2013).
14. D. B. Kolesky, K. A. Homan, M. A. Skylar-Scott, J. A. Lewis, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 3179–3184 (2016).
15. X. Cui, T. Boland, *Biomaterials* **30**, 6221–6227 (2009).
16. J. S. Miller *et al.*, *Nat. Mater.* **11**, 768–774 (2012).
17. W. Wu *et al.*, *Soft Matter* **6**, 739–742 (2010).
18. D. L. L. Ho *et al.*, *Adv. Healthc. Mater.* **11**, e2201138 (2022).
19. M. S. Olufsen *et al.*, *Ann. Biomed. Eng.* **28**, 1281–1299 (2000).
20. B. Kaimovitz, Y. Lanir, G. S. Kassab, *Ann. Biomed. Eng.* **33**, 1517–1535 (2005).
21. M. A. Galarreta-Valverde, M. M. G. Macedo, C. Mekkaoui, M. P. Jackowski, in *Proceedings Volume 8669 Medical Imaging 2013: Image Processing*, S. Ourselin, D. R. Haynor, Eds. (SPIE, 2013), p. 86691I.
22. R. W. Glenny, *J. Appl. Physiol.* **110**, 1119–1129 (2011).
23. M. Georg, T. Preusser, H. K. Hahn, "Global constructive optimization of vascular systems" (Technical Report WUCSE-2010-11, Washington University in St. Louis, 2010).
24. H. J. Kim, H. C. Rundfeldt, I. Lee, S. Lee, *Biomech. Model. Mechanobiol.* **22**, 1095–1112 (2023).
25. J. Shen, A. H. Faruqi, Y. Jiang, N. Maftoon, *IEEE Access* **9**, 20648–20661 (2021).
26. W. Schreiner *et al.*, *Med. Eng. Phys.* **28**, 416–429 (2006).
27. B. Kerautret, P. Ngo, N. Passat, H. Talbot, C. Jaquet, *Image Processing On Line* **13**, 258–279 (2023).
28. E. Jessen, M. C. Steinbach, D. Schilinger, [sarXiv:2410.06002 [physics.bio-ph] (2024).
29. C. Jaquet *et al.*, *IEEE Trans. Biomed. Eng.* **66**, 946–955 (2019).
30. Materials and methods are available as supplementary materials.
31. A. R. Pries, T. W. Secomb, P. Gaehtgens, *Am. J. Physiol.* **275**, H349–H360 (1998).
32. A. A. Guy, A. W. Justin, D. M. Aguilar-Garza, A. E. Markaki, *IEEE Trans. Biomed. Eng.* **67**, 1650–1663 (2020).
33. The Database Center for Life Science, BodyParts3D/Anatomography; http://lifesciencedb.jp/bp3d/.
34. G. D. M. Talou, S. Safaei, P. J. Hunter, P. J. Blanco, *Sci. Rep.* **11**, 6180 (2021).
35. A. Vardanian *et al.*, unum-cloud/usearch: v0.5.0. Zenodo (2023); https://doi.org/10.5281/zenodo.7949416.
36. V. T. Nguyen, T. W. Secomb, *FASEB J.* **36**, fasebj.2022.36.S1.R5364 (2022).
37. A. Updegrove *et al.*, *Ann. Biomed. Eng.* **45**, 525–541 (2017).
38. A. L. Marsden, M. Esmaily-Moghadam, *Appl. Mech. Rev.* **67**, 030804 (2015).
39. Y. Huo, G. S. Kassab, *Am. J. Physiol. Heart Circ. Physiol.* **291**, H1074–H1087 (2006).
40. M. R. Pfaller *et al.*, *Int. J. Numer. Methods Biomed. Eng.* **38**, e3639 (2022).
41. R. Oberhoffer, D. Lang, K. Feilen, *Eur. J. Pediatr.* **148**, 389–392 (1989).
42. J. P. Alberding, T. W. Secomb, *PLOS Comput. Biol.* **17**, e1009164 (2021).
43. T. Bhattacharjee *et al.*, *Sci. Adv.* **1**, e1500655 (2015).
44. M. A. Alioglu *et al.*, *Small Methods* **8**, e2301325 (2024).
45. Y. Fang *et al.*, *Adv. Mater.* **35**, e2205082 (2023).
46. A. L. Marsden, *Annu. Rev. Fluid Mech.* **46**, 519–546 (2014).
47. K. Menon *et al.*, *npj Imaging* **2**, 9 (2024).
48. Z. A. Sexton *et al.*, svVascularize: Initial release of SVV package, Version 0.1.0. Zenodo (2025); https://doi.org/10.5281/zenodo.15151168.
49. Z. A. Sexton *et al*., Supplemental Data: Rapid model-guided design of organ-scale synthetic vasculature for biomanufacturing [Dataset]. Zenodo (2025); https://doi.org/10.5281/zenodo.15588568.

ACKNOWLEDGMENTS

We thank E. Calderon for support with printing perfusion models. We also thank Stanford University and the Stanford Research Computing Center (Sherlock) for high-performance computing resources (3D CFD analysis) and the Stanford Center for Innovations in In vivo Imaging (SCi3) for micro-CT work using a Bruker SkyScan 1276 CT scanner. **Funding:** This project was supported, in part, by funding from the following entities: National Science Foundation (NSF) Graduate Research Fellowship Program grant DGE-1656518 (Z.A.S., S.S.); National Institutes of Health (NIH) CHIP T32 grant T32HL166155 (Z.A.S.); Swiss National Science Foundation grant P500PM_214204 (D.R.); Berg Scholars Program (J.E.H.); Alpha Omega Alpha Carolyn L. Kuckein Student Research Fellowship (J.E.H.); Dorothy Dee and Marjorie Helene Boring Trust Award (J.E.H.); Medical Scholars Research Program (J.E.H.); National Heart, Lung, and Blood Institute grant K99HL155777 (D.J.S.); Stanford Graduate Fellowship (S.S., A.M.); NIH grant R01EB029362-04 (A.M., J.P.); Parker B. Francis Fellowship (J.M.S.); American Heart Association Established Investigator Award (S.M.W.); Hoffmann/Schroepfer Foundation (S.M.W.); Joan and Sanford I. Weill Scholar Fund (S.M.W.); NSF RECODE grant 2134897 (S.M.W.); Breakthrough T1D grant 2-SRA-2021-1024-S-B (A.W.F.); National Heart, Lung, and Blood Institute grant DP2HL168563 (M.A.S.-S.); NIH grant 5R01HL141712-05 (A.M.); NSF SIMCARDIO grant (A.M.); Advanced Research Projects Agency for Health (ARPA-H) under award number AY1AX000002; content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or ARPA-H; Additional Ventures Cures Collaborative; Stanford Cardiovascular Institute; and NIH S10 Shared Instrumentation Grant 1S10OD02349701 **Author contributions:** Conceptualization: Z.A.S., D.R., J.E.H., S.M.W., A.W.F., M.A.S.-S., A.L.M.; Methodology: Z.A.S., D.R., J.E.H., A.R.H., S.S., A.M., J.M.S., S.M.W., M.A.S.-S., A.L.M.; Software: Z.A.S., J.E.H., S.S., F.S.S., J.P.; Validation: Z.A.S., D.R., J.E.H., J.D.; Formal analysis: Z.A.S., D.R., J.D., D.J.S.; Investigation: Z.A.S., D.R., J.E.H., A.R.H., S.S., J.D., D.J.S., A.M., F.S.S., J.M.S., S.M.W.; Visualization: Z.A.S., D.R., J.E.H., A.R.H., J.D., D.J.S.; Resources: Z.A.S., J.E.H., A.R.H., S.S., F.S.S., S.M.W., A.L.M.; Data curation: Z.A.S., D.R., J.E.H.; Funding acquisition: S.M.W., A.W.F., M.A.S.-S., A.L.M.; Project administration: M.A.S.-S., A.L.M.; Supervision: S.M.W., A.W.F., M.A.S.-S., A.L.M.; Writing – original draft: Z.A.S., J.E.H., A.R.H., J.D., D.J.S., J.M.S., S.M.W., M.A.S.-S., A.L.M.; Writing – review & editing: Z.A.S., D.R., J.E.H., J.D., J.P., A.W.F., J.M.S., S.M.W., M.A.S.-S., A.L.M. **Competing interests:** A.R.H. and A.W.F. are employees and D.J.S. is a consultant of FluidForm Bio, Inc., a start-up commercializing FRESH 3D printing. A.R.H., A.W.F., and D.J.S. have equity in FluidForm Bio, Inc. FRESH 3D printing is patented (US Patents 10,150,258, 11,672,887, and others). M.A.S.-S. received consulting fees from 3D Systems and holds shares in Formlabs, Inc., and stock options in Acoustica Bio, Inc. M.A.S.-S. serves on the scientific advisory board for Additional Ventures Foundation and served on the scientific advisory board for Acoustica Bio, Inc. M.A.S.-S. is listed as an inventor on a patent filing related to the nested biofabrication approach. A.L.M. is on the scientific advisory board of Additional Ventures Foundation and is a founder and consultant for BioGraft, Inc. All other authors declare no competing interests. **Data and materials availability:** The code is available in GitHub at https://github.com/SimVascular/svVascularize and is also archived and available on Zenodo (*48*). Additional data obtained in support of the generated figures is available on Zenodo (*49*). The project package is also available as a python module, which can be pip installed from the Python Package Index available at https://pypi.org/project/svv/.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adj6152
Materials and Methods; Supplementary Text; Figs. S1 to S15; Tables S1 to S6; References (*50*–*89*); Movies S1 to S7

Submitted 10 August 2023; resubmitted 26 October 2024; accepted 24 April 2025
10.1126/science.adj6152

NITROGEN FIXATION

Global importance of nitrogen fixation across inland and coastal waters

Robinson W. Fulweiler[1,2]*, Shelby Rinehart[3], Jason Taylor[4], Michelle C. Kelly[5], Megan E. Berberich[5], Nicholas E. Ray[6], Autumn Oczkowski[7], Sawyer Balint[1], Mar Benavides[8,9,10], Matthew J. Church[11], Brianna Loeks[12], Silvia Newell[13], Malin Olofsson[14], Jimmy Clifford Oppong[15], Sarah S. Roley[16], Carmella Vizza[17], Samuel T. Wilson[18], Subhadeep Chowdhury[9,10], Peter Groffman[19,20], J. Thad Scott[21], Amy M. Marcarelli[5]

Biological nitrogen fixation is a key driver of global primary production and climate. Decades of effort have repeatedly updated nitrogen fixation estimates for terrestrial and open ocean systems, yet other aquatic systems in between have largely been ignored. Here we present an evaluation of nitrogen fixation for inland and coastal waters. We demonstrate that water column and sediment nitrogen fixation is ubiquitous across these diverse aquatic habitats, with rates ranging six orders of magnitude. We conservatively estimate that, despite accounting for less than 10% of the global surface area, inland and coastal aquatic systems fix 40 (30 to 54) teragrams of nitrogen per year, equivalent to 15% of the nitrogen fixed on land and in the open ocean. Inland systems contribute more than half of this biological nitrogen fixation.

Nitrogen is essential to life because it is a fundamental component of nucleic acids and proteins (*1*). Globally, nitrogen limits carbon fixation and biological growth across both geological and biological timescales. Ironically, Earth is bathed in nitrogen, as its atmosphere is a large reservoir of dinitrogen (N_2) gas. However, N_2 gas is broadly inert, making it unusable to most, but not all, organisms on the planet. Biological nitrogen fixation, a microbially mediated process (*2*, *3*), converts N_2 into biologically available nitrogen—connecting the large atmospheric pool of N_2 to the biosphere. This biologically fixed nitrogen fuels primary production and enhances carbon storage in terrestrial systems and carbon export in the ocean, thus ultimately affecting Earth's climate through atmospheric carbon dioxide regulation (*4*–*6*). Because of its importance, much research has focused on quantifying rates of nitrogen fixation, and we now have a broad understanding of the role it plays in terrestrial and open ocean environments (*7*), which make up ~90% of Earth's surface.

Biological nitrogen fixation also occurs in the inland and coastal aquatic systems that cycle and transport materials between land and sea. While inland and coastal waters account for <10% of Earth's surface area, they play a disproportionate role in other biogeochemical processes [e.g., carbon cycling (*8*–*12*)]. Yet estimates of global nitrogen fixation have relied on a limited number of studies from inland and coastal systems or have ignored these environments altogether [e.g., (*13*, *14*)].

The poor integration of inland and coastal aquatic ecosystems with global nitrogen fixation estimates is the result of various factors (*15*). Whereas some early research demonstrated that rates could be high in these systems (*16*–*19*), synthesis work across inland and coastal waters concluded that nitrogen fixation rates correlated with trophic state and were dominated by planktonic cyanobacteria (*20*). Yet cyanobacteria blooms experience constraints that limit fixed nitrogen inputs to ecosystems (*21*), and sediment nitrogen fixation is not uncommon in these systems (*22*–*25*). Although streams and rivers were mostly excluded from early syntheses, these ecosystems have since been shown to have substantial and dynamic patterns of nitrogen fixation (*26*, *27*). Earlier research also often assumed that because nitrogen fixation is an energy-intensive process, it would only occur under conditions of nitrogen limitation. More recently, we have come to understand that the enzyme responsible for nitrogen fixation, nitrogenase, may benefit microorganisms in capacities beyond just nitrogen acquisition (e.g., maintaining intracellular redox equilibrium), implying that nitrogen fixation can affect ecosystem nitrogen availability regardless of nitrogen limitation (*28*–*30*). Finally, it may be the result of a broader research trend of the 1960s and 1970s, focused on quantifying the impacts of human activities on inland and coastal waters. Specifically, excess nutrient loading was starting to be linked to a range of deleterious consequences such as eutrophication. Since earlier syntheses concluded that nitrogen fixation likely only contributed a small percentage of the total nitrogen input for inland and coastal nitrogen budgets, research efforts focused on quantifying other inputs as well as nitrogen removal pathways (*31*). More recent estimates suggest that nitrogen fixation can account for a consequential portion of nitrogen input to these systems (*32*–*35*).

More than three decades beyond the original synthesis efforts (*20*, *36*), the paradigm that nitrogen fixation is dominated by photosynthetic organisms, controlled by inorganic nitrogen availability, unimportant in sediments, and inconsequential for inland and coastal systems lingers in large-scale syntheses of the global nitrogen cycle (*13*, *14*, *37*). Here we demonstrate that this paradigm is incorrect. Instead, we show that nitrogen fixation is widely detected across aquatic habitats, hourly rates are on par with and even exceed those reported for terrestrial and open ocean systems, and it commonly occurs in sediments. We also estimate the contribution of inland and coastal aquatic nitrogen fixation to the global reactive nitrogen budget. Further, we proffer a different paradigm: Nitrogen fixation is ubiquitous across the aquatic land-ocean continuum and is an important process in sediments, and inland and coastal waters have a disproportionate impact on the global nitrogen budget, contributing more nitrogen through fixation than their areal coverage would suggest.

Inland and coastal water nitrogen fixation rates

We leveraged a recently synthesized global database (*38*, *39*) of nitrogen fixation rates published between 1990 and 2022 for inland and coastal systems to investigate the distribution, magnitude, and global importance of nitrogen fixation across these environments. For this analysis, we define inland and coastal waters to include the water column and benthos of lakes (including reservoirs), rivers (including streams), freshwater wetlands, salt marshes, mangroves, seagrasses, tidal flats, estuaries, and continental shelves. We include >4500 nitrogen fixation rates that provide global coverage, although, unsurprisingly, most

[1]Department of Earth and Environment, Boston University, Boston, MA, USA. [2]Department of Biology, Boston University, Boston, MA, USA. [3]Department of Biodiversity, Earth and Environmental Science, Drexel University, Philadelphia, PA, USA. [4]Water Quality and Ecology Research Unit, National Sedimentation Laboratory, Agricultural Research Service, United States Department of Agriculture, Oxford, MS, USA. [5]Department of Biological Sciences, Michigan Technological University, Houghton, MI, USA. [6]School of Marine Science and Policy, University of Delaware, Lewes, DE, USA. [7]Atlantic Coastal Environmental Sciences Division, United States Environmental Protection Agency, Narragansett, RI, USA. [8]National Oceanography Centre, European Way, Southampton, UK. [9]Aix Marseille Univ, Université de Toulon, CNRS, IRD, MIO UM 110, Marseille, France. [10]Turing Center for Living Systems, Aix-Marseille University, Marseille, France. [11]Flathead Lake Biological Station, University of Montana, Polson, MT, USA. [12]Water Resources Science, University of Minnesota, St. Paul, MN, USA. [13]School for Environment and Sustainability, University of Michigan, Ann Arbor, MI, USA. [14]Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Uppsala, Sweden. [15]Institute of Environmental Studies, Charles University, Prague, Czech Republic. [16]School of the Environment, Washington State University, Richland, WA, USA. [17]Department of Natural Science, Hawai'i Pacific University, Honolulu, HI, USA. [18]School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, UK. [19]Earth and Environmental Science, CUNY Advanced Science Research Center, Brooklyn College, New York, NY, USA. [20]Cary Institute of Ecosystem Studies, Millbrook, NY, USA. [21]Department of Biology, Baylor University, Waco, TX, USA. ***Corresponding author. Email: rwf@bu.edu**

rates are from the Northern Hemisphere (Fig. 1). This geographic bias is consistently observed in a variety of disciplines and is yet another clear call for expanding the geography of our collective research efforts (*40*–*42*). Nonetheless, this collection of nitrogen fixation rates is the most complete record ever assembled for inland and coastal systems and enables the most thorough evaluation of their distribution and importance to date.

Overall, we find that nitrogen fixation rates across the land-ocean continuum are highly variable both within and across habitats and that the distribution of these rates are habitat specific (Fig. 2). Water column rates are most variable for lakes, estuaries, and continental shelves, with lakes and rivers exhibiting significantly ($P < 0.001$) higher water column nitrogen fixation compared with other habitats, although river water column fixation rates are infrequently quantified (Fig. 2A and fig. S1A). Similarly, benthic rates are quite variable (Fig. 2B). Lake benthic nitrogen fixation rates are significantly ($P < 0.001$) lower than all other habitats, except for rivers (Fig. 2B and fig. S1B). Tidal flat benthic nitrogen fixation rates are significantly ($P < 0.001$) higher than all other habitats, except for freshwater wetlands (Fig. 2B and fig. S1B).

While most of the reported nitrogen fixation data are from temperate zones, the limited data from polar and tropical regions suggest that substantial nitrogen fixation can occur in their water columns (table S1) and benthos (table S2). Polar and tropical lake water columns exhibit the highest rates across all habitats and climate zones, with polar lake rates being seven times higher than those of tropical lakes (table S1). In contrast, water column rates from continental shelves, regardless of climate zone, as well as temperate freshwater wetlands and seagrasses were two orders of magnitude lower (table S1). For the benthos, tropical rivers and lakes exhibit the highest rates, followed by polar continental shelves and temperate tidal flats (table S2). Polar and temperate freshwater wetland rates are two orders of magnitude higher than their tropical counterparts (table S2). The lack of data from polar and tropical climates limits our ability to assess the role of climate in driving inland and coastal aquatic nitrogen fixation rates. Nevertheless, these data add to the emerging understanding of the importance of tropical and polar systems in global nitrogen cycling and are similar to patterns reported for terrestrial (*43*) and open ocean ecosystems (*44*).

The variability in nitrogen fixation rates is likely driven by a combination of factors, none of which are mutually exclusive, including methodological differences and temporal and spatial variability in ecosystem conditions. The dataset used for this study includes rates measured by the acetylene reduction assay (ARA), $^{15}N_2$ labeling, and the net

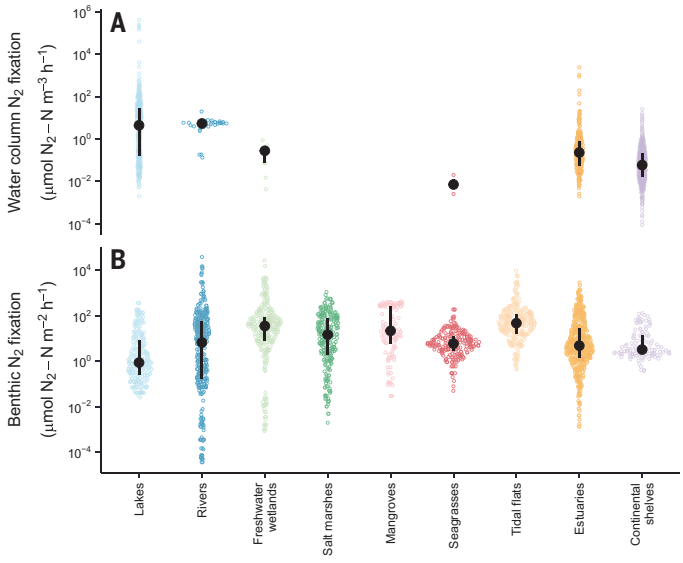


Fig. 2. Inland and coastal aquatic system biological nitrogen fixation rates. Distribution of hourly nitrogen fixation rates in the water column (**A**) and benthos (**B**) of the inland and coastal habitats studied here. Median (Q1–Q3) are also shown (black dot with bars). Unit conversion details can be found in the methods in the supplementary materials.

N_2/Ar technique (fig. S2). These methods rely on different assumptions, and each assumption has biases that influence our understanding of nitrogen fixation. These methods are, however, the common ways in which nitrogen fixation rates have historically been and are currently being measured. ARA was the most common method, constituting 64% of the total observations we present here. While logistically simple and relatively inexpensive (*45*), this method has known weaknesses (*46*), particularly in sediments. Acetylene is known to inhibit nitrogen-fixing microbes, including methanogens (*47*), and it has been shown to both up- and down-regulate sulfate reducers (*48*). Another challenge relates to the conversion ratio of ethylene produced to N_2 fixed equivalents. A ratio of 3:1 (molar) is often assumed. This ratio can vary substantially however, and there is no consensus on what drives this variation nor how variable it is across systems and time (*20*, *49*). The $^{15}N_2$ method, which accounted for 32% of the measurements, is often considered the gold standard for quantifying nitrogen fixation. Recent efforts in open ocean pelagic nitrogen fixation research have modified the technique and in doing so found potential underestimations of water column nitrogen fixation (*50*, *51*). There are major challenges for using $^{15}N_2$ labeling in sediments, including the assumption that the isotope tracer is quickly and evenly distributed throughout the sample. It is also difficult to detect the isotopic label against a large background of sediment organic nitrogen. Finally, the N_2/Ar technique, which emerged as an additional method in the mid-1990s, makes up 4% of the rate dataset. This technique provides a net measurement of N_2 production through denitrification and consumption through nitrogen fixation and thus is surely an underestimate of gross nitrogen fixation rates. Nevertheless, this method requires fewer assumptions, has no known impact on the microbial community, and can be applied across experimental scales

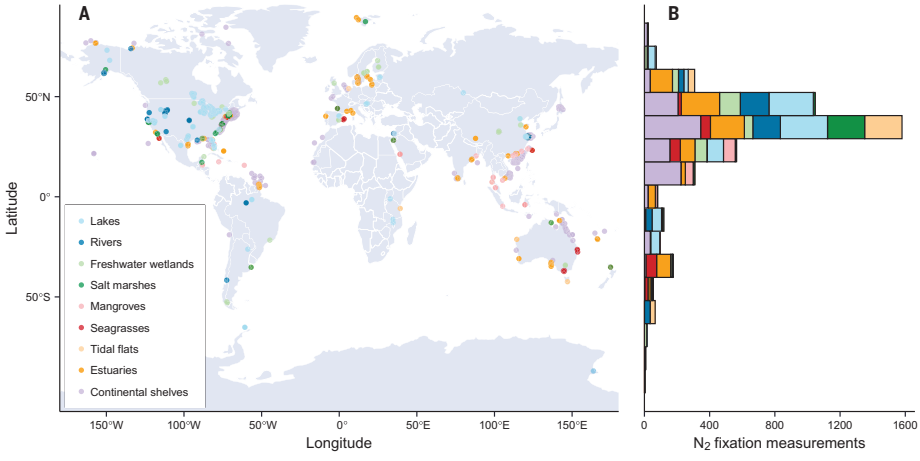


Fig. 1. Distribution of biological nitrogen fixation rates across inland and coastal aquatic systems included in this study. Global distribution of study sites with colors highlighting habitat type (**A**) and the number of nitrogen fixation measurements by latitude and habitat type (**B**).

Table 1. Estimate of the total annual contribution of nitrogen to the global nitrogen budget by the water column and benthos of inland and coastal waters. See methods in the supplementary materials for details on scaling assumptions.

Region	Habitat	Substratum	*n*	Median N_2 fixation (g N m^{-2} $year^{-1}$)	Annual N contribution (Tg N $year^{-1}$)
Inland	Lakes	Water column	534	0.296 (0.214–0.412)	0.948 (0.686–1.32)
		Benthos	321	1.73 (1.43–2.11)	5.53 (4.58–6.76)
	Rivers	Water column	36	0.0154 (0.0121–0.0196)	0.00966 (0.00756–0.0123)
		Benthos	426	6.26 (3.27–11.2)	3.92 (2.05–7.01)
	Freshwater wetlands	Water column	11	0.000423 (0.000423–0.0014)	0.00355 (0.00355–0.0121)
		Benthos	323	1.63 (1.25–2.05)	13.7 (10.5–17.2)
	Inland total		**1650**	**1.97 (1.46–2.64)**	**24.1 (17.8–32.3)**
Coastal	Salt marshes	Water column	0	–	–
		Benthos	274	1.16 (0.935–1.48)	0.064 (0.0514–0.0814)
	Mangroves	Water column	0	–	–
		Benthos	139	2.02 (1.8–2.12)	0.297 (0.265–0.312)
	Seagrasses	Water column	2	0.0000638 (0.0000289–0.0000944)	0.0000102 (0.00000463–0.0000151)
		Benthos	234	0.596 (0.492–0.725)	0.0954 (0.0778–0.116)
	Tidal flats	Water column	0	–	–
		Benthos	323	2.44 (2.07–2.69)	0.312 (0.265–0.344)
	Estuaries	Water column	274	0.027 (0.0239–0.0335)	0.0275 (0.0244–0.0342)
		Benthos	575	0.356 (0.286–0.424)	0.363 (0.292–0.432)
	Continental shelves	Water column	919	0.0169 (0.0138–0.0202)	0.445 (0.364–0.534)
		Benthos	135	0.557 (0.413–0.746)	14.7 (10.9–19.7)
	Coastal total		**2880**	**0.584 (0.437–0.774)**	**16.3 (12.3–21.6)**
Inland and coastal systems combined total			**4530**	**1.01 (0.75–1.34)**	**40.4 (30.1–53.9)**

from microcosms to ecosystems (*35*, *52*, *53*). As we move nitrogen fixation research forward, there is an urgent need for intercalibration of these methods across habitats and substrates. Additionally, there are best practices for data collection and reporting that can be implemented for improved quantification of nitrogen fixation in the future (*39*).

The variance in nitrogen fixation rates is also driven by temporal and spatial variability. Most nitrogen fixation data are collected during the growing season, missing inter- and intra-annual changes. Yet both short- and long-term studies for a variety of biogeochemical processes can vary substantially over hourly, monthly, and yearly timescales (*32*, *54*–*60*). These data highlight that nitrogen fixation follows similar dynamics. Inland and coastal systems are naturally dynamic areas that routinely experience a variety of physical, chemical, and biological shifts that may influence rates of nitrogen fixation. For example, runoff, freshwater input, stratification, tidal cycles, storm events, and diel changes in light, pH, and oxygen are all temporally shifting conditions that can affect biological nitrogen fixation. Of course, these conditions also shift spatially. For example, water columns can exhibit strong horizontal and vertical gradients in light, oxygen, pH, suspended particles, and nutrient concentrations. These conditions can affect abundance of cyanobacteria and particle-associated and free-living nitrogen fixers (*61*–*66*). Benthic habitats are particularly challenging because they are often extremely heterogeneous. For example, in rivers and other shallow freshwater habitats, the benthos can include organic sediments, rocks, and submerged woody debris, all of which host different assemblages of nitrogen-fixing taxa (*62*, *67*). In ecosystems with rooted aquatic macrophytes, both the leaves and the rhizosphere can host nitrogen fixers (*68*–*73*), while the roots can supply sediments with oxygen and dissolved organic carbon excretions that affect nitrogen fixation (*74*, *75*). Similarly, both living and detrital macroalgae can host nitrogen fixers (*76*–*78*). Nitrogen-fixing microbes are also associated with epi- and infauna (*79*, *80*), although their importance to host organisms and to ecosystem nitrogen availability is unknown. We suspect that much of the variability in rates depends on where and when we sample—a one-time or seasonal sampling scheme likely fails to capture the full variability of these systems.

These data also highlight a need for a reevaluation of the role of nitrogen fixation in the benthos, particularly in sediments (Fig. 2B). Indeed, the conditions that often characterize sediments (e.g., anoxic, high organic carbon concentrations) may provide an optimal environment for nitrogen fixation. Energy limitation is unlikely to be a concern for nitrogen fixers in shallow benthic environments that contain high concentrations of organic matter and other reduced substrates that can be used as electron donors. Sediments also support a variety of metabolic pathways associated with nitrogen fixation. Surface sediments may support cyanobacterial nitrogen fixation, whereas subsurface sediments can support both chemoheterotrophic and chemoautotrophic nitrogen fixation (*28*). This diversity of nitrogen-fixing metabolisms may help us understand why sediment nitrogen fixation is observed even under high concentrations of porewater inorganic nitrogen (*81*, *82*); it is possible that these high concentrations of inorganic nitrogen may even be due to nitrogen fixation. Additionally, sediments are often rich in the metals needed to synthesize nitrogenase, the enzyme catalyzing the reduction of N_2 to ammonia. Finally, nitrogenase is inhibited by oxygen and thus requires nitrogen fixers in oxygenated environments to expend additional resources protecting this enzyme.

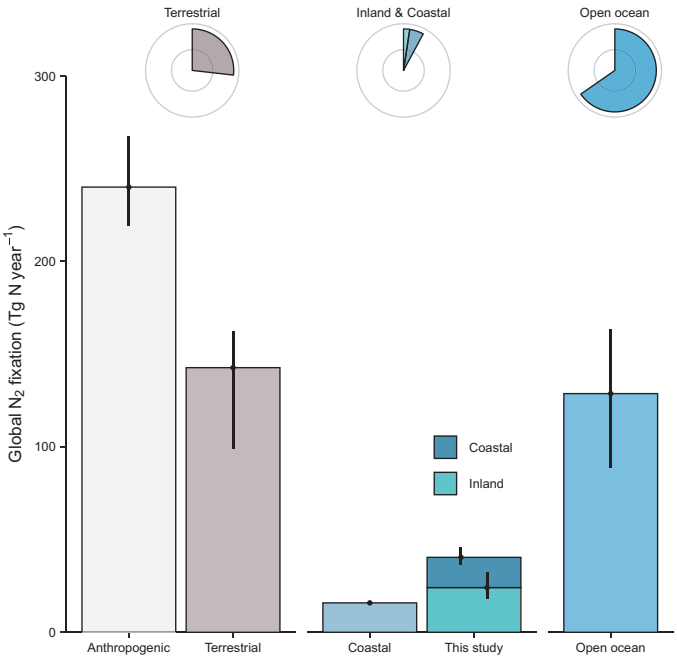


Fig. 3. Inland and coastal waters have a disproportionate impact on global nitrogen cycling. A comparison of the median (Q1–Q3) of commonly reported nitrogen fixation estimates from anthropogenic (n = 4), terrestrial (n = 11), coastal (n = 20), and open ocean (n = 20) systems and those reported in this study for inland and coastal waters. Above each system, we show the area percentage of the globe for each system. Note that in this study we report a median estimate, whereas most other studies report an arithmetic mean or, in a few cases, a geometric mean. References compiled for use in this figure can be found in table S7.

Because sediments are typically anoxic within the top few centimeters to millimeters (especially those that are organic-rich, like those found inland and coastal waters), nitrogen fixers in these environments may expend little to no energy protecting nitrogenase.

Together, these data demonstrate that nitrogen fixation is a highly dynamic process that likely plays a key, and yet largely unknown, role in inland and coastal ecosystem function. The variance of these rates suggests that nitrogen fixation fluctuates across spatial and temporal scales such as microhabitats or triggering events, making it hard to detect.

Inland and coastal water contribution to the global nitrogen budget

Here we report an evaluation of the total annual contribution of nitrogen fixation by inland and coastal waters. To do this, we used the latitude associated with each nitrogen fixation rate to bin them into three primary climate zones: polar, temperate, and tropical. Rates were then scaled to common units of grams of nitrogen per square meter per year (g N m^{-2} $year^{-1}$) for each climate zone, following the methods outlined in the last review of inland and coastal water nitrogen fixation (*20*). Given the high variability in nitrogen fixation rates, habitat-specific statistical distributions, methodological differences, and geographic bias of data reported in the literature, we report medians for each habitat and substratum as a conservative estimate (Table 1). We note that terrestrial and open ocean estimates commonly report means, which are more easily biased by outlier values. This analysis reveals that hourly nitrogen fixation rates for inland and coastal waters (Table 1) are on par with those reported for terrestrial and open ocean systems (*83*–*86*).

Because of the paucity of data for many climate zones, we chose to scale habitats in climate zones with limited data using the entire dataset for that habitat (see methods in the supplementary materials). We then multiplied habitat-specific annual rates by ecosystem area within each climate zone to estimate the contribution of inland and coastal aquatic system nitrogen fixation to the global nitrogen budget (table S6; methods). Our global synthesis reveals a median [quartile 1 to quartile 3 (Q1–Q3)] contribution from inland and coastal aquatic ecosystem nitrogen fixation of 40 (30 to 54) Tg N $year^{-1}$ (Table 1).

This conservative estimate of the annual contribution of nitrogen fixation from inland and coastal waters is equivalent to ~30% of the average estimated contributions made by nitrogen fixation on land (~143 Tg N $year^{-1}$) or in the open ocean (~130 Tg N $year^{-1}$) (Fig. 3 and Table 1). Inland and coastal aquatic systems account for ~8% of global surface area, yet they contribute ~15% of total global biological nitrogen fixation. We propose that most of this nitrogen represents a distinct contribution to the budget, as terrestrial estimates do not typically include contributions made from lakes and rivers, nor do open ocean budgets typically incorporate the coastal systems included here. The addition of inland and coastal aquatic system biological nitrogen fixation to estimates from terrestrial and marine ecosystems increases the total contribution of biological nitrogen fixation to the global nitrogen budget from ~270 to ~310 Tg N $year^{-1}$.

Most global nitrogen studies ignore nitrogen fixation in lakes and rivers and only account for a relatively minor contribution from freshwater wetlands (*20*). Here we estimate that inland aquatic ecosystems contribute 24 Tg N $year^{-1}$, which accounts for 60% of the total nitrogen contribution through nitrogen fixation from inland and coastal aquatic systems (Table 1). This inland water contribution is driven by benthic nitrogen fixation across all freshwater habitats (Table 1). High rates in some of these habitats, such as lake benthos, have been observed since nitrogen fixation measurements were first made in these systems (*16*).

We estimate that coastal systems provide ~16 Tg N $year^{-1}$. We note that the most commonly cited previous estimate did not include tidal flats, which we do include here, but did include coral reefs, which we did not include (*36*). Combining the previously estimated coral reef value (2.8 Tg N $year^{-1}$) with our estimate increases the total coastal contribution to 19 Tg N $year^{-1}$. The coastal contribution estimated here is driven by sediment nitrogen fixation, particularly from the continental shelf. Another recent effort estimated a similar rate for coastal water columns alone (*87*), reinforcing that the coastal ocean is a key area for nitrogen fixation.

This study demonstrates that nitrogen fixation is a dynamic and important process across inland and coastal aquatic habitats. Nitrogen fixation commonly occurs across a wider range of habitats than previously thought, and the benthos plays a particularly important role in global production of reactive nitrogen. Nitrogen fixation in these ecosystems appears to contribute substantially to the global nitrogen budget, despite their small areal coverage. The fate and ecological role of this fixed nitrogen is largely unknown. The fixed nitrogen may help alleviate organism nitrogen limitation, potentially through sharing the nitrogen in microbial consortia or in symbiotic and syntrophic relationships (*28*). It likely also affects other important biogeochemical processes (e.g., carbon cycling) within inland and coastal systems and beyond. Biological nitrogen fixation in terrestrial systems and in the open ocean is responsible for a substantial portion of carbon accumulation (*5*), new production (*87*), and export to the deep ocean (*88*). The potential for biological nitrogen fixation to affect inland and coastal water carbon (and other element) cycling is large and a critical topic for future studies.

We anticipate that this study will motivate future research that improves detection and quantification techniques for nitrogen fixation, expands our spatial and temporal coverage of nitrogen fixation rates globally with an emphasis on studying underrepresented climate zones, and better constrains environmental drivers of nitrogen fixation (*39*). We also urge research focused on system-specific as well as global modeling of nitrogen fixation for more nuanced spatial and temporal upscaling and inclusion of inland and coastal aquatic habitats in global nitrogen budget estimates.

REFERENCES AND NOTES

1. H. R. Rucker, B. Kaçar, *Trends Microbiol.* **32**, 554–564 (2024).
2. H.-W. Pi *et al.*, *Mol. Biol. Evol.* **39**, msac181 (2022).
3. T. H. Coale *et al.*, *Science* **384**, 217–222 (2024).
4. Z. Wen *et al.*, *Sci. Adv.* **8**, eabl7564 (2022).
5. J. Peng *et al.*, *J. Clean. Prod.* **202**, 1158–1166 (2018).
6. N. Gruber, in *The Ocean Carbon Cycle and Climate*, M. Follows, T. Oguz, Eds., vol. 40 of *NATO Science Series: IV* (Springer, 2004), pp. 97–148.
7. X. Zhang, B. B. Ward, D. M. Sigman, *Chem. Rev.* **120**, 5308–5351 (2020).
8. L. Tranvik, J. J. Cole, Y. T. Prairie, *Limnol. Oceanogr. Lett.* **3**, 41–48 (2018).
9. G. Rocher-Ros *et al.*, *Nature* **621**, 530–535 (2023).
10. P. Regnier, L. Resplandy, R. G. Najjar, P. Ciais, *Nature* **603**, 401–410 (2022).
11. J. A. Rosentreter *et al.*, *Nat. Clim. Chang.* **13**, 579–587 (2023).
12. J. J. Cole *et al.*, *Ecosystems* **10**, 172–185 (2007).
13. D. Fowler *et al.*, *Philos. Trans. R. Soc. London Ser. B* **368**, 20130164 (2013).
14. M. Voss *et al.*, *Philos. Trans. R. Soc. London Ser. B* **368**, 20130121 (2013).
15. A. M. Marcarelli, R. W. Fulweiler, J. T. Scott, *Limnol. Oceanogr. Lett.* **7**, 1–10 (2022).
16. P. L. Brezonik, C. L. Harper, *Science* **164**, 1277–1279 (1969).
17. M. A. Keirn, P. L. Brezonik, *Limnol. Oceanogr.* **16**, 720–731 (1971).
18. R. Dugdale, V. Dugdale, J. Neess, J. Goering, *Science* **130**, 859–860 (1959).
19. J. J. Goering, J. C. Neess, *Limnol. Oceanogr.* **9**, 530–539 (1964).
20. R. W. Howarth, R. Marino, J. Lane, J. J. Cole, *Limnol. Oceanogr.* **33**, 669–687 (1988).
21. J. T. Scott, M. J. McCarthy, *Limnol. Oceanogr.* **55**, 1265–1270 (2010).
22. E. M. Grantz, A. Kogo, J. T. Scott, *Limnol. Oceanogr.* **57**, 925–935 (2012).
23. E. K. Eberhard, S. M. Techtmann, C. V. Baxter, A. M. Marcarelli, *Ecosystems* **28**, 18 (2025).
24. S. Wang *et al.*, *ISME J.* **18**, wrae119 (2024).
25. Y. Xin *et al.*, *Water Res.* **264**, 122239 (2024).
26. N. B. Grimm, K. C. Petrone, *Biogeochemistry* **37**, 33–61 (1997).
27. A. Patoine, M. D. Graham, P. R. Leavitt, *Limnol. Oceanogr.* **51**, 1665–1677 (2006).
28. R. W. Fulweiler, *Global Biogeochem. Cycles* **37**, e2023GB007777 (2023).
29. V. J. Bertics *et al.*, *Biogeosciences* **10**, 1243–1258 (2013).
30. H. M. Joshi, F. R. Tabita, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14515–14520 (1996).
31. S. W. Nixon, *Hydrobiologia* **629**, 5–19 (2009).
32. R. W. Fulweiler, E. M. Heiss, *Oceanography* **27**, 184–195 (2014).
33. S. E. Newell, M. J. McCarthy, W. S. Gardner, R. W. Fulweiler, *Estuaries Coasts* **39**, 1626–1638 (2016).
34. U. Larsson, S. Hajdu, J. Walve, R. Elmgren, *Limnol. Oceanogr.* **46**, 811–820 (2001).
35. R. W. W. Fulweiler, S. W. W. Nixon, B. A. A. Buckley, S. L. L. Granger, *Nature* **448**, 180–182 (2007).
36. D. G. Capone, E. J. Carpenter, *Science* **217**, 1140–1142 (1982).
37. W. H. Schlesinger, E. S. Bernhardt, *Biogeochemistry: An Analysis of Global Change* (Academic Press, ed. 4, 2020).
38. R. W. Fulweiler *et al.*, Global dataset of nitrogen fixation rates across inland and coastal waters based on a coordinated synthesis effort ver 2, Environmental Data Initiative (2025); https://doi.org/10.6073/pasta/333f651ca721da657d5fd0c393d26cf8.
39. R. W. Fulweiler *et al.*, *Limnol. Oceanogr. Lett.* 10.1002/lol2.10459 (2025).
40. A. M. Vieillard, S. E. Newell, S. F. Thrush, *J. Geophys. Res. Biogeosci.* **125**, e2020JG005766 (2020).
41. H. L. S. Cheung *et al.*, *Limnol. Oceanogr.* **69**, 621–637 (2024).
42. L. Boyero, A. Ramírez, D. Dudgeon, R. G. Pearson, *J. N. Am. Benthol. Soc.* **28**, 397–403 (2009).
43. Xu-Ri, I. C. Prentice, *Biogeosciences* **14**, 2003–2017 (2017).
44. L. W. von Friesen, L. Riemann, *Front. Microbiol.* **11**, 596426 (2020).
45. W. D. Stewart, G. P. Fitzgerald, R. H. Burris, *Proc. Natl. Acad. Sci. U.S.A.* **58**, 2071–2078 (1967).
46. R. J. Flett, R. D. Hamilton, N. E. R. Campbell, *Can. J. Microbiol.* **22**, 43–51 (1976).
47. G. D. Sprott, K. F. Jarrell, K. M. Shaw, R. Knowles, *Microbiology* **128**, 2453–2462 (1982).
48. R. W. Fulweiler *et al.*, *Front. Microbiol.* **6**, 418 (2015).
49. S. P. Seitzinger, J. H. Garber, *Mar. Ecol. Prog. Ser.* **37**, 65–73 (1987).
50. T. Großkopf *et al.*, *Nature* **488**, 361–364 (2012).
51. S. T. Wilson, D. Böttjer, M. J. Church, D. M. Karl, *Appl. Environ. Microbiol.* **78**, 6516–6523 (2012).
52. W. S. Gardner *et al.*, *Limnol. Oceanogr.* **51**, 558–568 (2006).
53. J. M. Taylor, I. M. Andersen, A. K. Hoke, P. T. Kelly, J. T. Scott, *Biogeochemistry* **166**, 283–301 (2023).
54. J. Tucker, A. E. Giblin, C. S. Hopkinson, S. W. Kelsey, B. L. Howes, *Estuar. Coast. Shelf Sci.* **151**, 54–68 (2014).
55. W. R. Boynton, M. A. C. Ceballos, C. L. S. Hodgkins, D. Liang, J. M. Testa, *Estuaries Coasts* **46**, 356–375 (2023).
56. B. J. Roberts, P. J. Mulholland, *J. Geophys. Res.* **112**, G04002 (2007).
57. J. W. Newbold, J. W. Elwood, R. V. O'Neill, A. L. Sheldon, *Ecology* **64**, 1249–1265 (1983).
58. P. J. Mulholland, J. D. Newbold, J. W. Elwood, L. A. Ferren, J. R. Webster, *Ecology* **66**, 1012–1023 (1985).
59. E. J. Chua, R. W. Fulweiler, *Front. Environ. Sci.* **10**, 1028405 (2023).
60. V. A. Billaud, *J. Fish. Res. Board Can.* **25**, 2101–2110 (1968).
61. N. M. Hayes *et al.*, *Freshw. Biol.* **64**, 269–283 (2019).
62. A. M. Marcarelli, W. A. Wurtsbaugh, *Biogeochemistry* **94**, 95–110 (2009).
63. A. Tilstra *et al.*, *Front. Microbiol.* **9**, 2351 (2018).
64. W. A. Wurtsbaugh, A. J. Horne, *Can. J. Fish. Aquat. Sci.* **40**, 1419–1429 (1983).
65. J. N. Pedersen, D. Bombar, R. W. Paerl, L. Riemann, *Front. Microbiol.* **9**, 2759 (2018).
66. M. R. Hamersley *et al.*, *Aquat. Microb. Ecol.* **63**, 193–205 (2011).
67. F. J. Triska, J. R. Sedell, K. Cromack Jr., S. V. Gregory, F. M. McCorison, *Ecol. Monogr.* **54**, 119–140 (1984).
68. J. L. Rolando *et al.*, *Nat. Commun.* **15**, 3607 (2024).
69. J. Cui *et al.*, *Front. Microbiol.* **13**, 851424 (2022).
70. R. Marino *et al.*, *Biogeochemistry* **166**, 211–226 (2023).
71. W. Mohr *et al.*, *Nature* **600**, 105–109 (2021).
72. L. R. Finke, H. W. Seeley Jr., *Appl. Environ. Microbiol.* **36**, 129–138 (1978).
73. H. Šantrůčková, E. Rejmánková, B. Pivničková, J. M. Snyder, *Biogeochemistry* **101**, 295–310 (2010).
74. K. J. McGlathery, N. Risgaard-Petersen, P. B. Christensen, *Mar. Ecol. Prog. Ser.* **168**, 245–258 (1998).
75. D. T. Welsh, *Ecol. Lett.* **3**, 58–71 (2000).
76. W. D. Head, E. J. Carpenter, *Limnol. Oceanogr.* **20**, 815–823 (1975).
77. Y. Raut, D. G. Capone, *Environ. Microbiol.* **23**, 4372–4388 (2021).
78. M. R. Hamersley, J. A. Sohm, J. A. Burns, D. G. Capone, *Aquat. Bot.* **125**, 57–63 (2015).
79. M. Zilius *et al.*, *Sci. Rep.* **10**, 13966 (2020).
80. S. König *et al.*, *Nat. Microbiol.* **2**, 16193 (2016).
81. A. N. Knapp, *Front. Microbiol.* **3**, 374 (2012).
82. J. Gier *et al.*, *Biogeosciences* **13**, 4065–4080 (2016).
83. P. M. Vitousek, D. N. L. Menge, S. C. Reed, C. C. Cleveland, *Philos. Trans. R. Soc. London Ser. B* **368**, 20130119 (2013).
84. C. C. Cleveland *et al.*, *Global Biogeochem. Cycles* **13**, 623–645 (1999).
85. J. P. Zehr, D. G. Capone, in *Marine Nitrogen Fixation* (Springer Cham, 2021), pp. 143–156.
86. Z. Shao *et al.*, *Earth Syst. Sci. Data* **15**, 3673–3709 (2023).
87. W. Tang *et al.*, *Nat. Commun.* **10**, 831 (2019).
88. W.-L. Wang, J. K. Moore, A. C. Martiny, F. W. Primeau, *Nature* **566**, 205–211 (2019).
89. R. Fulweiler *et al.*, Supplemental Information for: Global Importance of Nitrogen Fixation Across Inland and Coastal Waters, version 1.2.1, Zenodo (2025); https://doi.org/10.5281/zenodo.15546493.

ACKNOWLEDGMENTS

We thank the Department of Earth and Environment at Boston University for enabling us to host our first Research Coordination Network workshop. Special thanks to F. George and L. Consalvo, who helped us with workshop logistics. We thank L. Simpson for PRISMA meta-analysis methodological assistance. Thanks to N. Bartolucci, L. Champlin, C. Mahoney, E. Wilson, S. Glastein, M. Hagy, L. Jefferson, F. Johnson II, and B. Webster, who provided helpful comments on this manuscript. Thanks to A. Geisser, who provided artistic guidance with Fig. 3. Many thanks to M. Dietze for consultation and review of our statistical analyses. R.W.F. gives special thanks to V. Farcus, who assisted on a variety of manuscript tasks. We thank two anonymous reviewers whose insightful comments and careful edits much improved our manuscript. Disclaimer: The use of trade, firm, or corporation names is solely for the information and convenience of the reader. Mention of names does not constitute an official endorsement or approval by the USDA or the Agricultural Research Service of any product or service to the exclusion of others that may be suitable. The USDA prohibits any discrimination in all its programs and activities on the basis of race, color, national origin, age, disability, sex, marital status, familial status, parental status, religion, sexual orientation, genetic information, political beliefs, reprisal, or because any part of an individual's income is derived from any public assistance program. **Funding:** Funding for this work was provided by National Science Foundation grant 2015825 to J.T.S., R.W.F., and A.M.M. R.W.F. also received funding from Rhode Island Sea Grant for this work. **Author contributions:** Conceptualization: R.W.F., A.M.M., J.T.S.; Methodology: R.W.F., S.R., J.T., S.B., A.M.M., J.T.S.; Investigation: All authors; Visualization: S.B., M.C.K., M.E.B., N.E.R., R.W.F.; Funding acquisition: J.T.S., A.M.M., R.W.F.; Supervision: R.W.F.; Writing – original draft: R.W.F.; Writing – review & editing: All authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The nitrogen fixation observations used in this study have been deposited to the Environmental Data Initiative (*38*). The code used to perform global scaling and generate the figures in this manuscript is available at Zenodo (*89*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adt1511
Materials and Methods; Figs. S1 and S2; Tables S1 to S7; References (*90*–*126*)

Submitted 13 September 2024; accepted 16 April 2025

10.1126/science.adt1511

CLIMATE

Observed trend in Earth energy imbalance may provide a constraint for low climate sensitivity models

Gunnar Myhre[1]*, Øivind Hodnebrog[1], Norman Loeb[2], Piers M. Forster[3]

Climate forcings by greenhouse gases and aerosols cause an imbalance at the top of the atmosphere between the net incoming solar radiation and outgoing longwave radiation from Earth. This Earth energy imbalance has strengthened over the period 2001 to 2023 with satellite data. Here, we show that low climate sensitivity models fail to reproduce the trend in Earth energy imbalance, particularly in the individual longwave and shortwave contributions to the imbalance trend. The inability to produce a strong positive shortwave and strong negative longwave Earth energy imbalance trend is found to be a robust feature in the low climate sensitivity models, especially for models with a climate sensitivity below 2.5 kelvin. The negative longwave contribution to Earth energy imbalance is driven by surface temperature increases and is therefore most pronounced in high climate sensitivity models, whereas the shortwave contribution is generally positive and amplified by greater surface warming.

A long-standing research question in climate science is how sensitive the climate is to increases in greenhouse gases (GHGs) (*1*–*3*). This climate sensitivity is taken as the surface temperature rise for a doubling of the CO_2 concentration (*4*, *5*). In the latest Intergovernmental Panel on Climate Change (IPCC) report, the best estimate of the equilibrium climate sensitivity (ECS) was assessed as 3°C, with a likely range from 2.5° to 4°C and a very likely range from 2° to 5°C (*4*). How clouds change in a warmer world is the main cause of the uncertainty in the climate sensitivity (*4*, *5*), with divergent results from observational studies (*6*–*9*). The recent warming over the first one to two decades of this century has been used as arguments for low climate sensitivity models being most realistic, in particular how feedback processes are represented for the recent warming trend (*10*, *11*). However, the pattern of observed sea-surface warming in the Pacific may have biased some of these findings (*12*).

The past decades have seen a continued increase in GHGs (*4*) combined with a reversal of the aerosol effect (*13*). A reduction in the cooling effect of aerosols has thus a warming effect, and the total effective radiative forcing has been accelerating over the past decades (*14*). The Earth energy imbalance (EEI) is increasing (*15*, *16*) and will likely give an accelerated warming over the coming years (*17*). Hodnebrog *et al.* (*16*) showed that climate models forced with observed sea-surface temperatures (SSTs) reproduce the satellite-retrieved strengthening in EEI from the Clouds and the Earth's Radiant Energy System (CERES), but all models have a weaker trend than the observed trend. Schmidt *et al.* (*18*) showed that the EEI trend, split into longwave (LW) and shortwave (SW) trends, differed markedly between the CERES satellite and in different configurations of a climate model. Here we use a large set of coupled climate models from the Coupled Model Intercomparison Project Phase 6 (CMIP6) (*19*) to illustrate that low climate sensitivity models have an EEI trend behavior that is inconsistent with the satellite-derived EEI trend.

[1]CICERO Center for International Climate Research, Oslo, Norway. [2]NASA Langley Research Center, Hampton, VA, USA. [3]Priestley Centre for Climate Futures, University of Leeds, Leeds, UK. *Corresponding author. Email: gunnar.myhre@cicero.oslo.no

Trend in EEI

Figure 1 shows the EEI over the period with CERES satellite data and compared with coupled climate model simulations from CMIP6. The model simulations are a combination of the historical simulation until 2014 combined with a SSP5-8.5 scenario from 2015 onward. The SSP5-8.5 scenario includes reductions in aerosols combined with a strong increase in GHG concentrations. The CERES data show a stronger trend in EEI than the multimodel CMIP6 mean and higher EEI in 2023 than any of the CMIP6 models. However, for individual CMIP6 models and ensembles, EEI is comparable to or higher at other periods than the CERES value in 2023. Interannual variability in EEI is clearly shown for the CERES data and the climate models. In simulations with observed SST, the interannual variability in the CERES data is largely reproduced by the climate models (*16*).

The strengthened EEI from CERES is further supported by an accelerated trend in the ocean heat content (OHC) (*14*, *20*–*22*). Discrepancies exist in the degree of acceleration among various OHC datasets, with best agreement between CERES and OHC datasets having better ocean coverage and filling in data in data-sparse regions (*23*).

Relationship of LW and SW EEI trend

To illustrate climate model differences and robustness between models, we use idealized CMIP6 experiments. In Fig. 2, the EEI is shown for a range of models for the experiment with a 1%/year increase in CO_2 concentrations (named 1pctCO2). All models show an increasing net EEI (Fig. 2A) but with a much larger model diversity when net EEI is split into LW (Fig. 2B) and SW EEI (Fig. 2C). LW and SW EEI are positive when reducing outgoing radiation at the top of the atmosphere, typically caused by an increase in absorption by GHGs (both anthropogenic and as climate feedback). An increase in surface temperature causes an increase in outgoing LW radiation and thus a negative LW EEI. The majority of models have a negative LW EEI after some years because the increase in surface temperature and more outgoing LW radiation overwhelm the positive effect from the increase in CO_2. However, several models have a positive or very weak LW EEI even after more than 100 years. With one exception, the models have a positive SW EEI mostly throughout the time period of increase in CO_2 caused by less snow and ice (*24*) and contributions from water vapor absorption (*25*) and clouds for several models (see discussions in next sections). Figure 2D shows the trend in LW EEI versus SW EEI for the CMIP6 models, with uncertainties reflecting variation among four 23-year periods over the model simulations. Consistent with Schmidt *et al.* (*18*), we find a robust linear relationship in the LW and SW EEI trends among model members and a marked spread in the trends. Periods of 23 years are selected to match the length of data available from CERES.

Linking trends in EEI with climate sensitivity

Figure 3A shows the 1pctCO2 experiment and the abrupt quadrupling of CO_2 (abrupt-4xCO2) experiment, with colors reflecting the ECS. The LW and SW EEI trends are calculated as a mean of four 23-year trends and ECS from regressions using the abrupt-4xCO2 experiment over 150 years (*26*) and thus consistent with ECS values derived elsewhere (*4*, *27*). The 1pctCO2 and abrupt-4xCO2 experiments show different trend relationships for LW and SW EEI, with the former having a weaker LW negative trend gradient as a sustained increase in CO_2 contributes to LW EEI imbalance. The shading in Fig. 3A shows that the net EEI is generally increasing in the 1pctCO2 experiment and decreasing in the abrupt-4xCO2 experiment. Figure 3A also includes atmosphere-only simulations with observed SST fields from 2001 to

2019 from Hodnebrog *et al.* (*16*). Atmosphere-only simulations in which climate drivers (GHGs and aerosols) have been kept constant for the period 2001 to 2019 align with the abrupt-4xCO2 experiment, and atmosphere-only simulations with an increase in GHGs (but constant aerosols) over the 2001 to 2019 period show LW and SW trends similar to those of the 1pctCO2 experiment. Notably, the low climate sensitivity models have much weaker changes in LW and SW EEI trends than the other models. In particular, the models with ECS below 2.5 K all show very weak LW and SW EEI trends. For models with an ECS of 4 K or higher, there is little alignment with the LW and SW EEI trends, and the models are widely spread along the regression line. Nevertheless, it is notable that none of these models show very weak LW and SW EEI trends.

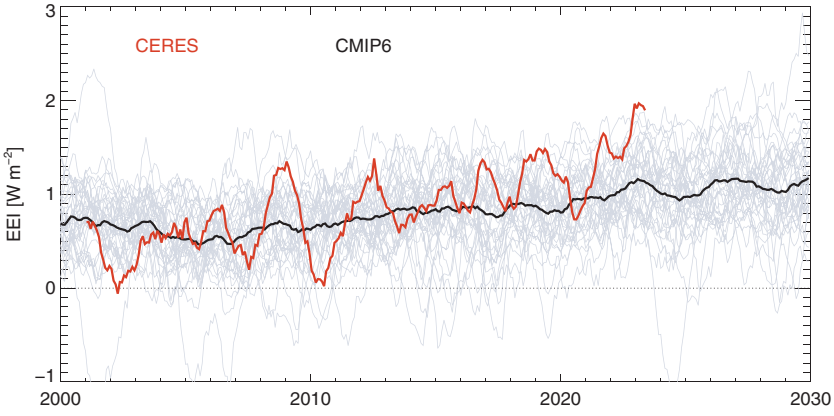


Fig. 1. Trend in EEI in CERES and CMIP6 models. The CERES data are shown from 2001 to 2023. The CMIP6 data are shown from 2000 to 2030. All EEI are given as 12-month running means. CMIP6 model mean is shown by a thick black line and individual models are shown in thin gray lines. Only one ensemble member for each of the models is shown.

In Fig. 3B, the CERES satellite LW and SW EEI trend is shown together with the CMIP6 intramodel ensemble mean LW and SW EEI trend for the period 2001 to 2023. CMIP6 simulations are from a combination of the historical simulation until 2014 and the SSP5-8.5 scenario from 2015. Results are very similar for other scenarios with aerosol reductions (e.g., SSP2-4.5). Additionally, atmosphere-only simulations with changes in GHG and aerosols from Hodnebrog *et al.* (*16*) are included in Fig. 3B; note that these are for the period 2001 to 2019. The number of ensemble members for each model is quite variable (table S1). Figure 3C shows results for all ensemble members included in this study. Figure 3, B and C, show systematic weak LW and SW EEI trends from low climate sensitivity models, consistent with Fig. 3A. Note that the relationship between climate sensitivity and net EEI trends shows no systematic pattern, and various intramodel ensemble members exhibit a wide range in the net EEI trend (fig. S1). All models with climate sensitivity below 2.5 K have very weak LW and SW EEI trends. It is worth mentioning some of the EEI trends of lighter colors in Fig. 3, B and C. FGOALS-f3-L has a LW EEI trend similar to that of CERES and a SW EEI trend 0.2 W m^{-2}/decade weaker than that of CERES, but this model has a climate sensitivity of 3.0 K (Fig. 3B). In Fig. 3C, one ensemble member out of 50 for MIROC6 has a SW EEI trend above 0.4 W m^{-2}/decade. MIROC6 has a climate sensitivity of 2.6 K. GISS-E2-2-G is among the very low climate sensitivity models, with an ECS of 2.4 K, and has one out of five ensemble members with a SW EEI trend above 0.4 W m^{-2}/decade (0.41) but with a near-zero LW EEI. The atmosphere-only simulations in Fig. 3B are closer to CERES LW and SW EEI trends than the fully coupled simulations of the same model. This can be illustrated by HadGEM3, which shows slightly weaker trends than CERES where observed SST is used, and much stronger trends in the coupled simulations. Similarly, NorESM2

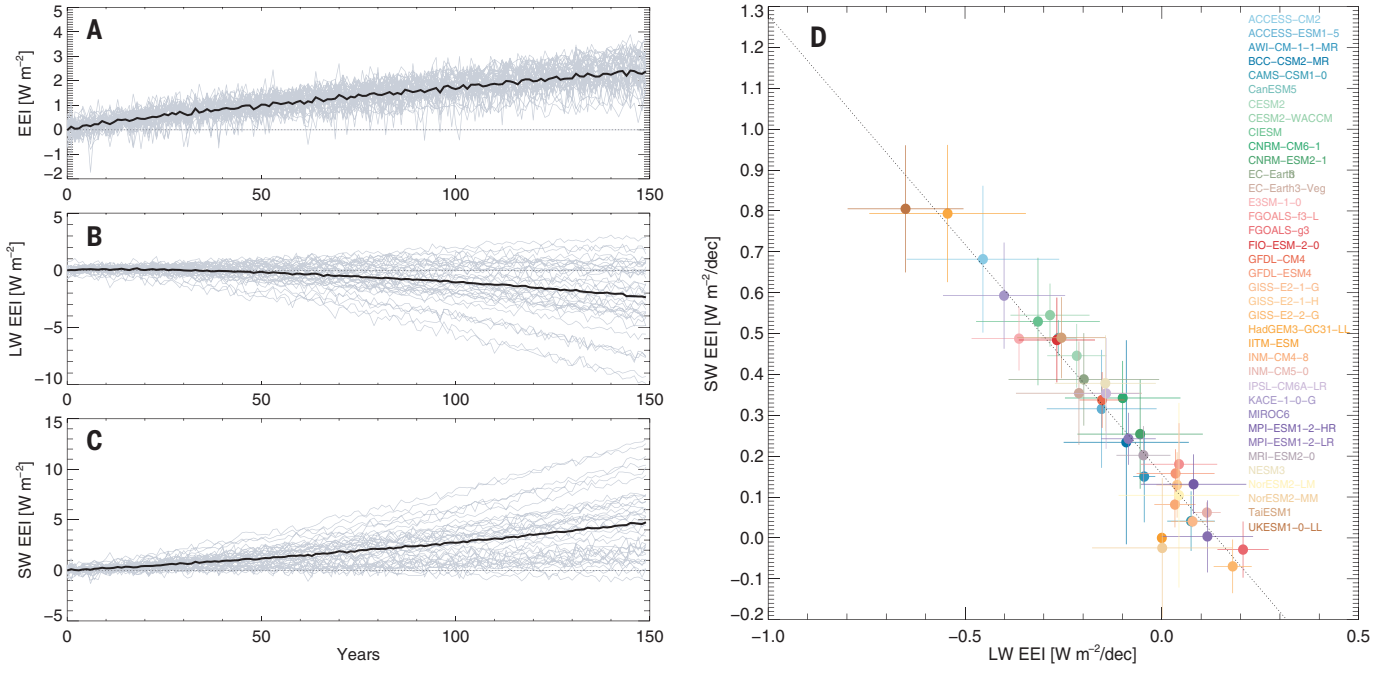


Fig. 2. EEI in CMIP6 models of 1%/year increase in CO_2 (1pctCO2). Net EEI (**A**), LW EEI (**B**), SW EEI (**C**), and SW EEI versus LW EEI trends where trends are derived from four 23-year intervals (**D**). Uncertainty ranges shown in (D) represent the standard deviation among the four 23-year intervals.

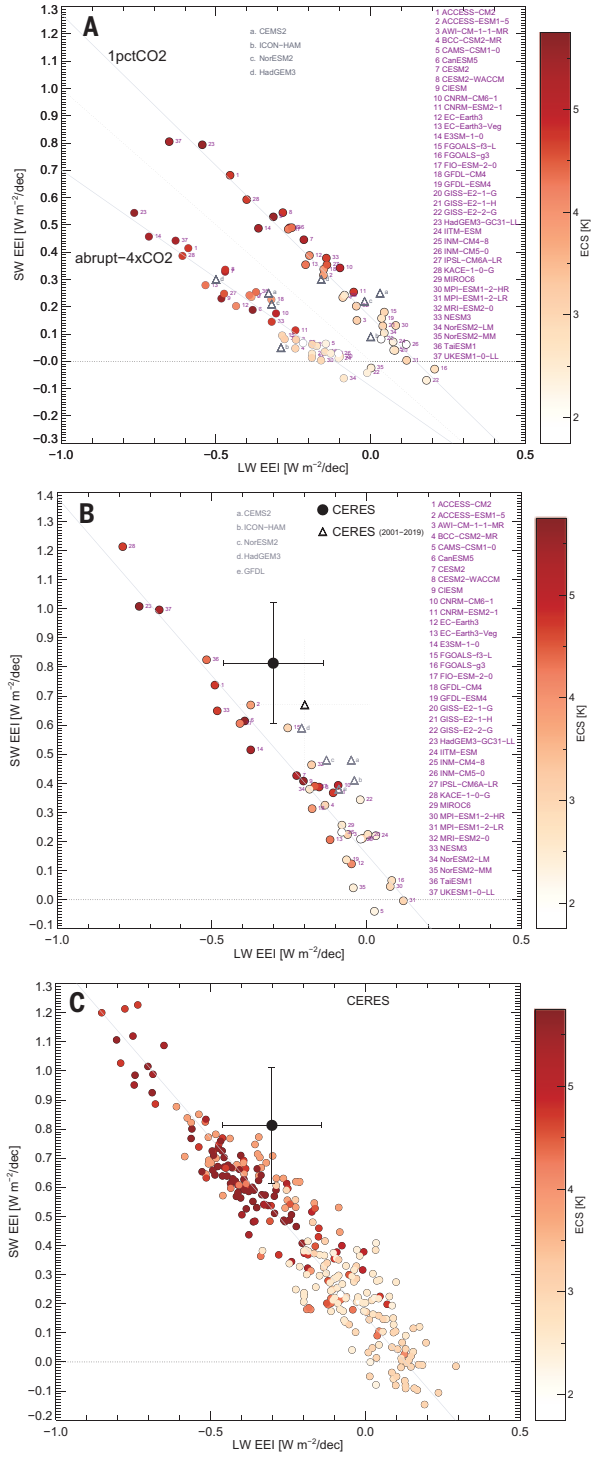


Fig. 3. SW EEI versus LW EEI trends and their relation to climate sensitivity in CMIP6 models. Primary ensemble member from each CMIP6 model for the abupt-4xCO2 and 1pctCO2 experiments with results from Hodnebrog *et al*. (*16*) on all driver constants (following abupt-4xCO2) and only GHG changes included (following 1pctCO2) (**A**); CERES and ensemble mean from each CMIP6 models (historical+SSP5-8.5) for 2001 to 2023 with results from Hodnebrog *et al*. (*16*) shown in gray colors for the period 2001 to 2019 (see further description in supplementary text) (**B**); and CERES and all individual ensemble members from CMIP6 models for 2001 to 2023 (**C**). The range of ECS in the CMIP6 models is from 1.9 to 5.6 K. In (A), the 1pctCO2 simulations have thicker lines around the circles than the abrupt-4xCO2 simulations. Yellow-gray shaded area in (A) shows where net EEI trend is negative. Lines around the CERES trends are 90% confidence intervals.

aligns better with CERES data when using observed SST simulations, showing stronger LW and SW EEI trends than in fully coupled simulations. However, differences with CERES are also evident among model simulations using observed SSTs, indicating that both the atmospheric and ocean components of the climate models contribute to differences in the LW and SW EEI trends.

Clear-sky LW and SW EEI trends show similar patterns to all-sky trends, with strong negative LW EEI trends for high climate sensitivity models and positive SW EEI trends in idealized CMIP6 experiments (fig. S2). The largest difference between the clear-sky and all-sky results are for SW EEI trends (Fig. 3A and fig. S2).

EEI trends and surface warming

The surface temperature warming differs substantially between the CMIP6 models, and Fig. 4 (and fig. S3 and fig. S4) investigates whether this alters the relationships shown in Fig. 3. The most notable result is that the CERES data show a higher SW EEI trend per degree warming than any of the CMIP6 models, and only the models with a strong negative LW EEI trend per degree warming are close to CERES (Fig. 4). The low climate sensitivity models have consistently much weaker LW and SW EEI trends per degree warming than the CERES data.

Discussion

The analysis above relates to models' long-term climate sensitivity estimated from abrupt 4xCO2 experiments. This can differ from their effective climate sensitivity, estimated from changes over the recent historical period. Previous studies find that climate models are unable to capture the recent pattern of East Pacific warming observed (*28*), and this is associated with reduced effective feedbacks and reduced effective climate sensitivity (*12*). Hodnebrog *et al*. (*16*) find that applying SST patterns to models improves EEI trends across models. When comparing the observed SST simulations (triangles) to their coupled modeled counterparts (circles) in Fig. 3B, low-sensitivity models such as NorESM exhibit relatively modest differences in the EEI trends and remain well below the CERES trend even with the observed SST pattern applied. Hence the particular observed SST pattern is unlikely to play a large role for the low climate sensitivity models, substantially underestimating the LW EEI and SW EEI trends compared to CERES.

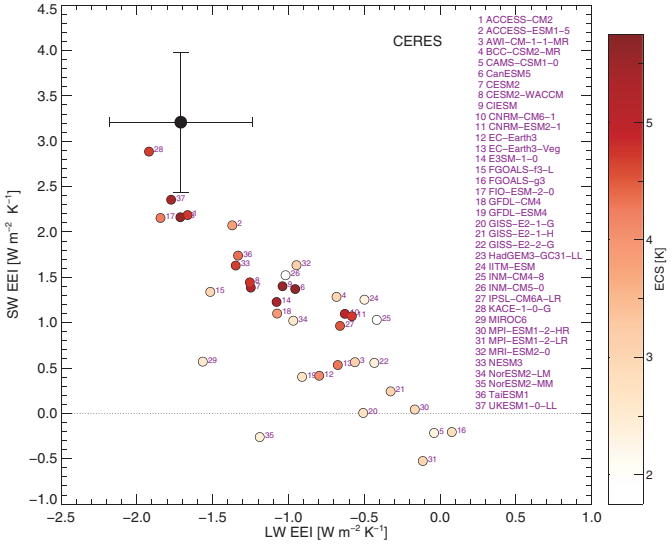


Fig. 4. SW EEI per degree warming trends versus LW EEI trends per degree warming in CMIP6 models and CERES satellite data. Colors show climate sensitivity of each CMIP6 model. Temperatures for the CERES data are taken from GISTEMP (*31*). Lines around the CERES trends are 90% confidence intervals.

Climate models consistently show a robust feature of a relationship between LW EEI and SW EEI trends, which varies depending on climate drivers involved in the simulations (Fig. 3, A and B). This relationship is also evident under clear-sky conditions (see fig. S2), where the positive SW EEI trend often is modest and driven by reduced surface albedo from less snow and ice in addition to contributions from SW absorption by water vapor. The negative LW EEI trend is driven by surface temperature increase and moderated if GHGs are increasing during the simulations. Cloud changes further amplify the clear-sky relationship between LW EEI and SW EEI trends. Figure S5 shows previously derived SW and LW cloud feedbacks (*27*) that exhibit a similar shape of relationship to that of LW EEI and SW EEI trends but slightly different gradient. Notably, SW cloud feedback demonstrates greater model diversity than LW cloud feedback, although the latter also shows substantial variability. A negative correlation between LW and SW cloud feedbacks can be expected as a result of the opposing warming and cooling effect of clouds on the climate system (*29*).

We show, using a large set of climate models, that trend in net EEI has no clear relationship to climate sensitivity. Consequently, we argue that the trends in net EEI and surface warming trend over the first two decades of this century provide little constraint on climate sensitivity. However, we present robust findings for trends in LW and SW EEI. These trends, and their relationship to climate sensitivity, are more physically based than the net EEI trend. The model distribution of EEI trends compared to CERES is shown in fig. S6. All models, given as the 99.999% level of the distribution, with an ECS of 2.93 K or below, are outside the CERES range. The models have a positive aerosol radiative forcing trend similar to that observed, around 0.16 W m^{-2} per decade (*16*, *30*). This would need to be underestimated by at least 50% to make the SW EEI trend from models with an ECS of 2.5 K match the CERES range, making such a low ECS unlikely.

REFERENCES AND NOTES

1. R. Knutti, M. A. A. Rugenstein, G. C. Hegerl, *Nat. Geosci.* **10**, 727–736 (2017).
2. J. Hansen *et al.*, *Science* **213**, 957–966 (1981).
3. G. H. Roe, M. B. Baker, *Science* **318**, 629–632 (2007).
4. P. Forster *et al.*, in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change,* V. Masson-Delmotte *et al.*, Eds. (Cambridge Univ. Press, 2021).
5. S. C. Sherwood *et al.*, *Rev. Geophys.* **58**, RG000678 (2020).
6. P. Ceppi, P. Nowack, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2026290118 (2021).
7. G. V. Cesana *et al.*, *Commun. Earth Environ.* **5**, 181 (2024).
8. I. Tan, T. Storelvmo, M. D. Zelinka, *Science* **352**, 224–227 (2016).
9. R. Vogel *et al.*, *Nature* **612**, 696–700 (2022).
10. A. Modak, T. Mauritsen, *Geophys. Res. Lett.* **48**, e2020GL091779 (2021).
11. D. Olonscheck, M. Rugenstein, *Geophys. Res. Lett.* **51**, e2023GL106909 (2024).
12. K. C. Armour *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2312093121 (2024).
13. J. Quaas *et al.*, *Atmos. Chem. Phys.* **22**, 12221–12239 (2022).
14. P. M. Forster *et al.*, *Earth Syst. Sci. Data* **15**, 2295–2327 (2023).
15. N. G. Loeb *et al.*, *Geophys. Res. Lett.* **48**, e2021GL093047 (2021).
16. Ø. Hodnebrog *et al.*, *Commun. Earth Environ.* **5**, 166 (2024).
17. J. E. Hansen *et al.*, *Oxford Open Climate Change* **3**, kgad008 (2023).
18. G. A. Schmidt *et al.*, *Front. Clim.* **5**, 1202161 (2023).
19. V. Eyring *et al.*, *Geosci. Model Dev.* **9**, 1937–1958 (2016).
20. Z. Li, M. H. England, S. Groeskamp, *Nat. Commun.* **14**, 6888 (2023).
21. A. Minière, K. von Schuckmann, J.-B. Sallée, L. Vogt, *Sci. Rep.* **13**, 22975 (2023).
22. K. von Schuckmann *et al.*, *Earth Syst. Sci. Data* **15**, 1675–1709 (2023).
23. N. G. Loeb *et al.*, *Surv. Geophys.* **45**, 1757–1783 (2024).
24. A. Duspayev, M. G. Flanner, A. Riihelä, *Geophys. Res. Lett.* **51**, e2024GL109608 (2024).
25. A. Donohoe, K. C. Armour, A. G. Pendergrass, D. S. Battisti, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 16700–16705 (2014).
26. J. M. Gregory *et al.*, *Geophys. Res. Lett.* **31**, L03205 (2004).
27. M. D. Zelinka *et al.*, *Geophys. Res. Lett.* **47**, e2019GL085782 (2020).
28. M. Rugenstein *et al.*, *Geophys. Res. Lett.* **50**, e2023GL105488 (2023).
29. V. Ramanathan *et al.*, *Science* **243**, 57–63 (1989).
30. P. M. Forster *et al.*, *Earth Syst. Sci. Data* **16**, 2625–2658 (2024).
31. N. J. L. Lenssen *et al.*, *J. Geophys. Res. Atmos.* **124**, 6307–6326 (2019).
32. G. Myhre, Fig3b,c [Data set], Zenodo. https://doi.org/10.5281/zenodo.15087491.

ACKNOWLEDGMENTS

Funding: The European Union's Horizon 2020 research and innovation program under grant agreement 820829 (CONSTRAIN) (G.M., Ø.H., P.M.F.) and project grants no. 325270 by the Norwegian Research Council (G.M., Ø.H.). **Author contributions:** Conceptualization: G.M. Methodology: G.M., Ø.H., N.L., P.M.F. Investigation: G.M., Ø.H., N.L., P.M.F. Formal analysis: G.M. Visualization: G.M. Validation: G.M., Ø.H., N.L., P.M.F. Writing – original draft: G.M. Writing – review & editing: G.M., Ø.H., N.L., P.M.F. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** CMIP6 model data used in this study are freely available from the CMIP6 repository on the Earth System Grid Federation nodes (https://esgf-node.llnl.gov/search/cmip6/, World Climate Research Programme, 2020). Derived LW and SW EEI from CMIP6 data presented in Fig. 3, B and C, are available at Zenodo (*32*). CERES EBAF-TOA Edition4.2 data were obtained from the NASA Langley Research Centre CERES ordering tool at https://ceres.larc.nasa.gov/data/. Global temperatures are from GISTEMP Team, 2024: GISS Surface Temperature Analysis (GISTEMP), version 4, NASA Goddard Institute for Space Studies, dataset accessed 2024-07-05 at https://data.giss.nasa.gov/gistemp/. Results for observed SST simulations are from Hodnebrog *et al.* (*16*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adt0647
Materials and Methods; Supplementary Text; Figs. S1 to S6; Tables S1 and S2; References (*33*, *34*)

Submitted 12 September 2024; accepted 7 April 2025

10.1126/science.adt0647

MICROBIOLOGY

Antagonism as a foraging strategy in microbial communities

Astrid K. M. Stubbusch[1,2,3]*†, François J. Peaudecerf[4,5], Kang Soo Lee[6,5], Lucas Paoli[7,8], Julia Schwartzman[9], Roman Stocker[5], Marek Basler[10], Olga T. Schubert[1,2], Martin Ackermann[1,2,11]‡, Cara Magnabosco[3]‡, Glen G. D'Souza[1,2,12,13]‡

In natural habitats, nutrient availability limits bacterial growth. We discovered that bacteria can overcome this limitation by acquiring nutrients by lysing neighboring cells through contact-dependent antagonism. Using single-cell live imaging and isotopic markers, we found that during starvation, the type VI secretion system (T6SS) lysed neighboring cells and thus provided nutrients from lysing cells for growth. Genomic adaptations in antagonists, characterized by a reduced metabolic gene repertoire, and the previously unexplored distribution of the T6SS across bacterial taxa in natural environments suggest that bacterial antagonism may contribute to nutrient transfer within microbial communities in many ecosystems.

Bacteria require nutrients to synthesize cellular building blocks such as nucleotides, amino acids, and fatty acids for growth. In nature, nutrients are often found only at low concentrations or in the form of complex polymers that require specific enzymatic degradation before they can be used. Under such conditions, the biomass of other cells that are rich in common metabolites and cellular building blocks could serve as a valuable nutrient source (*1*). Killing neighboring cells can give access to these nutrients.

Bacteria have evolved various systems to lyse neighboring bacteria by transferring toxins such as the type IV and type VI secretion systems (T4SS and T6SS, respectively) (*2*–*4*). These contact-dependent antagonistic systems are widespread in cultured bacteria (*3*). Their dominant ecological role is usually viewed as mediating competition for space and nutrients (*5*–*7*), leading to their characterization as the "weapons" of bacterial "warfare" (*3, 7*–*11*).

In this study, we investigated whether in addition to their role in killing competitors, contact-dependent antagonism mechanisms enable bacteria to acquire nutrients from target cells. This has been hypothesized previously (*12*, *13*), but a direct role in nutrient foraging has never been shown experimentally for nonpredatory bacteria. We focused on the T6SS, a well-studied bacteriophage-like molecular machinery that enables the transfer of diverse toxins into neighboring cells and is present in ~25% of all sequenced Gram-negative bacteria (*14*). The distribution of T6SS$^+$ taxa across natural environments remains uncharted.

Contact-dependent antagonism enables growth during starvation

To determine whether the biomass of target cells provides a source of nutrients for antagonistic cells, we studied a simple community consisting of the marine isolates *Vibrio cyclitrophicus* ZF270 (target cells) and the T6SS-encoding *Vibrio anguillarum* (formerly known as *Vibrio ordalii*) FS144 (T6SS$^+$ cells) as an ecological model system. We investigated the growth of T6SS$^+$ cells and the lysis of target cells at single-cell resolution within a microfluidic device, where cells can grow in a monolayer in shallow microfluidic chambers (fig. S1). When we provided the T6SS$^+$ cells with a carbon source that they could metabolize, *N*-acetylglucosamine (GlcNAc) (fig. S2), they grew exponentially with a doubling rate of ~0.2 hours^{-1} (Fig. 1, A and B; fig. S3; and movie S1). The doubling rate is a measure that is independent of the initial number of cells per chamber, which varies among chambers (fig. S3). In co-culture with target cells, we observed a similar doubling rate (Fig. 1B and fig. S3). When we provided a carbon source that only the target cells could metabolize (alginate; fig. S2), T6SS$^+$ cells failed to multiply in monoculture but were able to grow in coculture with target cells at a doubling rate of ~0.07 hours^{-1} (Fig. 1B, fig. S3, and movies S2 and *3*).

To investigate whether the growth of T6SS$^+$ cells without a metabolizable carbon source is facilitated by the T6SS-mediated killing of target cells rather than by the exchange of secreted nutrients (cross-feeding), we tested whether the growth was contingent on a functional T6SS. We used a genetic model system, *V. cholerae* 2740-80 and its T6SS-deletion mutant (*15*, *16*), paired with *Escherichia coli* as target cells. Again, we provided T6SS$^+$ cells and target cells in microfluidic chambers with a carbon source that T6SS$^+$ cells could not metabolize (melibiose; fig. S1 and movie S4). We found that the growth rate of the T6SS-deletion mutant population was lower than that of the T6SS$^+$ population (figs. S4 and S5, A and B, and movies S5 and S6), suggesting that T6SS-mediated cell lysis increased the growth of T6SS$^+$ cells under carbon starvation conditions. This finding was confirmed by growth rate measurements in bulk culture experiments (−0.1 and 0.3 hours^{-1} for the T6SS-deletion population and T6SS$^+$ population, respectively; fig. S5C).

To test whether the observed growth increase in T6SS$^+$ cells was caused by the acquisition of nutrients from target cells, we used stable isotope probing–Raman microspectroscopy (*17*). We first labeled *E. coli* target cells with deuterium and then co-cultured them with unlabeled *V. cholerae* T6SS$^+$ cells or the T6SS-deletion mutants. After 9 hours, T6SS$^+$ cells showed significant levels of deuterium incorporation compared with the T6SS-deletion cells (Fig. 1C and fig. S5, D and E), further confirming that the T6SS indeed facilitated access to nutrients.

Slow cell lysis increases the nutrient uptake by surrounding cells

A more detailed examination of T6SS-driven antagonism revealed that lysis was not instantaneous. We observed that the cell shape of attacked target cells changed from rod-shaped to round before bursting, likely because T6SS effectors compromised cell wall integrity (*18*) (tables S2 and S3). These round *V. cyclitrophicus* target cells persisted for 1.43 ± 0.20 hours in alginate medium and for 0.33 ± 0.09 hours in GlcNAc medium before bursting (mean ± 95% confidence interval) (Fig. 2A). By staining cells with propidium iodide (PI), which does not permeate intact cell membranes but can enter cells with leaky membranes (*19*, *20*), we found that round target cells incorporated more PI than target cells in their usual rod shape (Fig. 2B). Together, these findings suggest that the action of carbon-starved T6SS$^+$ cells causes slow lysis of target cells, likely leading to gradual rather than instantaneous nutrient release.

[1]Institute of Biogeochemistry and Pollutant Dynamics, Department of Environmental Systems Science, ETH Zurich, Zurich, Switzerland. [2]Department of Environmental Microbiology, Eawag: Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland. [3]Geological Institute, Department of Earth and Planetary Sciences, ETH Zurich, Zurich, Switzerland. [4]Univ Rennes, CNRS, IPR (Institut de Physique de Rennes) - UMR 6251, Rennes, France. [5]Institute of Environmental Engineering, Department of Civil, Environmental and Geomatic Engineering, ETH Zurich, Zurich, Switzerland. [6]Department of Mechanical Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan, South Korea. Ulsan, South Korea. [7]Department of Biology, Institute of Microbiology and Swiss Institute of Bioinformatics, ETH Zurich, Zurich, Switzerland. [8]Global Health Institute, School of Life Sciences, École Polytechnique Fédéral de Lausanne (EPFL), Lausanne, Switzerland. [9]Department of Biological Sciences, University of Southern California, Los Angeles, USA. [10]Biozentrum, University of Basel, Basel, Switzerland. [11]Laboratory of Microbial Systems Ecology, School of Architecture, Civil and Environmental Engineering (ENAC), École Polytechnique Fédéral de Lausanne (EPFL), Lausanne, Switzerland. [12]Center for Fundamental and Applied Microbiomics, Biodesign Institute, Arizona State University, Tempe, AZ, USA. [13]School of Molecular Sciences, Arizona State University, Tempe, AZ, USA. ***Corresponding author. Email: astrid.stubbusch@monash.edu** †Present address: Department of Microbiology, Biomedicine Discovery Institute, Monash University, Clayton, VIC, Australia. ‡These authors contributed equally to this work.

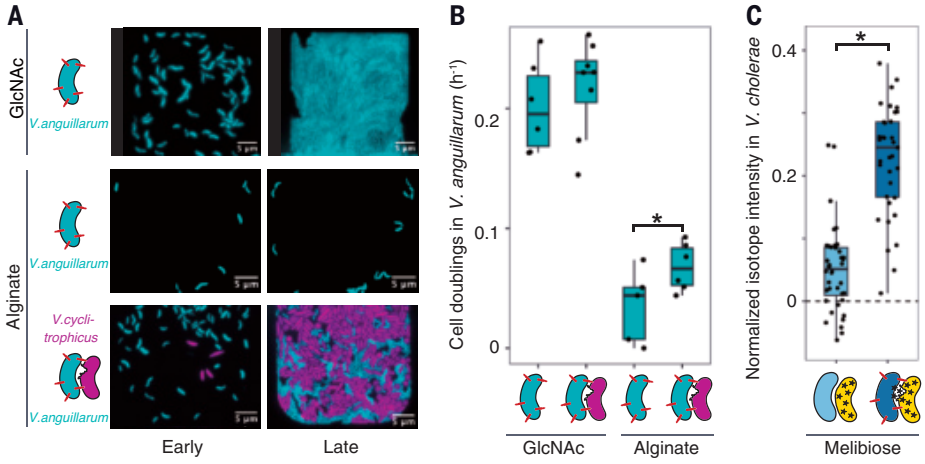


Fig. 1. Growth benefit and acquisition of nutrients released by contact-dependent antagonism through the T6SS. (**A**) Representative images of T6SS$^+$ cells (*V. anguillarum* FS144, cyan) and target cells (*V. cyclitrophicus* ZF270, magenta) supplied with different carbon sources in microfluidic chambers at the start (0 hours) and end time point (12, 24, and 24 hours, top to bottom, respectively, based on the time it took for *V. anguillarum* cells to fill the chamber). Scale bar, 5 μm. (**B**) Growth rate of T6SS$^+$ cells (*V. anguillarum* FS144) in microfluidic chambers estimated from the exponential increase of cell numbers over time shown in fig. S3. Points show the estimates from individual growth chambers (fig. S3), and box plots indicate the median and interquartile range (n = 6, 8, 5, and 6 chambers, respectively (left to right); asymptotic Wilcoxon-Mann-Whitney tests, P = 0.04). (**C**) Isotope incorporation by *V. cholerae* without T6SS (light blue) and with T6SS (dark blue) from deuterium-labeled *E. coli* (yellow). The deuterium signal in individual *V. cholerae* cells was quantified using Raman microspectroscopy. The co-culture medium contained deuterium and melibiose as a carbon source for the continued growth of labeled *E. coli*. The values were normalized by subtracting the signal of the respective *V. cholerae* cells in this medium in monoculture (fig. S5E). T6SSs are shown as red spikes and deuterium as star symbols (n = 40 and 35 cells, respectively (left to right); Wilcoxon rank-sum tests, $P = 8 \times 10^{-13}$).

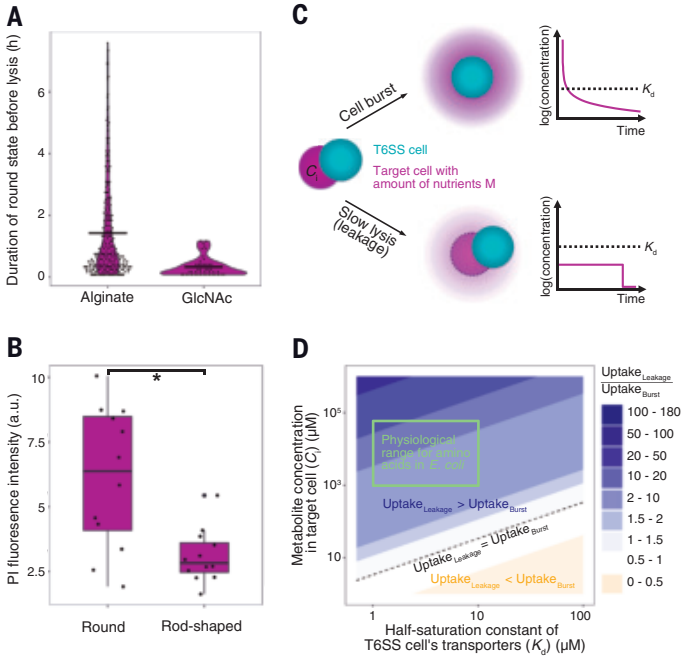


Fig. 2. Slow cell lysis increases the nutrient gain for T6SS cells. (**A**) *V. anguillarum* T6SS cells lyse the *V. cyclitrophicus* target cells slowly, especially under starvation conditions in alginate medium. (**B**) Induction of a round shape in target cells is associated with leakiness of the cell membrane, as estimated by the PI staining intensity of round (length-to-width ratio of 1 to 1.8) and rod-shaped (length-to-width ratio of >1.8 to 6) target cells. Each dot represents the mean value from one microfluidic chamber (from two independent experiments with six chambers each). Jitter was applied in the horizontal direction (n = 12 chambers for round and rod-shaped cells; Wilcoxon signed-rank exact test, $P = 5 \times 10^{-4}$). (**C**) Schematic diagram of the mathematical model comparing the nutrient concentration field available to T6SS cells through fast lysis (top) or slow lysis (bottom), alongside the nutrient release from the target cell over time (right). (**D**) Model predictions comparing the approximated total nutrient uptake by an antagonistic cell through slow and fast cell lysis, suggesting that slow lysis increases nutrient uptake in the physiological parameter range. For slow lysis, we assume that the surface concentration of leaked metabolites is small compared with K_d of the T6SS cell. The green box depicts literature values for C_i and K_d in *E. coli* (table S4).

To test whether slow cell lysis improves nutrient acquisition from target cells, we modeled the nutrient concentration field of a cell that bursts instantaneously compared with that of a cell that releases the same amount of nutrients slowly over time (Fig. 2C). We then modeled the nutrient uptake rate of a neighboring T6SS$^+$ cell using Michaelis-Menten kinetics to take into account that membrane transport can reach saturation. Considering the typical intracellular concentration C_i of amino acids and the half-saturation constant K_d of amino acid transporters of *E. coli* (*21*, *22*), we estimated that the approximate total nutrient gain of T6SS$^+$ cells by slow lysis is 2-fold to 50-fold greater than that through fast lysis (Fig. 2, C and D). Fast lysis led to larger initial concentrations of nutrients but exceeded the uptake capacity of the T6SS$^+$ cell's transporters, thus limiting its uptake. T6SS$^+$ cells would acquire more nutrients from a bursting target cell than from a slowly lysing target cell only if the target cell's C_i was >100-fold smaller or the K_d was >10-fold higher than for amino acids in *E. coli*. These results suggest that T6SS$^+$ cells gain more nutrients from slow cell lysis.

We found that the conversion of biomass from slowly lysing cells into new T6SS$^+$ cells was efficient. In densely populated microfluidic chambers, on average, one new T6SS$^+$ cell formed for every two to three lysing target cells (fig. S6B). This biomass conversion was substantially more efficient than that previously reported for various taxa in well-mixed systems, which required 13 to 2000 heat-killed cells for the growth of one new cell (*23*). This high efficiency may partly be explained by the ~2-fold larger volume of the target cells compared with the T6SS$^+$ cells in our system (fig. S6, C to E). However, it also suggests that T6SS$^+$ cells more efficiently acquire nutrients from slowly lysing cells in sessile microbial communities.

Environmental significance of the use of the T6SS for nutrient foraging

If the T6SS is generally used to generate sources of nutrients by lysing cells, then one would expect to observe a reduction of genes needed for the degradation of complex organic matter and synthesis of cellular building blocks, as observed in pathogenic and predatory bacteria (*24*, *25*). To test this idea, we compared the gene repertoire of T6SS$^+$ and T6SS$^-$ genomes within the genus *Vibrio*. We analyzed ~6000 publicly available high-quality genomes. First, we screened the genomes for the presence of a T6SS (fig. S7) and then de-replicated the genomes into 141 operational taxonomic units (OTUs) (fig. S8).

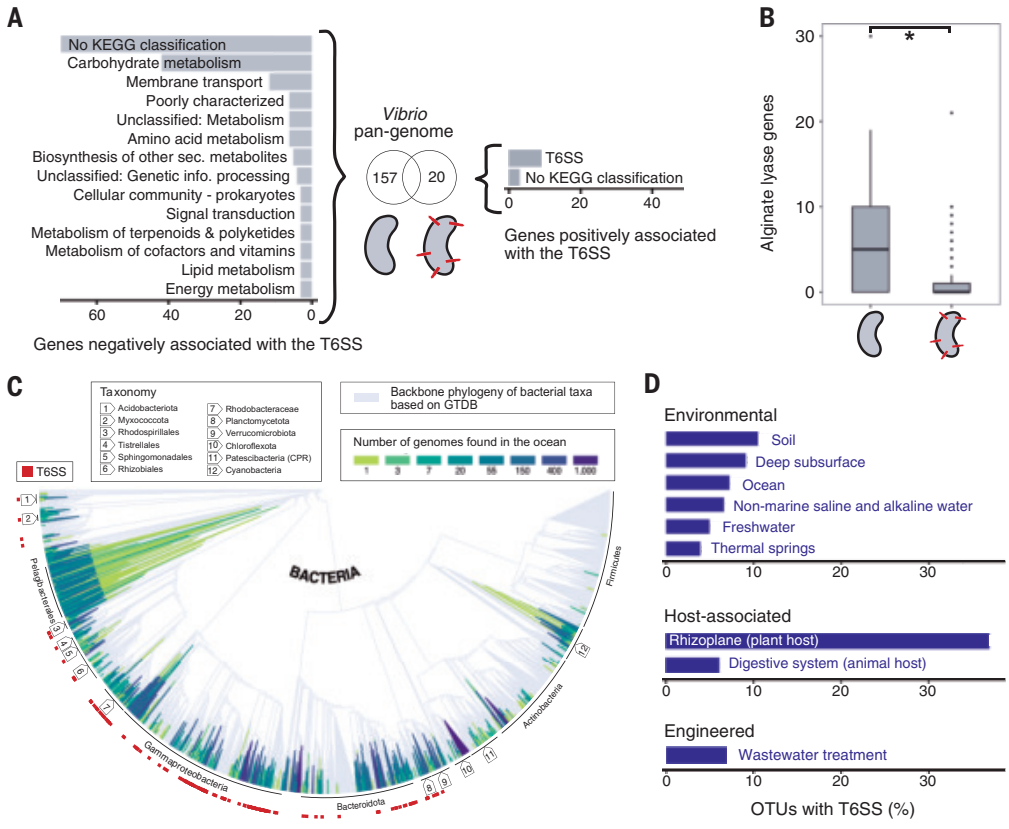


Fig. 3. *Vibrio* with T6SSs have a reduced metabolic gene repertoire, and bacteria with T6SSs are prevalent in natural environments. (**A**) Counts of genes that are positively or negatively associated with the presence of the T6SS in 141 *Vibrio* species-level OTUs, grouped into functional categories according to the Kyoto Encyclopedia of Genes and Genomes (KEGG) classification. A gene can be part of several KEGG classes. Functional categories with fewer than three significant genes are not shown. (**B**) Number of alginate lyase genes across *Vibrio* OTUs with ($n = 82$) and without ($n = 39$) the T6SS (Wilcoxon rank sum test, $P = 1 \times 10^{-5}$). (**C**) Genomes of the OMD (green to blue) placed onto the bacterial Genome Taxonomy Database (GTDB) backbone tree (gray) [modified from Fig. 2C in Paoli *et al*. (*28*)], with red squares indicating the T6SS-encoding genomes. For visualization, the last 15% of the nodes were collapsed (color codes indicate the number of genomes collapsed). (**D**) Prevalence of T6SS-encoding OTUs in different ecosystems based on data from the genomic catalog of Earth's microbiomes (*30*).

We found T6SS$^+$ genomes in 82 OTUs, genomes without T6SS in 39 OTUs, and 20 OTUs that had both T6SS$^+$ and T6SS$^-$ genomes (fig. S9). A pan-genome–wide association study (*26*) on representative genomes of the 82 T6SS$^+$ and 39 T6SS$^-$ OTUs was used to identify T6SS-associated genes. Because many bacterial traits are conserved among related taxa, we removed associations that arose from the underlying phylogenetic structure of the OTUs using a post hoc label-switching permutation test (*26*) (fig. S10). We found 20 genes that had significantly positive and 157 genes that had significantly negative associations with the T6SS. The genes positively associated with the T6SS primarily encode the T6SS itself (Fig. 3A). The genes negatively associated with the T6SS are mainly involved in metabolic processes such as carbohydrate metabolism, metabolite transport, amino acid metabolism, and biosynthesis of secondary metabolites (Fig. 3A). This finding indicates that these metabolic functions are reduced in T6SS$^+$ species. As an example, for carbohydrate metabolism, we found that T6SS$^+$ *Vibrio* genomes contained significantly fewer enzymes to degrade alginate (1.54 ± 0.75 alginate lyases in T6SS$^+$ and 6.03 ± 2.24 in T6SS$^-$ genomes) (Fig. 3B and fig. S11E), and the pan-genomic association study showed that alginate lyases of the PL7, PL12/PL17, and PL15/35 families were significantly negatively associated with the T6SS. The genome size and genome completeness were, however, not significantly different between T6SS$^+$ and T6SS$^-$ OTUs (fig. S11, A to C). Overall, the reduced repertoire of metabolic genes that we observed, particularly for carbohydrate metabolism, in T6SS$^+$ genomes is consistent with genomic adaptation toward acquisition of nutrients and macromolecules from external sources, rather than their biosynthesis from simple metabolites obtained from the decomposition of complex organic matter.

In the ocean, cell lysis induced by viruses is known to be an important contributor to carbon and nutrient cycling, a phenomenon known as the "viral shunt" (*27*). Cell lysis through bacterial antagonism may enhance this flux of carbon from living bacterial cells to lysis products. Therefore, we screened for T6SS genes in the Ocean Microbiomics Database (OMD) (*28*), a consolidated collection of metagenome-assembled genomes from four major global ocean expeditions and two marine time series, as well as in single-cell amplified genomes and marine microbial isolate genomes (fig. S12). In the combined 7610 bacterial species-level OTUs, we found T6SS in 281 OTUs, i.e., in ~3.7%. In 229 of these OTUs, more than half of the OTU members had a T6SS, i.e., in ~3.0% (Fig. 3C). Apart from clades that are known to be rich in T6SS, such as *Gammaproteobacteria* and *Bacteroidota*, we also found that *Planctomycetota* and *Rhodobacteraceae*, which are abundant in the coastal ocean (*29*), frequently encode T6SS. Overall, the prevalence of the T6SS in marine taxa from the global oceans suggests that T6SS-mediated nutrient foraging is common and may thus substantially influence nutrient cycling in the ocean.

A wider screen of natural environments showed that T6SS$^+$ bacteria are common in diverse habitats. We screened for the presence of T6SS genes in a catalog of Earth's microbiomes containing 45,599 species-level OTUs that include metagenome-assembled bacterial genomes from >10,000 samples from diverse habitats (*30*). We found T6SS genes in 4 to 7% of OTUs in aquatic systems, in 9 to 11% of OTUs in terrestrial systems, and in 37% of OTUs in the rhizoplane (Fig. 3D). This is consistent with previous findings suggesting that T6SS$^+$ bacteria are more common among particle-attached than among planktonic bacteria (*13*). In addition, soils have the largest average number of cells per volume of any known ecological environment (*31*, *32*), and the rhizosphere is considered to be the richest niche within this habitat (*5*, *33*), which may promote contact-dependent interactions in this environment. In the rhizoplane, most bacteria are thought to be commensals, but plant pathogens were found that use the T6SS to dominate the root microbiota and increase their virulence against the host plant (*5*). In environments that are comparatively rich in organic biomass, such as the digestive systems of animals and wastewater treatment plants, 6 and 7% of OTUs were T6SS$^+$, respectively. These

findings indicate that the T6SS is widespread across both nutrient-rich and nutrient-poor natural environments.

Discussion

In this work, we found that contact-dependent antagonism can be used by bacteria to access nutrients from cell lysate, which can be rapidly assimilated at low metabolic cost. In contrast to the difficult-to-digest, complex organic compounds prevalent in natural environments, such as polysaccharides, this source contains many readily available cellular building blocks, such as nucleotides and amino acids (25 and 57% of cell content, respectively) (*34*). Directly polymerizing these building blocks into functional cell components represents an energy savings of >90% compared with first having to synthesize them and also reduces proteomic cost (*34*). This mode of nutrient foraging is likely effective in microbial communities with a high frequency of cell-cell contact and limited diffusional loss of released nutrients, such as those in biofilms, soil aggregates, marine snow, the rhizosphere, and the intestinal mucus layer. It is less likely to bring an advantage in planktonic cells, but we have not assessed the efficiency of nutrient foraging in planktonic cells.

Nutrient foraging is an alternative to the mediation of bacterial competition and other functions that have been ascribed to the T6SS, such as acquisition of genetic material (*12*), defense (*35*), and metal scavenging (*36*). These functions are not mutually exclusive, but the genomic signatures of T6SS$^+$ species in *Vibrio* match the expectations that arise from a role in nutrient foraging (*24*, *25*). Nonetheless, it remains unknown whether antagonistic foraging in bacterial communities is the cause of this genomic adaptation or if other factors dominate. For example, some T6SSs are active against eukaryotic hosts (*9*), which may contribute to an association between T6SSs and the reduced metabolic capacity in a host-associated lifestyle. The reduced repertoire of enzymes for carbohydrate degradation renders cells unable to grow on certain carbon sources, such as alginate in the case of *V. anguillarum*. In these conditions, they are not in competition with other cells for the carbon source, but the possession of a T6SS enables them to acquire carbon from nearby cells, thus extending their ecological niche. Functional and genomic analyses of other genera are now necessary to determine whether this function is important in other T6SS-encoding taxa. Under a broad definition of predation, i.e., killing cells for nutrient uptake, this interaction may be classified as facultative predation; however, there is no evidence that T6SS$^+$ cells actively hunt target cells, so this does not meet the stricter definitions of predation (*37*). In contrast to predation through ixotrophy, a recently described hunting strategy found in some filamentous T6SS$^+$ bacterial species of the *Saprospiria* and *Cytophagia* classes with "grappling hooks" to attach to prey cells (*38*), the T6SS$^+$ cells studied here did not grasp target cells.

Because the T6SS is found in diverse taxa and diverse environments, we expect interbacterial cell lysis to also affect large-scale biogeochemical processes. Cell lysis releases both labile dissolved organic matter and carbon-rich particulate organic matter (*27*). The pool of labile substances provided by the lysate can be readily reused by bacteria (*39*), whereas the more recalcitrant particulate organic matter may preferentially sink into deeper ocean layers and thus contribute toward long-term carbon storage in the ocean, as proposed for the "viral shunt" (*27*). Our analyses suggest that there may be a parallel "bacterial shunt" that participates in marine biogeochemical cycling.

REFERENCES AND NOTES

1. J. E. González-Pastor, E. C. Hobbs, R. Losick, *Science* **301**, 510–513 (2003).
2. S. B. Peterson, S. K. Bertolli, J. D. Mougous, *Curr. Biol.* **30**, R1203–R1214 (2020).
3. E. T. Granato, T. A. Meiller-Legrand, K. R. Foster, *Curr. Biol.* **29**, R521–R537 (2019).
4. T. A. Klein, S. Ahmad, J. C. Whitney, *Trends Microbiol.* **28**, 387–400 (2020).
5. R. Gallegos-Monterrosa, S. J. Coulthurst, *FEMS Microbiol. Rev.* **45**, fuab033 (2021).
6. D. L. MacIntyre, S. T. Miyata, M. Kitaoka, S. Pukatzki, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 19520–19524 (2010).
7. M. Basler, B. T. Ho, J. J. Mekalanos, *Cell* **152**, 884–894 (2013).
8. B. W. Anderson, J. D. Wang, *Nature* **575**, 599–600 (2019).
9. S. Coulthurst, *Microbiology* **165**, 503–515 (2019).
10. R. E. Hernandez, R. Gallegos-Monterrosa, S. J. Coulthurst, *Cell. Microbiol.* **22**, e13241 (2020).
11. T. G. Sana, K. A. Lugo, D. M. Monack, *PLOS Pathog.* **13**, e1006325 (2017).
12. S. Borgeaud, L. C. Metzger, T. Scrignari, M. Blokesch, *Science* **347**, 63–67 (2015).
13. M. W. Kempnich, M. P. Sison-Mangus, *PLOS ONE* **15**, e0244217 (2020).
14. F. Boyer, G. Fichant, J. Berthod, Y. Vandenbrouck, I. Attree, *BMC Genomics* **10**, 104 (2009).
15. M. Basler, J. J. Mekalanos, *Science* **337**, 815 (2012).
16. J. P. Schneider *et al.*, *EMBO J.* **38**, e100825 (2019).
17. K. S. Lee *et al.*, *Nat. Rev. Methods Primers* **1**, 80 (2021).
18. T. Dörr, B. M. Davis, M. K. Waldor, *PLOS Pathog.* **11**, e1004850 (2015).
19. L. Boulos, M. Prévost, B. Barbeau, J. Coallier, R. Desjardins, *J. Microbiol. Methods* **37**, 77–86 (1999).
20. C. Kirchhoff, H. Cypionka, *J. Microbiol. Methods* **142**, 79–82 (2017).
21. A. J. Roe, D. McLaggan, I. Davidson, C. O'Byrne, I. R. Booth, *J. Bacteriol.* **180**, 767–772 (1998).
22. J. R. Piperno, D. L. Oxender, *J. Biol. Chem.* **243**, 5914–5920 (1968).
23. I. Nioh, A. Furusaka, *J. Gen. Appl. Microbiol.* **14**, 373–385 (1968).
24. V. Merhej, M. Royer-Carenzi, P. Pontarotti, D. Raoult, *Biol. Direct* **4**, 13 (2009).
25. Z. Pasternak *et al.*, *ISME J.* **7**, 756–769 (2013).
26. O. Brynildsrud, J. Bohlin, L. Scheffer, V. Eldholm, *Genome Biol.* **17**, 238 (2016).
27. C. A. Suttle, *Nat. Rev. Microbiol.* **5**, 801–812 (2007).
28. L. Paoli *et al.*, *Nature* **607**, 111–118 (2022).
29. B. J. Campbell, L. Yu, J. F. Heidelberg, D. L. Kirchman, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 12776–12781 (2011).
30. S. Nayfach *et al.*, *Nat. Biotechnol.* **39**, 499–509 (2021).
31. P. D. Schloss, J. Handelsman, *PLOS Comput. Biol.* **2**, e92 (2006).
32. L. F. W. Roesch *et al.*, *ISME J.* **1**, 283–290 (2007).
33. A. M. Hirsch, *Symbiosis* **37**, 345–363 (2004).
34. M. A. Lever *et al.*, *FEMS Microbiol. Rev.* **39**, 688–728 (2015).
35. S. Pukatzki *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1528–1533 (2006).
36. X. Yang, H. Liu, Y. Zhang, X. Shen, *Front. Microbiol.* **12**, 756136 (2021).
37. J. Pérez, A. Moraleda-Muñoz, F. J. Marcos-Torres, J. Muñoz-Dorado, *Environ. Microbiol.* **18**, 766–779 (2016).
38. Y.-W. Lien *et al.*, *Science* **386**, eadp0614 (2024).
39. I. Beidler *et al.*, *Nat. Commun.* **15**, 4048 (2024).
40. A. K. M. Stubbusch, Code and data for: Antagonism as a foraging strategy in microbial communities, ERIC Open (2025); https://doi.org/10.25678/000D09.
41. A. K. M. Stubbusch, Stable isotope probing Raman microspectroscopy data for: Antagonism as a foraging strategy in microbial communities, MicrobioRaman (2025); https://www.ebi.ac.uk/biostudies/microbioraman/studies/S-MBRS13.

ACKNOWLEDGMENTS

We thank current and previous members of the Schubert and Ackermann and Magnabosco labs, the PriME collaboration, and M. Gabrielli for helpful discussions; H. Wildschutte for providing the antibiotic deletion of *V. anguillarum*; F. Oschmann and J. Fluri for the development of the segmentation and tracking tool *MIDAP*; K. Drescher and N. Netter for advice on single-cell segmentation of the T6SS cells; and R. Naisbit for editorial comments. We declare the use of ChatGPT-4 to inspire specific coding solutions. **Funding:** This work was supported by ETH Zurich and Eawag (M.A., O.T.S.); Simons Foundation grants 542379FY22 and 542395FY22 as part of the Principles of Microbial Ecosystems (PriME) Collaboration (R.S., O.T.S., M.A.) and PriME Opportunity Grant from the PriME Collaboration (G.G.D., J.S.); ETH Zurich (C.M.); ETH fellowship and a Marie Curie Actions for People COFUND program fellowship FEL-3716-1 (G.G.D.); ETH Career SEED grant SEED-14 18-1 (G.G.D.); and Gordon and Betty Moore Foundation Symbiosis in Aquatic Systems Initiative Investigator Award GBMF9197 to R.S.). **Author contributions:** Conceptualization: A.K.M.S., M.A., C.M., G.G.D.; Funding acquisition: M.A., O.T.S., G.G.D., R.S., J.S.; Investigation: A.K.M.S., G.G.D., F.J.P., K.S.L.; Methodology: A.K.M.S., G.G.D., F.J.P., K.S.L., J.S., M.B., L.P.; Project administration: A.K.M.S.; Supervision: M.A., O.T.S., C.M., G.G.D.; Visualization: A.K.M.S.; Writing – original draft: A.K.M.S.; Writing – review & editing: All authors. **Competing interests:** R.S. is also affiliated with PhAST Inc., Zurich. K.S.L. is also affiliated with the Department of Biological Sciences and the Department of Biomedical Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan, South Korea. The remaining authors declare no competing interests. **Data and materials availability:** Code and data to reproduce the results and figures in this paper are available on ERIC Open (*40*). The stable isotope probing Raman microspectroscopy data are also deposited in the MicrobioRaman database (*41*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adr8286
Materials and Methods; Figs. S1 to S12; Tables S1 to S8; References (*42*–*76*); Movies S1 to S6; MDAR Reproducibility Checklist

Submitted 19 November 2024; accepted 17 April 2025

10.1126/science.adr8286

VIROLOGY

Pan-viral ORFs discovery using massively parallel ribosome profiling

Shira Weingarten-Gabbay[1,2,3]*†, Matthew R. Bauer[4]‡, Alexandra C. Stanton[1,5,6], Yingpu Yu[3], Catherine A. Freije[1,5], Nicole L. Welch[1,5], Chloe K. Boehm[1], Susan Klaeger[1], Eva K. Verzani[1], Daniel López[7], Lisa E. Hensley[8], Karl R. Clauser[1], Steven A. Carr[1], Jennifer G. Abelin[1], Charles M. Rice[3], Pardis C. Sabeti[1,2,6,9,10]*

Defining viral proteomes is crucial to understanding viral life cycles and immune recognition but the landscape of translated regions remains unknown for most viruses. We have developed massively parallel ribosome profiling (MPRP) to determine open reading frames (ORFs) across tens of thousands of designed oligonucleotides. MPRP identified 4208 unannotated ORFs in 679 human-associated viral genomes. We found viral peptides originating from detected noncanonical ORFs presented on class-I human leukocyte antigen in infected cells and hundreds of upstream ORFs that likely modulate translation initiation of viral proteins. The discovery of viral ORFs across a wide range of viral families—including highly pathogenic viruses—expands the repertoire of vaccine targets and reveals potential cis-regulatory sequences.

Despite advances in sequencing viral genomes, functional annotations of these genomes have lagged behind. Beyond the annotated canonical open reading frames (ORFs), viral genomes encode noncanonical ORFs that do not fulfill the classical definition of ORFs—i.e., they do not start with ATG nucleotides (nt) and/or are shorter than 100 amino acids (aa) in length. These noncanonical ORFs and the resulting microproteins modulate viral infection (*1*–*3*), contribute to the immune response to viruses (*4*–*8*), and regulate gene expression (*9*–*11*). However, the deviation from classical ORF features complicates their detection by computational approaches and relies mostly on experimental measurements.

The development of ribosome profiling (also termed Ribo-seq) has transformed our ability to detect translated regions across genomes (*12*). Ribosome profiling utilizes deep sequencing of ribosome-bound mRNA fragments to determine ribosome occupancy, indicating translated regions at single nucleotide resolution. It has uncovered many noncanonical ORFs in mammalian cells, yeast, bacteria, and viruses, including upstream ORFs (uORFs) and upstream-overlapping ORFs (uoORFs) in 5′UTRs, short ORFs in noncoding RNAs, and overlapping internal ORFs in annotated coding sequence (iORFs) (*13*, *14*).

The landscape of translated regions, however, is still unknown for most viruses. Ribosome profiling has only been used to profile a handful of viruses as a result of the following challenges: Each virus demands a specific culturing system, some viruses cannot be cultured in the lab, and highly pathogenic viruses require high-containment facilities. Moreover, most viruses are genetically diverse, necessitating a method that can evaluate multiple variants in parallel.

We have leveraged an oligo synthesis library (*15*–*17*) and combined it with ribosome profiling to perform pan-viral ORF discovery. We measured the translation of 20,170 synthetic sequences from 679 viral genomes in two human cell lines, under stress conditions associated with viral infection, and when expressed by cap-dependent or internal ribosome entry site (IRES)-dependent translation. We estimated ORF discovery using the annotated coding sequences (CDSs) and previously reported noncanonical ORFs. We compared ribosome footprints in the synthetic library and natural infection with four viruses. We then examined the function of the detected ORFs in two processes: class-I human leukocyte antigen (HLA-I) presentation and uORF-mediated translation regulation.

Results

Massively parallel ribosome profiling to identify ORFs

We developed massively parallel ribosome profiling (MPRP) to measure translated regions across hundreds of human viruses (Fig. 1A). We used oligonucleotide library synthesis technology to encapsulate thousands of viral sequences in a single pooled experiment. Each oligo contained 200 nt of viral sequence, flanked by constant primers. We cloned the library into an overexpression plasmid downstream of a CMV promoter and upstream of a Woodchuck hepatitis virus post-transcriptional regulatory element used to further enhance expression. To monitor the translation of ORFs in the designed oligos, we excluded ATG start codons on the plasmid. We transfected the pooled library plasmid into HEK293T or A549 cells. We then performed a modified protocol of ribosome profiling (*18*) after treating cells with either cycloheximide (CHX) to inhibit elongating ribosomes or lactimidomycin (LTM) to inhibit initiating ribosomes. We mapped deep sequencing reads representing ribosome footprints to the synthetic library and identified ORFs using probabilistic inference of codon activities by an EM algorithm (PRICE) a computational method for the detection of ORFs in ribosome profiling experiments (*19*).

We tested the quality of ribosome footprints by examining host-mapped reads, showing the expected read length distribution, footprint enrichment in the CDS, and trinucleotide periodicity (fig. S1). To verify that our experimental system captures the translation of an ORF embedded within a synthetic oligo, we performed ribosome profiling on cells transiently transfected with a full-length or truncated green fluorescent protein (fig. S2).

We designed a "pilot" library to estimate viral ORF discovery using MPRP. We tiled 30 mRNAs annotated in the genomic datasets of human cytomegalovirus (HCMV) and herpes simplex virus type 1 (HSV-1) (fig. S3A). We observed the expected low occupancy in the 5′UTR and enrichment of ribosomes at the start codon (fig. S3B). However, we also observed high occupancy in oligos along the CDSs that do not contain the CDS start codon. In the context of the full transcript, these alternative initiation sites are preceded by multiple start codons, making it unlikely that the ribosome will initiate at these positions. We reasoned that a tiling approach may result in false-positive ORF discovery and that MPRP has higher accuracy at the beginning of annotated CDSs and 5′UTRs. Thus, we decided to focus on this region for the design of the pan-viral library, which encompasses the majority of ORF initiation sites (*13*).

We designed a pan-viral library of 15,000 oligos to screen for novel ORFs in the 5′UTRs and the beginning of the CDS of 3976 genes in 679 viral genomes. For each gene, we designed three oligos: a wild-type (WT) oligo containing 60 nt of the 5′UTR and 140 nt of the CDS, a mutated oligo in which the annotated start codon was mutated to GCC nts, and a further-upstream oligo in the 5′UTR region, starting −260 nt relative to the CDS start codon (Fig. 1B).

To gauge the reproducibility of the MPRP measurements, we compared ribosome occupancy on the synthetic oligos in different experiments.

[1]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [2]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA. [3]Laboratory of Virology and Infectious Disease, The Rockefeller University, New York, NY, USA. [4]Harvard Program in Biological and Biomedical Sciences, Division of Medical Sciences, Harvard Medical School, Boston, Massachusetts, USA. [5]Harvard Program in Virology, Harvard Medical School, Boston, MA, USA. [6]Howard Hughes Medical Institute, Chevy Chase, MD, USA. [7]Unidad de Presentación y Regulación Inmunes, Centro Nacional de Microbiología, Instituto de Salud Carlos III, Majadahonda (Madrid), Spain. [8]Zoonotic and Emerging Disease Research Unit, National Bio and Agro-Defense Facility, US Department of Agriculture, Agricultural Research Service (ARS), Manhattan, KS, USA. [9]Massachusetts Consortium on Pathogen Readiness, Boston, MA, USA. [10]Department of Immunology and Infectious Disease, Harvard T.H. Chan School of Public Health, Boston, MA, USA. †Present address: Department of Microbiology, Harvard Medical School, Boston, MA, USA. ‡Present address: Harvard Law School, Harvard University, Cambridge, MA, USA. ***Corresponding author. Email: shira_weingarten@hms.harvard.edu (S.W.-G.); pardis@broadinstitute.org (P.C.S.)**

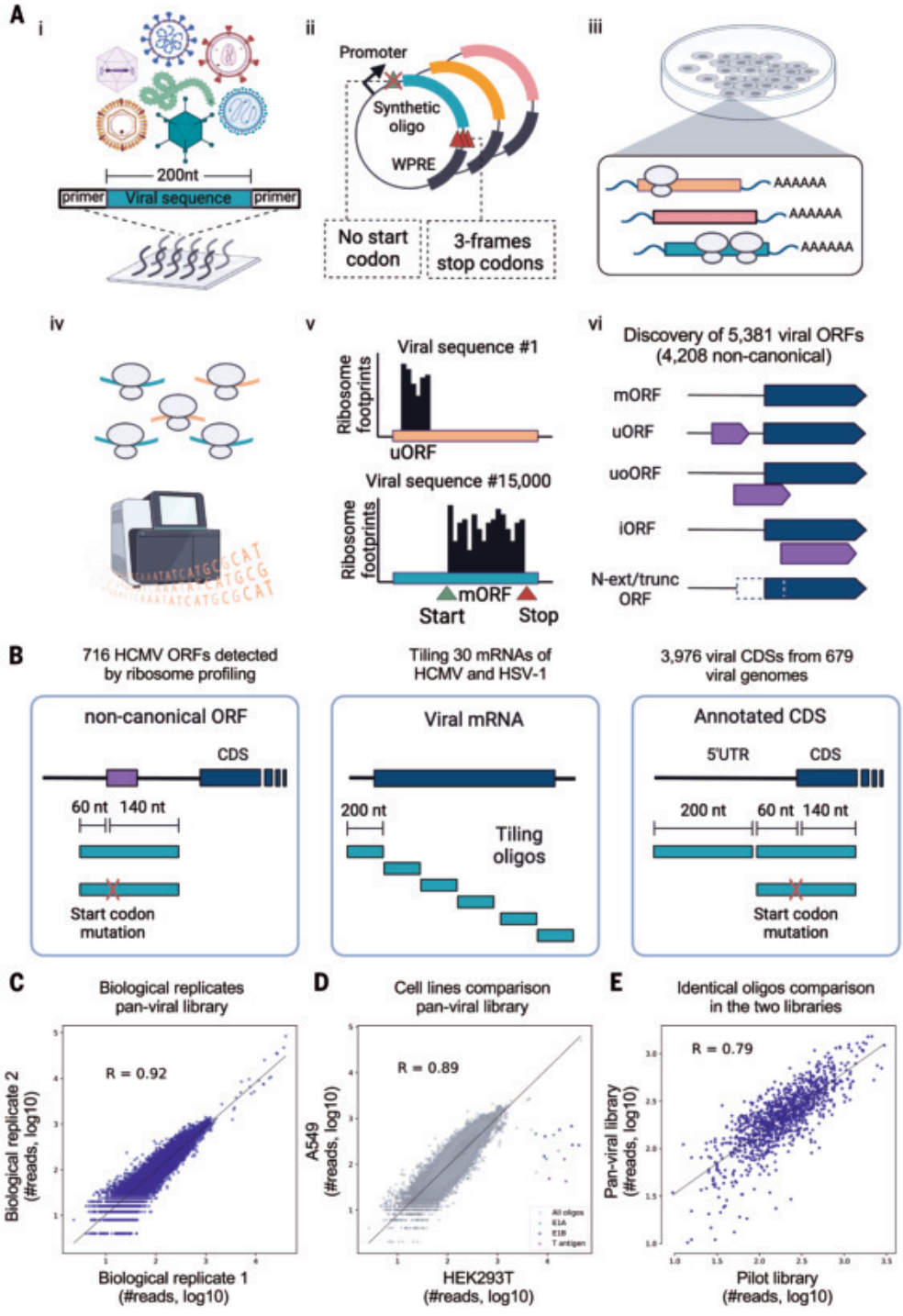


Fig. 1. Design of oligonucleotide synthetic library and MPRP measurements. (**A**) Illustration of the MPRP. (i) Synthetic library amplification using constant primers. (ii) Cloning of library into the overexpression vector. (iii) Transient transfection of plasmid pool into HEK293T or A549 cells for 24 hours. (iv) Treatment of cells with either LTM or CHX and performance of ribosome profiling protocol. (v) Mapping of deep sequencing reads to the synthetic library. (vi) Inferring of translated ORFs using PRICE (*19*). (**B**) Design of the tested synthetic oligonucleotides: (i) ORFs detected previously by ribosome profiling in infected cells with either the intact start codon or a GCC mutation. (ii) Tiling oligos encompassing complete viral transcripts. (iii) Oligos spanning the 5′UTR and the first 140 nt of annotated viral CDSs. For the region containing the CDS, two oligos were designed: the WT sequence and a start-codon–mutated oligo. (**C**) Comparison of the number of ribosome footprints mapped to 15,000 oligos in two biological replicates in HEK293T cells. R = 0.92, Pearson correlation. (**D**) Comparison of the number of ribosome footprints mapped to 15,000 oligos in HEK293T and A549 cells. R = 0.89, Pearson correlation. Oligos with shared sequences with the adenoviral E1A/B genes and the SV40 large T-antigen endogenously expressed by HEK293T cells are shown in green, blue, and pink. (**E**) Comparison of the number of ribosome footprints mapped to 1163 identical oligos in two synthetic libraries in HEK293T cells. R = 0.79, Pearson correlation.

Measurements of the pilot and the pan-viral libraries were consistent between biological replicates in HEK293T cells (R = 0.81 and R = 0.92, respectively; fig. S4 and Fig. 1C). We found high agreement between ribosome occupancies on the pan-viral library in HEK293T and A549 cells (R = 0.89, Fig. 1D). We also designed a set of 1163 oligos with identical sequences in both the pilot and pan-viral libraries and found good agreement between oligos that were independently synthesized, cloned, and measured (R = 0.79, Fig. 1E).

To imitate the cellular environment in infected cells, we repeated MPRP when inducing stress. We treated cells with poly(I:C) to model viruses that form double-strand RNA and Thapsigargin to induce endoplasmic reticulum stress. We confirmed the expected phosphorylation of translation regulators (PKR and eIF2alpha, fig. S5A) and innate immune sensors (IRF3 and STAT1, fig. S5B). Notably, we observed phosphorylation of PKR, eIF2alpha, IRF3, and STAT1 in cells expressing a plasmid without treatment, suggesting that transient transfection by itself induces a similar stress response. Accordingly, we found a significant agreement between MPRP measurements (fig. S5C) in nontreated and poly(I:C)-treated A549 and HEK293T cells (R = 0.773 and R = 0.767, respectively, fig. S5D and Fig. 5E). These results indicate that sequences tested in MPRP are exposed to a cellular environment associated with viral infection. MPRP measurements on the pan-viral library uncovered 5381 ORFs including 4208 noncanonical ORFs (table S1) for further study.

Estimating ORF discovery using annotated viral coding sequences

To estimate the detection of ORFs, we examined the distribution of ribosome footprints across oligos representing the first 140 nt of annotated viral CDSs (Fig. 1B). We performed metagene analysis for all the CDSs from each family by computing the average ribosome footprints in each position relative to the annotated start codon. In most of the 21 viral families tested, we found the expected enrichment of ribosome footprints in the CDS and in the correct reading frame, indicating a robust identification of annotated CDSs in MPRP across different viruses (Fig. 2A and fig. S6).

We next examined ribosome footprints in the presence of an LTM inhibitor,

which inhibits 80S complexes at the initiation site. We found a clear enrichment in the annotated start codons for 19 of the 21 viral families tested (Fig. 2A and fig. S7). Using reverse genetics, we found a significant 3.7-fold reduction in the number of footprints on 3777 oligos in which we mutated the annotated start codon to GCC ($P < 10^{-288}$, Fig. 2B and fig. S8).

We used the annotated CDSs to assess the performance of our computational pipeline for inferring ORFs in the MPRP experiment. When running PRICE, we did not indicate which viral oligos contain annotated CDSs, to evaluate their discovery in an unbiased fashion. PRICE successfully captured 1136 of 3976 viral CDSs initiating at the annotated start codon (28%, accounting for 31% of the total ORFs detected by PRICE, Fig. 2C), with a high correlation between the number of reads in two biological replicates (R = 0.93, fig. S9). The observance of 28% of the annotated CDSs may represent inherent limitations of the MPRP, assaying the translation of viral sequences outside the context of viral infection (see Discussion). Nevertheless, we observe specific enrichment at the annotated start codon (15.3 times higher than the next most abundant position).

Since many viruses rely on IRESs for protein translation, we tested whether the mode of ribosome recruitment affects ORF discovery. We cloned the pan-viral library into a bicistronic plasmid downstream of the EMCV or Polio IRES and repeated MPRP in HEK293T cells (figs. S10A and S11 and tables S2 and S3). We observed the expected enrichment of ribosomes on annotated CDSs and a reduction in start codon mutated oligos when expressing the library from both IRESs (fig. S10B). We found a significant overlap between PRICE-detected CDSs in the cap-dependent and IRES-dependent MPRPs (hypergeometric P-value $< 10^{-20}$ and P-value $< 10^{-8}$ for EMCV and Polio IRES, respectively) and agreement between the number of ribosome footprints on shared ORFs (R = 0.75 and R = 0.83 for EMCV and Polio IRES, respectively, fig. S10, C and D). Finally, we found a similar pattern of ribosome footprints in the cap- and IRES-dependent MPRPs on CDSs of viruses that rely on IRESs for translation initiation from the Flaviviridae and Picornaviridae families (fig. S10, E and F). These results suggest that at least for some proteins, ORF discovery was independent of the mechanism by which the ribosome was recruited to the transcript.

We then compared the pattern of ribosome footprints in MPRP and native viral infection. We performed ribosome profiling in cells infected with vesicular stomatitis virus [(VSV) Indiana strain], influenza A virus [(IAV) Puerto Rico 8 H1N1 strain], and Hepatitis C Virus [(HCV) genotype 2] (fig. S12). We found a high congruence between the location of footprints originating from MPRP and infected cells in CDSs from VSV (Fig. 2D and fig. S13), IAV (Fig. 2E and fig. S14), and HCV (fig. S15).

Noncanonical ORFs discovery

We estimated the discovery of noncanonical ORFs. As part of the pilot library, we included oligos comprising 60 nt upstream and 140 nt downstream of the initiation codons of 716 ORFs identified in HCMV-infected cells by ribosome profiling (*20*), termed here "Ribo-seq ORFs" (Fig. 1B). For each ORF, we also designed a mutated oligo in which we replaced the reported start codon with GCC.

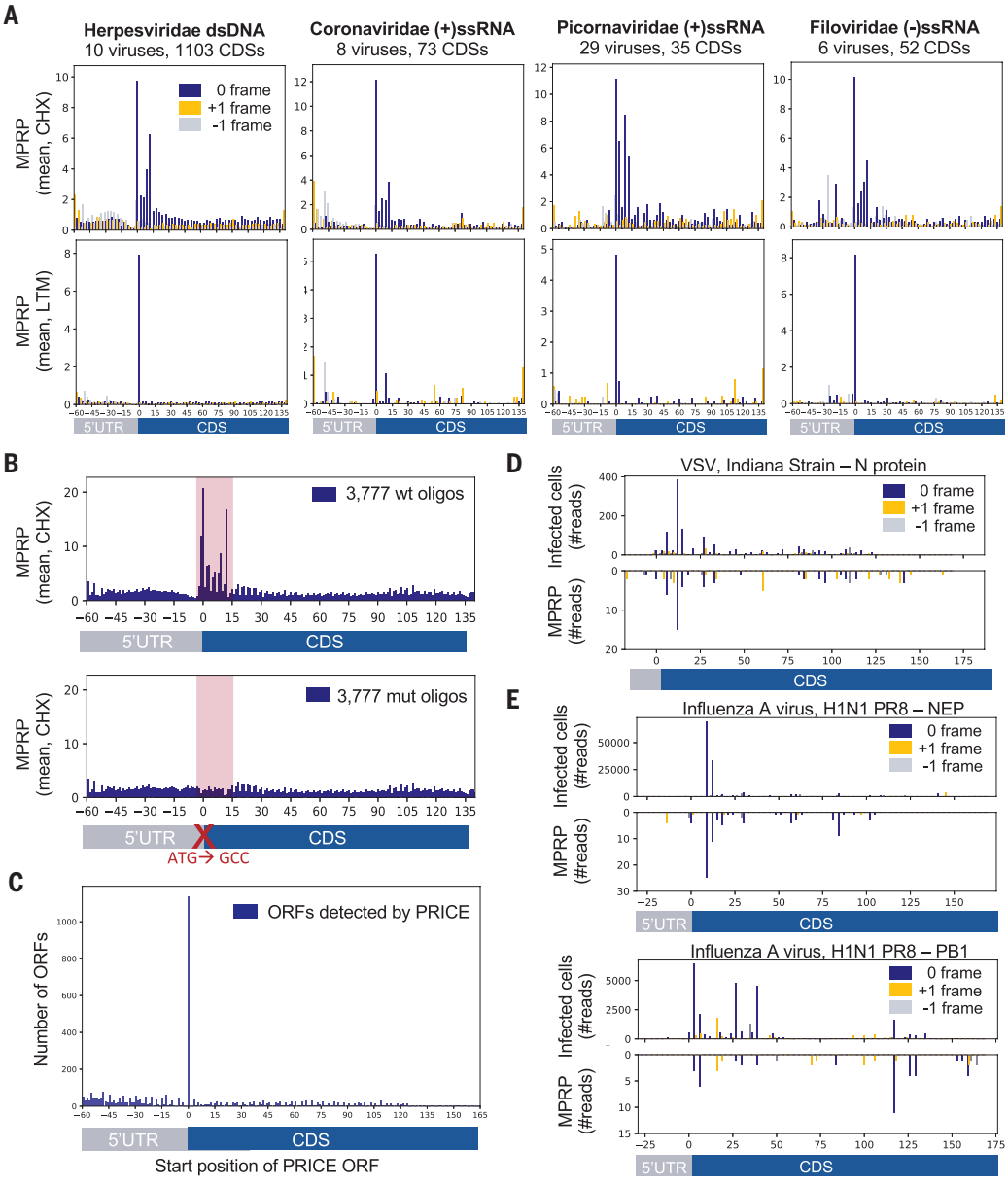


Fig 2. Annotated CDSs measurements in MPRP and infected cells. (**A**) Metagene analysis for four viral families with CHX and LTM inhibitors showing the average ribosome footprints in each position. See figs S6 and S7 for 21 viral families. (**B**) Comparison of the average number of ribosome footprints between oligos containing the WT start codon (upper graph) with those in which the annotated start codon was mutated to GCC (lower graph). Shown are the average ribosome footprints in each position across 3777 oligos. (**C**) ORF discovery using PRICE. Number of ORFs detected in each position (ORF start position). (**D**) Mirror plot showing the number of ribosome footprints mapped to the first 200 nt of the Nucleocapsid transcript in VSV-infected cells and MPRP. (**E**) Similar to (D) for IAV NEP and PB1 transcripts.

We found clear enrichment of ribosome footprints along the reported Ribo-seq ORFs with trinucleotide periodicity indicating translation in the correct reading frame. Notably, MPRP detected ribosome footprints on both canonical and noncanonical Ribo-seq ORFs, including ORFs with a non-AUG start codon and short ORFs in the length of 20 aa or less (Fig. 3, A to C). Mutating the start codon to GCC of 284 Ribo-seq ORFs resulted in a substantial reduction of ribosome footprints compared with WT oligos (Fig. 3D). Our findings confirmed the non-AUG start codons reported by Stern-Ginossar *et al.*, (*20*) and demonstrate how MPRP can be used to functionally characterize the initiation site of noncanonical ORFs. PRICE ORF prediction detected 152 of the 716 Ribo-seq ORFs (21%, accounting for 25% of the total ORFs detected by PRICE) with 11.7-fold enriched initiation at the reported start codon compared to the next abundant position (Fig. 3E).

We found similar ribosome footprints on PRICE-predicted uORFs and iORFs in MPRP and cells infected with HCMV, VSV, and IAV (Fig. 3, F to G, and fig. S16). Consistent with HCMV-infected cells, MPRP detected a prominent peak at the end of an uORF in the *UL4* 5′UTR. This uORF was shown to inhibit the translation of the main CDS by causing ribosome stalling (*21*) (Fig. 3F), indicating that MPRP can provide information on functional stalling sites. Ribosome profiling of cells infected with the H1N1 strain of IAV confirms an iORF in the +1 frame of M1 detected by MPRP (Fig. 3, G to H). Notably, the design of the pan-viral library allowed us to uncover this iORF in six additional influenza strains, including the H5N1 bird flu strain responsible for the current cattle outbreak (Fig. 3I).

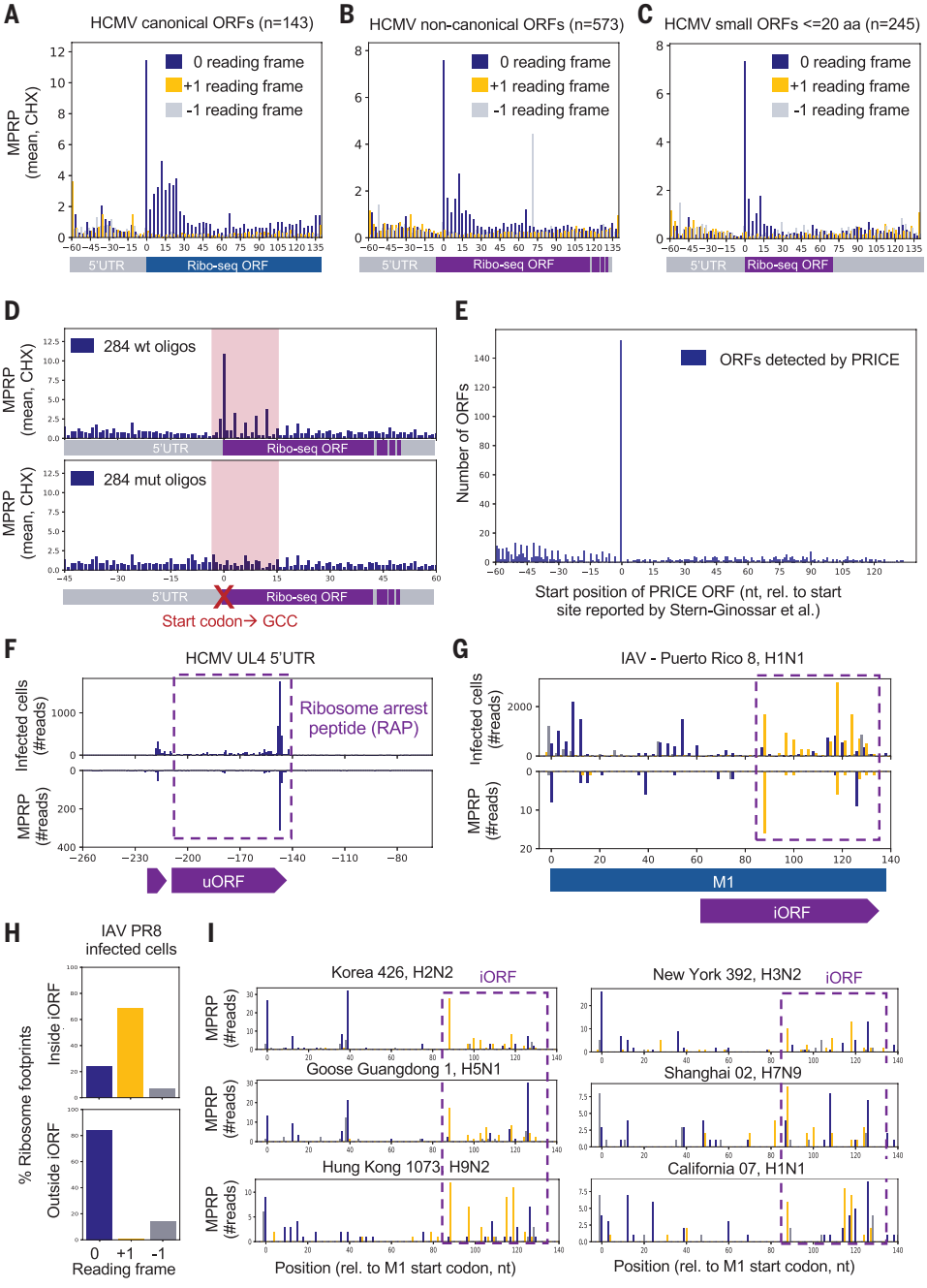


Fig 3. Noncanonical ORF measurements in MPRP and infected cells. (**A** to **C**) Metagene analysis of oligos containing the sequence of ORFs that were identified by ribosome profiling of HCMV-infected cells (*20*). The average of ribosome footprints in each position. (**D**) Comparison of the average number of ribosome footprints between oligos containing the WT start codon (upper graph) to those in which the reported start codon was mutated to GCC (lower graph). Shown are the average ribosome footprints across 284 oligos, containing Ribo-seq ORFs with lengths of 7 to 45 aa. (**E**) ORF discovery using PRICE. Number of ORFs detected in each position (ORF start position). (**F**) Mirror plot showing the number of ribosome footprints in HCMV-infected cells and MPRP. The purple box highlights uORF2, which encodes a ribosome arrest peptide (*21*) (**G**) Mirror plot showing the number of ribosome footprints in IAV-infected cells and MPRP. The purple box highlights an internal overlapping iORF in the +1 reading frame. (**H**) Percentages of ribosome footprints mapped to 0, +1, and −1 reading frames in the region encoding the internal ORF (upper panel) and outside this region (lower panel). (**I**) Ribosome footprints on the M1 coding sequences of six additional IAV strains from the MPRP experiment.

Expanding the repertoire of viral antigens

Noncanonical ORFs contribute to the pool of peptides presented on the HLA-I complex to cytotoxic T cells (*5*, *6*, *7*, *22*–*28*). Notably, peptides from noncanonical ORFs can be enriched on the HLA-I complex in comparison to canonical peptides, with some eliciting stronger T cell responses (*7*, *26*). Thus, noncanonical ORFs discovery can provide new targets for vaccines and insights on the interaction between viruses and the immune system.

To assess the contribution of noncanonical ORFs detected by MPRP to HLA-I presentation, we reanalyzed two immunopeptidome datasets from cells infected with either HCMV (*19*) or vaccinia virus (VACV) (*29*) (fig. S17). We found five distinct peptides from four noncanonical ORFs in HCMV: an uORF in the 5′UTR of UL4 (VLSAKKLS, and VLSAKKLSSL), an uORF in the 5′UTR of UL148 (FAKSKTIGL), an uoORF in the 5′UTR of UL135 (YPAPRPQAI), and an N-terminal extended isoform of the UL36 protein (VMDDLRDTL) (Fig. 4A). In VACV-infected cells, we found two distinct HLA-I peptides from an ORF located upstream to I7L (ILFFHVLLY) and from an iORF overlapping the coding region of L3L (HRNKIINAEK) (Fig. 4B).

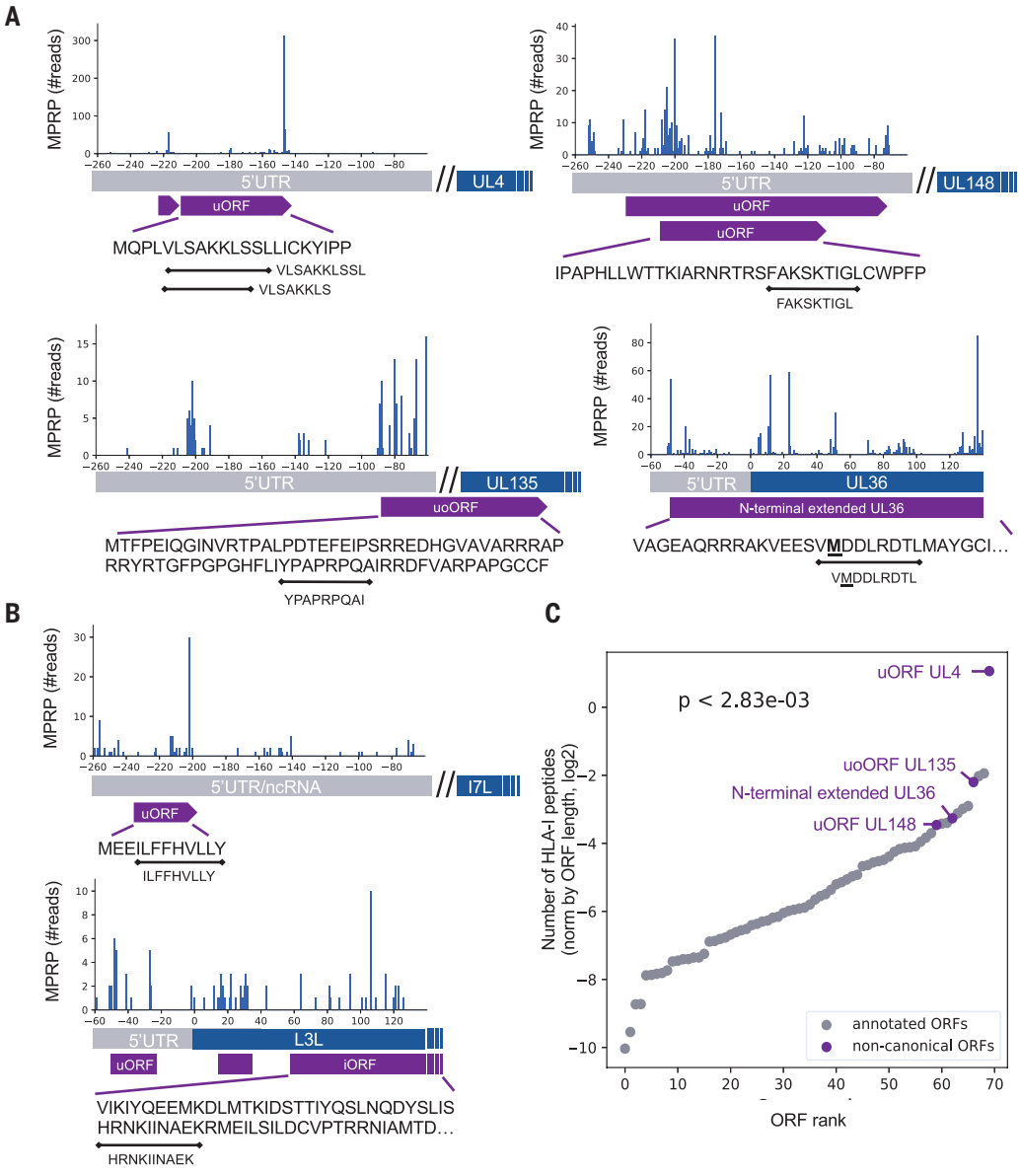


Fig 4. HLA-I peptides derived from noncanonical ORFs in HCMV and VACV. (**A**) HLA-I peptides detected in four noncanonical ORFs of HCMV identified by MPRP. (**B**) HLA-I peptides originating from two noncanonical ORFs in VACV. (**C**) Comparison of HLA-I presentation from annotated and noncanonical ORFs in HCMV. For each ORF, we present the number of total HLA-I peptides detected in HCMV immunopeptidome (not unique) normed by the ORF length. $P < 10^{-3}$, Wilcoxon rank-sum test.

Of the seven detected peptides, six were predicted to be good binders by HLAthena (*30*) (MSi rank ≤ 2) to at least one of the expressed HLA-I alleles, and all the noncanonical ORFs were supported by a peptide with a good prediction score (table S4).

Next, we evaluated the contribution of the noncanonical ORFs to HLA-I presentation. Appending MPRP-detected ORFs to the annotated HCMV ORFs increased the number of mapped HLA-I peptides by 7.4% (from 964 to 1035 peptides). Moreover, the noncanonical ORFs produced more HLA-I peptides, compared with most of the annotated ORFs (Wilcoxon rank-sum *P*-value $<10^{-3}$, Fig. 4C).

Exposing uORFs in viral 5′UTRs

uORFs can modulate the expression of viral proteins by attenuating translation initiation at the main CDS (*9*–*11*, *31*). Examining the distribution of ribosome footprints in 2418 viral genes, we identified two main clusters: a group of genes in which most of the footprints were observed in the 5′UTR with low occupancy in the CDS region (5′UTR cluster), and a group of genes in which most of the footprints were detected in the CDS with low occupancy in the 5′UTR (CDS cluster) (Fig 5, A and B). The trinucleotide periodicity observed in uORFs detected in the 5′UTR region indicates that they are actively translated by ribosomes (Fig 5C). Moreover, we found a strong signal of initiating ribosomes at the start codons of these uORFs in the presence of an LTM inhibitor.

We hypothesized that genes in the 5′UTR cluster have uORFs that attenuate translation from the main CDS. uORFs translation response to cellular stress, including viral infection, through the phosphorylation of eIF2alpha (fig. S18). Upon eIF2alpha phosphorylation, preinitiation complexes (PICs) are more likely to scan through the uORF start codon and initiate translation at the main CDS. To test if the detected uORFs are regulated by cellular stress, we treated cells with sodium arsenite, a potent inducer of eIF2alpha phosphorylation (*32*). We confirmed the increase in eIF2alpha phosphorylation and the uORF-regulated ATF4 protein (Fig. 5D).

We performed MPRP in HEK293T cells treated with sodium-arsenite. We found a relative decrease in the fraction of viral genes in which ribosomes were "held" at the 5′UTR (37% to 19% in untreated and treated cells, respectively, Fig. 5E) and an increase in the fraction of genes containing ribosome footprints in the CDS (63 to 81% in untreated and treated cells, respectively). This result indicates that in response to eIF2alpha phosphorylation, ribosomes were more likely to bypass the uORFs in the 5′UTR and initiate translation at the main CDS, as expected in the case of inhibitory uORFs.

Discussion

We present a method to comprehensively screen sequences of many viruses for translated regions in a single pooled experiment. Using MPRP, we uncovered thousands of potential ORFs and provided high resolution of ribosome footprints across the 5′UTR region and the beginning of the CDS in thousands of viral genes.

Although synthetic systems can never fully recapitulate naturally infected cells, MPRP bridges the gap between current computational annotations, which rely on outdated assumptions of ORF characteristics, and traditional ribosome profiling. This is critical given that the landscape of translated regions has not yet been determined for the majority of viruses. MPRP has the capacity to identify translated ORFs in a

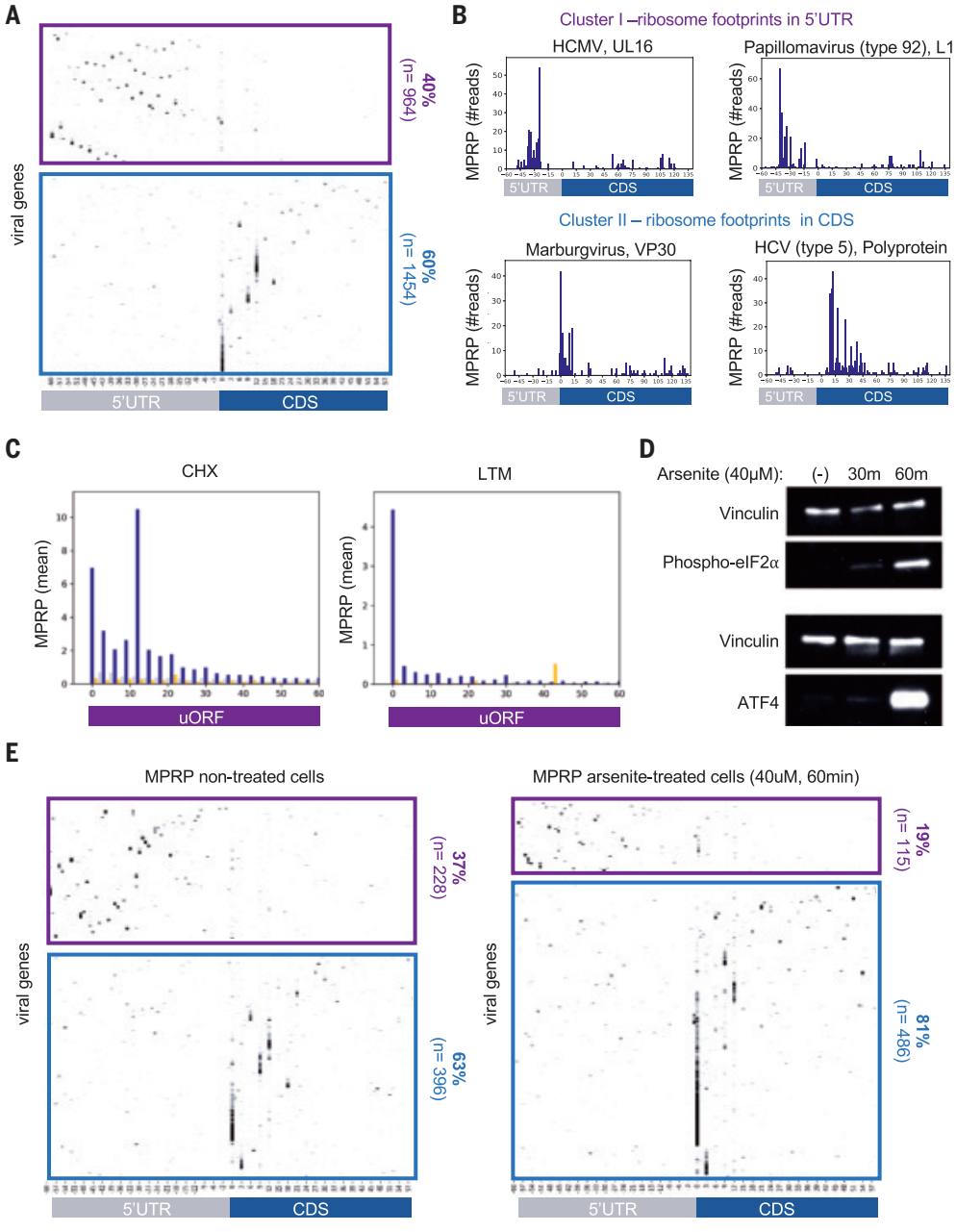


Fig 5. Ribosome densities on uORFs and CDSs in response to eIF2alpha phosphorylation. (**A**) Heatmap showing ribosome footprint densities across 2418 viral oligos. Each line represents a single viral gene and each column represents the position relative to the annotated start codon. Genes in the upper cluster (purple) had more footprints in the 5′UTR region than the CDS region and genes in the lower cluster (blue) had more footprints in the CDS than the 5′UTR. (**B**) Example of two individual genes from each cluster and the distribution of ribosome footprints observed in each position. (**C**) Metagene analysis showing the average ribosome footprints in each position along uORFs detected by PRICE, relative to the uORF start position. Shown for CHX (left) and LTM (right) inhibitors. (**D**) Western blot analysis of lysates from HEK293T cells treated with 40 uM sodium-arsenite for 30 and 60 min. Phosphorylated eIF2alpha was detected with a monoclonal phospho S51 antibody (upper panel). ATF4 protein was detected using a polyclonal antibody (lower panel). In both membranes, Vinculin was used as a loading control. (**E**) Repeat of the MPRP experiment in HEK293T cells that were treated with 40 uM sodium arsenite and in nontreated cells. Shown are heatmaps of ribosome densities across viral oligos and clusters similarly to the analysis in (A).

broad spectrum of viruses and provides a tool for investigating highly pathogenic viruses. Using 200-nt-long viral fragments omits the requirement for scarce high-containment facilities. Additionally, we demonstrated how MPRP can provide rapid insights during an outbreak by exposing a noncanonical iORF in the H5N1 bird flu genome. Within a few weeks, MPRP can detect ORFs in a newly discovered virus, independently of its culturing conditions.

We exposed additional sources of viral antigens that could contribute to HLA-I presentation and T cell recognition. While T cell assays almost exclusively assess responses against canonical proteins, some T cell epitopes from noncanonical ORFs induce more potent T cell responses compared with canonical epitopes (*7*). Thus, the incorporation of noncanonical ORFs into T cell assays has the potential to enhance their sensitivity and facilitate the identification of vaccine targets.

We also discovered viral uORFs that likely affect gene expression regulation. We found numerous potential uORFs that exhibited hallmarks of translation and eIF2alpha phosphorylation responsiveness. These uORFs might function in the temporal regulation of viral proteins as suggested for HHV-6 and HCMV (*33*). Moreover, MPRP can provide insights into the potential mechanism by which uORFs exert their function, such as ribosome stalling (*34*).

It is important to acknowledge the limitations inherent in this study. The viral sequences examined here were assessed independently of the broader genome context and were evaluated in noninfected cells. The 200-nt synthetic oligo might exclude cis-regulatory elements such as IRESs (*35*), VPg proteins (*36*), uORFs in long 5′UTRs (*37*), ribosome shunting (*38*), microRNA binding sites (*39*), and pseudoknots (*40*). MPRP does not capture the distinctive biology occurring within cells infected by each of the many viruses studied here and therefore lacks host and/or viral proteins that regulate translation, such as SARS-CoV-2 nsp1 (*41*). Hence, MPRP cannot accurately detect every ORF in each virus.

Nevertheless, we have substantive evidence supporting the identification of genuine ORFs. This evidence includes elongating and initiating ribosomes on annotated and reported noncanonical ORFs, the reproducibility of measurements across different cell types, the notable decrease in ribosome footprints upon mutating AUG and non-AUG start codons, the high similarity of ribosome footprints pattern in MPRP and native virus infection, the corroborative support from HLA-I peptides identified through mass spectrometry, and the responsiveness of uORFs to stress conditions.

In total, our study yields thousands of candidates for unexplored ORFs across hundreds of human viruses, which can enhance our understanding of viral biology and contribute to vaccine development.

REFERENCES AND NOTES

1. V. Lulla *et al.*, *Nat. Microbiol.* **4**, 280–292 (2019).
2. P. J. Ogden, E. D. Kelsic, S. Sinai, G. M. Church, *Science* **366**, 1139–1143 (2019).
3. V. Lulla, A. E. Firth, *Nat. Commun.* **11**, 4070 (2020).
4. A. Bansal *et al.*, *J. Exp. Med.* **207**, 51–59 (2010).
5. N. T. Ingolia *et al.*, *Cell Rep.* **8**, 1365–1379 (2014).
6. N. Yang *et al.*, *J. Immunol.* **196**, 3608–3617 (2016).
7. S. Weingarten-Gabbay *et al.*, *Cell* **184**, 3962–3980.e17 (2021).
8. S. Weingarten-Gabbay *et al.*, *Cell Rep.* **43**, 113596 (2024).
9. A. Chen, Y. F. Kao, C. M. Brown, *Nucleic Acids Res.* **33**, 1169–1181 (2005).
10. P. S. Gould, A. J. Easton, *J. Biol. Chem.* **280**, 21972–21980 (2005).
11. R. S. Shabman *et al.*, *PLOS Pathog.* **9**, e1003147 (2013).
12. N. T. Ingolia, S. Ghaemmaghami, J. R. S. Newman, J. S. Weissman, *Science* **324**, 218–223 (2009).
13. N. T. Ingolia, L. F. Lareau, J. S. Weissman, *Cell* **147**, 789–802 (2011).
14. N. T. Ingolia, J. A. Hussmann, J. S. Weissman, *Cold Spring Harb. Perspect. Biol.* **11**, a032698 (2019).
15. S. Weingarten-Gabbay *et al.*, *Science* **351**, aad4939 (2016).
16. S. Weingarten-Gabbay *et al.*, *Genome Res.* **29**, 171–183 (2019).
17. J. J. Seo, S.-J. Jung, J. Yang, D.-E. Choi, V. N. Kim, *Cell* **186**, 3291–3306.e21 (2023).
18. N. J. McGlincy, N. T. Ingolia, *Methods* **126**, 112–129 (2017).
19. F. Erhard *et al.*, *Nat. Methods* **15**, 363–366 (2018).
20. N. Stern-Ginossar *et al.*, *Science* **338**, 1088–1093 (2012).
21. K. Ito, S. Chiba, *Annu. Rev. Biochem.* **82**, 171–202 (2013).
22. J. Chen *et al.*, *Science* **367**, 1140–1146 (2020).
23. H. D. Hickman *et al.*, *J. Immunol.* **201**, 1222–1228 (2018).
24. N. J. Maness *et al.*, *J. Virol.* **84**, 11569–11574 (2010).
25. T. Ouspenskaia *et al.*, *Nat. Biotechnol.* **40**, 209–217 (2022).
26. M. V. Ruiz Cuevas *et al.*, *Cell Rep.* **34**, 108815 (2021).
27. S. R. Starck, N. Shastri, *Immunol. Rev.* **272**, 8–16 (2016).
28. J. Holly, J. W. Yewdell, *Curr. Opin. Immunol.* **83**, 102342 (2023).
29. E. Lorente *et al.*, *PLOS Negl. Trop. Dis.* **13**, e0007547 (2019).
30. S. Sarkizova *et al.*, *Nat. Biotechnol.* **38**, 199–209 (2020).
31. A. G. Hinnebusch, I. P. Ivanov, N. Sonenberg, *Science* **352**, 1413–1416 (2016).
32. D. E. Andreev *et al.*, *eLife* **4**, e03971 (2015).
33. Y. Finkel *et al.*, *eLife* **9**, e50960 (2020).
34. T. E. Dever, I. P. Ivanov, M. S. Sachs, *Annu. Rev. Genet.* **54**, 237–264 (2020).
35. N. Stern-Ginossar, S. R. Thompson, M. B. Mathews, I. Mohr, *Cold Spring Harb. Perspect. Biol.* **11**, a033001 (2019).
36. I. Goodfellow, *Curr. Opin. Virol.* **1**, 355–362 (2011).
37. A. G. Hinnebusch, *Annu. Rev. Microbiol.* **59**, 407–450 (2005).
38. L. A. Ryabova, M. M. Pooggin, T. Hohn, *Prog. Nucleic Acid Res. Mol. Biol.* **72**, 1–39 (2002).
39. P. Schult *et al.*, *Nat. Commun.* **9**, 2613 (2018).
40. I. Brierley, S. Pennell, R. J. C. Gilbert, *Nat. Rev. Microbiol.* **5**, 598–610 (2007).
41. K. Schubert *et al.*, *Nat. Struct. Mol. Biol.* **27**, 959–966 (2020).

ACKNOWLEDGMENTS

We thank T. Ouspenskaia, A. Nachshon, S. Reilly, J. Xue, A. Lin, H. Metsky, A. Lercher, and O. Mizrahi for many valuable discussions. We thank N. McGlincy and N. Ingolia for sharing their detailed ribosome profiling protocol with the broad community of researchers (*18*), which contributed to the development of MPRP. **Funding:** This study was supported in part by grants from the National Institute of Health (NIH) (U19AI110818 to P.C.S., P01CA206978 to S.A.C., and R01CA057973 to C.M.R.), the United States Department of Agriculture (58-3022-2-031 to P.C.S.), National Cancer Institute (NCI) Clinical Proteomic Tumor Analysis Consortium program (U24CA270823 and U01CA271402 to S.A.C.), and the Dr. Miriam and Sheldon G. Adelson Medical Research Foundation (to S.A.C.). S.W.-G. is the recipient of a Human Frontier Science Program fellowship (LT-000396/2018), EMBO nonstipendiary long-term fellowship (ALTF 883-2017), the Gruss-Lipper postdoctoral fellowship, the Zuckerman STEM Leadership Program fellowship, and the Rothschild Postdoctoral Fellowship. **Author contributions:** S.W.-G. conceptualized the study, designed the synthetic libraries, designed the experiments, performed experiments, analyzed the data, and wrote the manuscript. M.R.B., A.C.S., Y.Y., C.A.F., N.L.W., and C.K.B. performed the experiments. S.K., E.K.V., D.L., K.R.C., S.A.C., and J.G.A. analyzed the mass spectrometry data of HLA-I immunopeptidomes. L.E.H. and C.M.R. contributed to data interpretation. P.C.S. supervised the work and wrote the manuscript. **Competing interests:** S.W.-G., M.R.B., A.C.S., and P.C.S. are named as co-inventors on International Patent Application PCT/US2024/048478, claiming priority to US Provisional Application 63/540,279 related to this work filed by The Broad Institute that covers massively parallel methods and techniques for evaluating open reading frames in genomes, particularly viral genomes. These methods and data unveil new proteins, immune targets, and cis-regulatory elements that can serve in the design of vaccines and commercial overexpression platforms. M.R.B. is an employee of BioNTech SE. S.K. is an employee of Genentech. S.A.C. is a member of the scientific advisory boards of Kymera, PTM BioLabs, Seer, and PrognomIQ. J.G.A. is a paid consultant for Enara Bio and Moderna. P.C.S. is a cofounder of and consultant to Sherlock Biosciences and Delve Biosciences, is on the Board of Directors and a shareholder of Danaher Corporation and Polaris Genomics, and holds equity in all the companies. All other authors declare no competing interests. **Data and materials availability:** The raw sequencing data generated in this study have been submitted to the Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE272406.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ado6670
Materials and Methods; Figs. S1 to S18; References (*42*–*49*); Data S1 to S4; MDAR Reproducibility Checklist

Submitted 15 February 2024; accepted 10 March 2025

10.1126/science.ado6670

MOLECULAR BIOLOGY

RNA transcripts regulate G-quadruplex landscapes through G-loop formation

Koichi Sato[1,2]*, Jing Lyu[3,4], Jeroen van den Berg[1], Diana Braat[1,2], Victoria M. Cruz[1,2], Carmen Navarro Luzón[3,4], Joost Schimmel[2], Clara Esteban-Jurado[2], Maëlys Alemany[1], Jan Dreyer[5], Aiko Hendrikx[1], Francesca Mattiroli[5], Alexander van Oudenaarden[1], Marcel Tijsterman[2,6], Simon J. Elsässer[3,4], Puck Knipscheer[1,2]*

G-quadruplexes (G4s) are prevalent DNA structures that regulate transcription but also threaten genome stability. How G4 dynamics are controlled remains poorly understood. Here, we report that RNA transcripts govern G4 landscapes through coordinated G-loop assembly and disassembly. G-loop assembly involves activation of the ATM and ATR kinases, followed by homology-directed invasion of RNA opposite the G4 strand mediated by BRCA2 and RAD51. Disassembly of the G-loop resolves the G4 structure through DHX36-FANCJ–mediated G4 unwinding, which triggers nucleolytic incision and subsequent hybrid strand renewal by DNA synthesis. Inhibition of G-loop disassembly causes global G4 and R-loop accumulation, leading to transcriptome dysregulation, replication stress, and genome instability. These findings establish an intricate G-loop assembly-disassembly mechanism that controls G4 landscapes and is essential for cellular homeostasis and survival.

Eukaryotic transcription is a highly regulated process that is required to ensure cell identity. G-quadruplexes (G4s), four-stranded alternative DNA structures formed at G-rich genomic sequences, are thought to be transcriptional regulators enriched in active regulatory regions (*1*). Although mammalian genomes contain more than half a million DNA sequences that can adopt a G4 structure (potential G4-forming sequences, PQSs) (*2*), only a limited number of G4 structures form, and specific G4 subsets have been detected in different cell types (*3*, *4*). Misregulation of G4 positioning disrupts gene expression and embryonic differentiation (*5*, *6*) and has been linked to neurodegenerative disease, cancer, and accelerated aging (*7*, *8*). Furthermore, spontaneous G4s can arise during transcription and DNA replication (*9*, *10*). Such inadvertent G4s, as well as persistent regulatory G4s, cause DNA damage and are associated with genomic deletions and rearrangements (*11*–*13*), further underscoring the importance of G4 structure regulation.

Currently, we lack knowledge on how G4s are controlled to prevent transcription dysregulation and genome instability. G4s are often found in proximity to R-loops, three-stranded structures with an RNA-DNA hybrid and a displaced DNA strand (*14*). These R-loops have been proposed to reinforce the transcriptional regulatory functions of G4s by stabilizing the structure through an RNA-DNA hybrid opposite the G4 strand (*15*). However, this structure, called a G-loop, has not been verified directly beyond bacteria (*9*, *16*). By contrast, R-loops are also induced in the vicinity of ectopically induced G4s through unknown mechanisms and result in the rapid resolution of both structures (*17*). How proximal R-loops control the fate of G4s and whether G-loops are required for this process remain unknown.

G-loop assembly through RNA invasion

To explore the mechanism that couples G4s and R-loops, we used *Xenopus laevis* egg extracts that recapitulate physiological DNA replication, repair, and stress responses (*18*). In cells, the number of PQSs vastly exceeds the number of G4 structures, most likely because G4 formation requires temporal duplex DNA unwinding (*19*). We thus generated a pG4^{BOT} plasmid that contains a canonical G4 structure (G4^{G3N}) opposite a G4-stabilizing noncomplementary sequence, resembling a physiological G4-forming site (fig. S1, A and B). This template was incubated in nucleoplasmic extract (NPE) (*20*), isolated, and subjected to DNA-RNA immunoprecipitation and quantitative polymerase chain reaction (DRIP-qPCR) (fig. S1C). A marked R-loop signal was observed, which peaked at 60 min and declined to the basal level by 120 min, comparable to the half-life time of R-loops in human cells (*21*) (Fig. 1A). Treatment of products with RNaseH, which degrades RNA-DNA hybrids, abolished R-loop signals, validating the signal specificity (fig. S1, D and E). The R-loop signal was lost when the G4 motif was absent (pPolyT) or situated in a duplex region that prevented G4 structure formation (pdsG4^{BOT}) (Fig. 1A and fig. S1F). Moreover, placing a G4 structure on both strands (pG4^{DBL}) prevented R-loop formation, indicating RNA-DNA hybrid formation on the unstructured strand across from the G4 to form a G-loop. R-loop accumulation at the G4 region was also observed for a G4 structure from the telomeric sequence pTEL^{BOT} (fig. S1, A, B, and G to I). This telomeric G4 is less thermodynamically stable compared with G4^{G3N} (*22*), indicating that G-loop assembly is a general response to G4 structure formation regardless of G4 stability. The invading RNA is derived from oocyte transcripts (fig. S1J), because pretreatment of the extract with RNaseA prevented R-loop accumulation, whereas transcription inhibition by RNA polymerase II depletion and α-amanitin treatment had no effect (Fig. 1B and fig. S1, K and L). Chromatin association of RNA can be posttranscriptionally regulated by ribonucleoproteins (*23*, *24*). Consistent with this, depletion of the hnRNPA1 complex, a complex interacting with RNA transcripts involved in R-loop formation (*24*, *25*), prevented hybrid formation, and this was reversed by the addition of the hnRNPA1 complex (fig. S1, M and N). Thus, ribonucleoproteins promote RNA invasion into the non-G4 strand across from a G4 structure to assemble a G-loop (Fig. 2I).

Homology-directed G-loop assembly

Posttranscriptional R-loop formation can be mediated by the RAD51 recombinase (*24*), and, similar to G4 structures, both RAD51 and its loader BRCA2 are enriched in regulatory genomic regions through unknown mechanisms (*26*, *27*). Depletion of RAD51 and BRCA2 prevented R-loop accumulation in pG4^{BOT} (Fig. 1, C and D). Consistent with a role in G-loop formation, BRCA2 and RAD51 were rapidly recruited to the plasmid (Fig. 1, E and F) and specifically located at the G4 region (fig. S2, A to C). This was followed by hnRNPA1 accumulation, allowing G-loop formation (Fig. 1F). RAD51 recruitment required a single-stranded DNA (ssDNA) region, because it was lost when G4 structures were placed on both strands (fig. S2E). Furthermore, recruitment of both RAD51 and hnRNPA1 was prevented upon depletion of BRCA2 (fig. S2D). This suggests that BRCA2-mediated RAD51 loading on the displaced G4-opposing strand promotes homology-directed RNA invasion.

Analogous to its role in recombination activation (*28*), the accumulation of BRCA2 and RAD51 at pG4^{BOT} and subsequent G-loop assembly were promoted by two DNA damage response kinases, ATR and ATM (Fig. 1G and fig. S2, C and F). Rapid ATR activation was detected,

[1]Oncode Institute, Hubrecht Institute-KNAW & University Medical Center Utrecht, Utrecht, Netherlands. [2]Department of Human Genetics, Leiden University Medical Center, Leiden, Netherlands. [3]Science for Life Laboratory, Division of Genome Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden. [4]Ming Wai Lau Centre for Reparative Medicine, Stockholm node, Karolinska Institutet, Stockholm, Sweden. [5]Hubrecht Institute-KNAW & University Medical Center Utrecht, Utrecht, Netherlands. [6]Institute of Biology Leiden, Leiden University, Leiden, Netherlands. ***Corresponding author. Email: k.sato@hubrecht.eu (K.S.); p.knipscheer@hubrecht.eu (P.K.)**

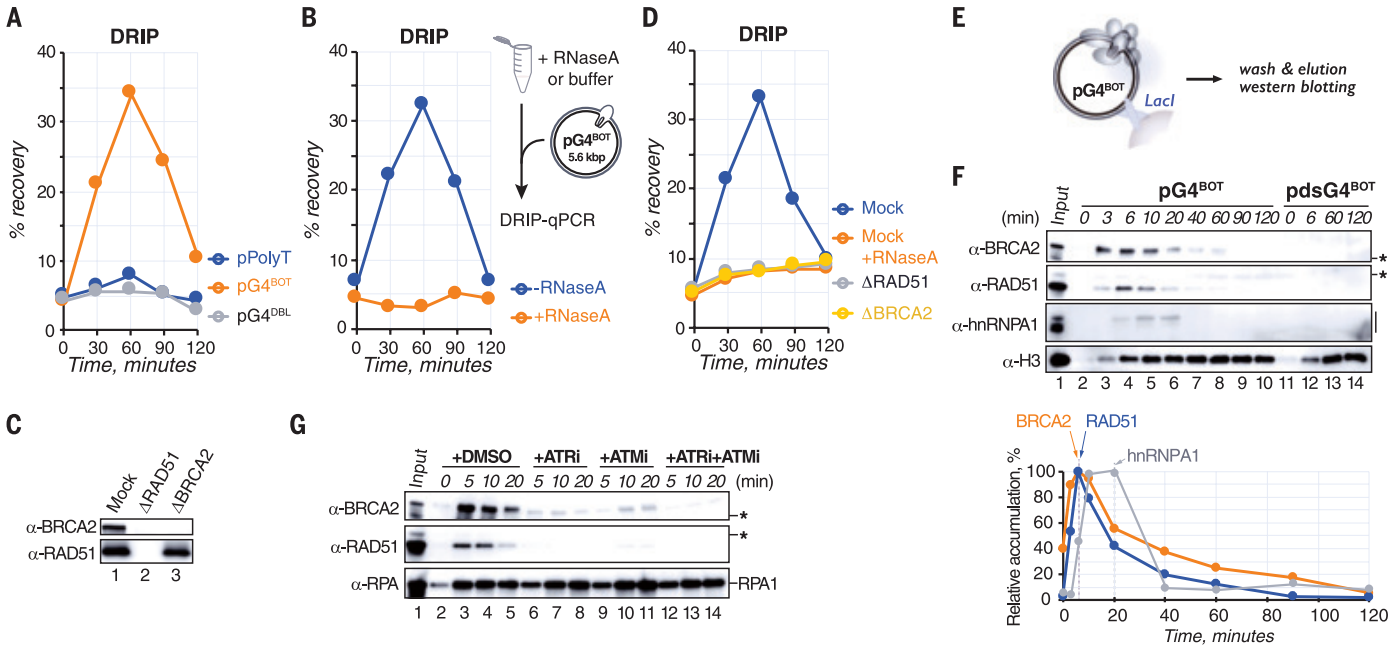


Fig. 1. Mechanism of G-loop establishment. (**A**) pPolyT, pG4^{BOT}, and pG4^{DBL} were incubated with pQuant in NPE, and products were analyzed by DRIP-qPCR with primers for the *G4* locus and the *pQuant* locus (for background signal measurement). After background subtraction, DRIP recovery percentages compared with input signals were plotted. (**B**) pG4^{BOT} was incubated in NPE that was pretreated with RNaseA or buffer, and products were analyzed as in (A). (**C**) Mock-, RAD51-, and BRCA2-depleted NPEs were analyzed by Western blot with RAD51 and BRCA2 antibodies. (**D**) pG4^{BOT} was incubated in the NPEs described in (C), and products were analyzed as in (A). Where indicated, NPE was pretreated with RNaseA. Co-depletion of BRCA2 with RAD51 complicated rescue experiments with recombinant proteins, so the involvement of BRCA2 and RAD51 in the G-loop assembly was not unequivocally established. (**E**) Scheme of the plasmid pull-down assay. In this assay, pG4^{BOT} was incubated in NPE, pulled down with streptavidin beads coated with biotinylated LacI, and bound proteins (oval) were analyzed by Western blot. (**F**) pG4^{BOT} and pdsG4^{BOT} were incubated in NPE and subjected to plasmid pull-down assay. Proteins were analyzed by Western blot with BRCA2, RAD51, hnRNPA1, and histone H3 antibodies (top). The band intensity of each protein was quantified, and the relative intensities compared with the highest signal among the conditions were plotted (bottom). (**G**) pG4^{BOT} was incubated in NPE supplemented with dimethylsulfoxide (DMSO), AZ20 (ATRi), and/or KU-55933 (ATMi); subjected to plasmid pull-down assay; and analyzed with BRCA2, RAD51, and RPA antibodies.

as measured by CHK1 phosphorylation (fig. S2G). ATR activation correlated with RPA binding to the unstructured strand and did not depend on BRCA2 and RAD51 (fig. S2, G to K). By contrast, RPA binding did not require ATR activation (Fig. 1G). Concurrent with RPA binding, both ATR and phosphorylated TOPBP1, a key ATR activator (*29*, *30*), accumulated on pG4^{BOT} (fig. S2, L and M). Therefore, G4-induced ATR activation to promote G-loop assembly resembles canonical checkpoint activation, in which ssDNA-bound RPA activates ATR through interaction with phosphorylated TOPBP1, stimulating recruitment of recombination factors (Fig. 2I).

G4 unwinding initiates G-loop disassembly

To investigate the fate of G-loops, we incubated pG4^{BOT} in egg extract and excised an 81–base pair (bp) fragment containing the G4 (fig. S3A). Native polyacrylamide gel electrophoresis (PAGE) analysis showed the initial noncomplementary bubble structure and an accumulating, faster-migrating, 81-bp linear duplex product (fig. S3, B to D). This structural conversion to duplex DNA must involve DNA synthesis past one of the strands of the G4-containing region (fig. S3A). Consistently, we observed active nucleotide incorporation in the G4 region (fig. S3, E and F). Differential digestion of the top and bottom strands showed preferential synthesis of the G4-complementary strand (Fig. 2, A and B, and fig. S3, C and D). This strand-specific DNA synthesis was blocked upon RAD51 and BRCA2 depletion, which was partly rescued by preloading of RAD51, but not the homologous pairing–deficient mutant (fig. S3, G to I). Thus, homology-mediated G-loop assembly promotes DNA synthesis. Similar G4 conversion was observed for a telomeric G4 (fig. S3J). Other DNA secondary structure motifs also induced R-loop formation and DNA synthesis, suggesting that this may be a general response to DNA secondary structures (fig. S3, K and L). Because DNA synthesis requires a free 3′ end, we hypothesized that this process may involve local nucleolytic incisions. Sequencing gel analysis of linearized reaction products indeed detected incisions specifically on the non-G4 strand one to eight nucleotides from the 3′ ssDNA–double-stranded DNA (dsDNA) junction, which depended on G-loop assembly (Fig. 2, C and D, and fig. S4A). On the basis of the substrate structure, these incisions are likely to be mediated by the XPF-ERCC1 or MUS81-EME1 endonuclease complexes (*31*). Depletion of XPF, but not MUS81, abrogated the incisions, which was rescued by the addition of recombinant XPF-ERCC1 (Fig. 2, E and F). Furthermore, XPF recruitment to the G4 region was dependent on G-loop assembly (fig. S4, B and C). Thus, G-loop assembly triggers the recruitment of XPF-ERCC1 to promote site-specific cleavage of the hybrid strand, leading to resynthesis of the G4-complementary strand.

Because XPF-ERCC1 acts most efficiently on ssDNA-dsDNA junctions with unstructured arms (*32*), we hypothesized that this incision requires prior G4 unfolding. To investigate this, we depleted two DNA helicases, DHX36 and FANCJ, which are also required for DNA replication–coupled G4 unwinding in *X. laevis* egg extracts (*33*). Loss of both helicases abolished G4 resolution and incision without affecting G-loop assembly and XPF-ERCC1 recruitment (Fig. 2, G and H, and fig. S4, B to H). The depletion also prevented G4 conversion, and these defects were rescued by recombinant wild-type (WT) DHX36 and FANCJ but not their ATPase-dead mutants (fig. S4, I and J). Moreover, recombinant XPF-ERCC1 catalytic activity is inhibited by a G4 structure on the 5′ ssDNA flap of a splayed arm model substrate (fig. S4K). G4 unwinding

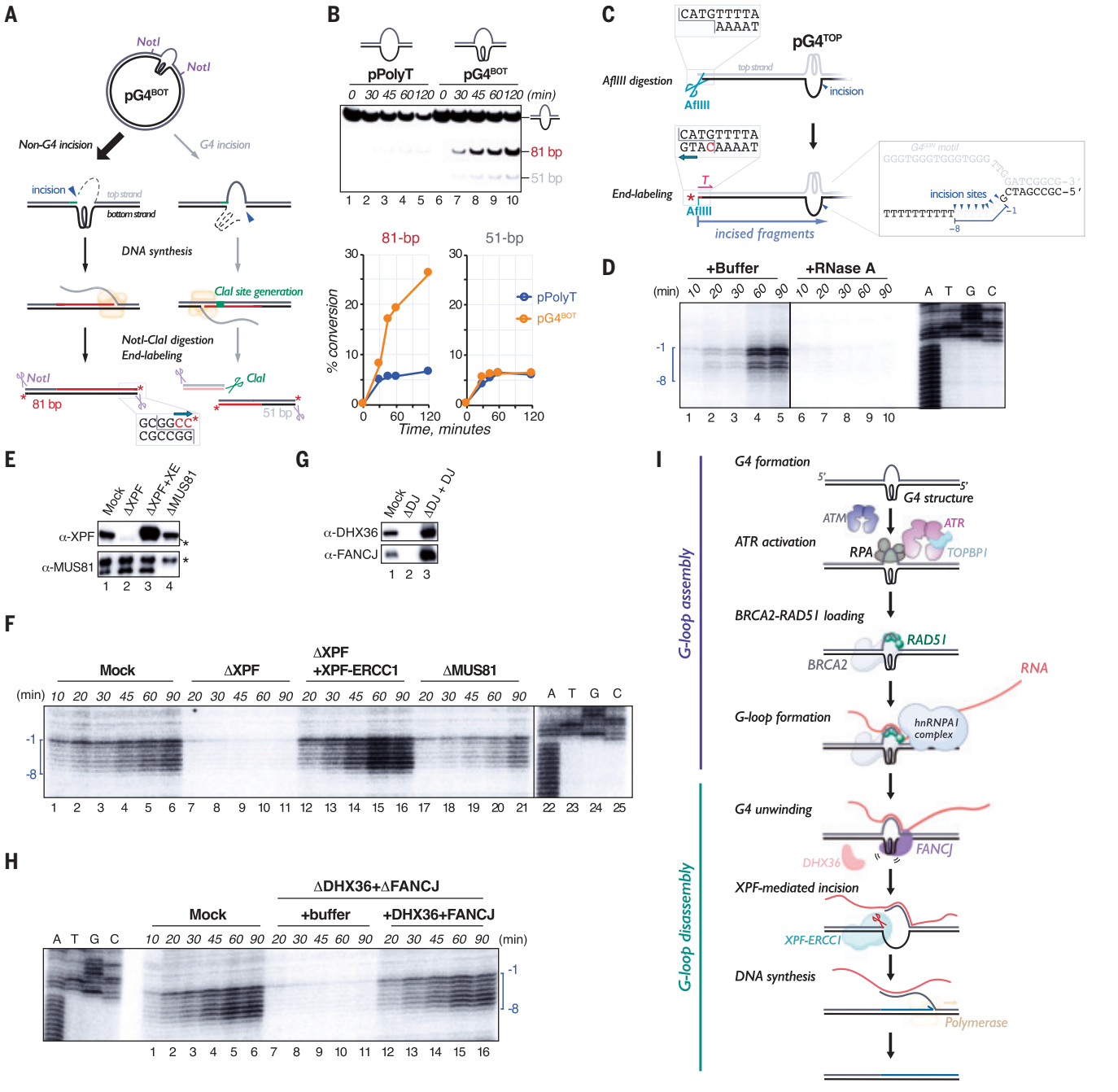


Fig. 2. Mechanism of G-loop disassembly. (**A**) Schematic of the G4 conversion assay. The G4-forming region of pG4^{BOT} is converted to complementary dsDNA in NPE. This conversion must involve DNA synthesis past at least one of the strands of the G4-containing region, which requires a free 3′ end that is likely generated by incision (blue arrowhead). The products were digested with NotI and ClaI and end-labeled (asterisk) by fill-in of the 5′-overhangs with Sequenase DNA polymerase (blue arrow) in the presence of ^{32}P-α-dCTP (see also fig. S3A). ClaI specifically incises the product when the G4 strand is resynthesized (see also fig. S1A). (**B**) pPolyT and pG4^{BOT} were incubated in NPE. At various times, DNA was isolated, digested with NotI and ClaI, end-labeled by fill-in of the 5′-overhangs with ^{32}P-α-dCTP, separated by native PAGE, and visualized by autoradiography (top). The 81- and 51-bp fragments were quantified, and the conversion percentage was calculated by comparing their intensity against the total intensity of all fragments at time point 0. Conversion was plotted against time (bottom). (**C**) Schematic representation of incision assay. After incubation of pG4^{TOP} in NPE, products were digested with AflIII and end-labeled by fill-in of the 5′-overhang with Sequenase DNA polymerase in the presence of ^{32}P-α-dCTP (blue arrow). The region surrounding the incision sites is enlarged and incision positions observed in (D) are indicated (blue arrowheads) (right). Incision products are numbered –8 to –1, where the –1 product corresponds to a product in which incision takes place between the first and the second nucleotide upstream of the ssDNA-dsDNA junction. (**D**) pG4^{TOP} was incubated in NPE that was pretreated with RNaseA or buffer. At various times, DNA was isolated, digested with AflIII, end-labeled by fill-in of the 5′-overhang with ^{32}P-α-dCTP, separated by denaturing PAGE, and visualized by autoradiography. A sequencing ladder derived from extension of primer *T* annealed to pG4^{TOP} (C) was used as a size marker. Incision products are indicated with a blue bracket. (**E**) Mock-, XPF-, and MUS81-depleted NPEs supplemented with buffer or XPF-ERCC1 were analyzed by Western blot with XPF and MUS81 antibodies (bottom). The asterisks represent nonspecific bands. (**F**) pG4^{TOP} was incubated in the NPEs described in (E), and products were analyzed as in (D). (**G**) Mock- and DHX36-FANCJ–depleted NPEs supplemented with buffer or WT DHX36 and FANCJ were analyzed by Western blot with DHX36 and FANCJ antibodies. (**H**) pG4^{TOP} was incubated in the NPEs described in (G), and products were analyzed as in (D). (**I**) Model for G4 suppression through the G-loop assembly-disassembly cycle.

is thus required for site-specific R-loop incision and G-loop disassembly. G4 unwinding coincided with G-loop assembly and required both BRCA2 and invading RNA but not downstream factors XPF-ERCC1 (fig. S4, E and L to O). Therefore, G4 resolution is a highly coordinated process that is initiated by G-loop assembly, leading to G4 structure unwinding followed by nucleolytic incision and DNA synthesis past the G4 sequence (Fig. 2I).

RNA sequence and abundance regulate G-loops

The sequence opposite the G4 in the pG4 substrates is a poly(T) stretch (fig. S1A). Therefore, the invading RNA could be poly(A) stretches in extract-derived RNA molecules. Consistent with this, replacement of the poly(T) with a poly(A) stretch in pG4^{BOT} blocked G-loop assembly (figs. S1A and S5A). To determine whether other sequences also support G-loop formation, we placed a 21-nucleotide unique DNA sequence opposite the G4 and incubated the plasmid (pG4^{BOT}-I) in NPE supplemented with a transcript containing a complementary sequence (transcript I) or a noncomplementary sequence (transcript II) (Fig. 3A). Upon incubation with the noncomplementary RNA, only minor RNA-DNA hybrid formation was detected (Fig. 3C), and the bubble structure underwent rapid degradation (Fig. 3B), as also seen when G-loop assembly was prevented (Fig. 2B and fig. S3, I and J). This degradation did not involve XPF-ERCC1 but was promoted by the MRE11 and DNA2 nucleases (fig. S5, B and C). Degradation was blocked by incubation with the complementary transcript I at a nearly 1:1 G4 plasmid-to-transcript ratio, which also supports G-loop formation and subsequent disassembly (Fig. 3, B and C). Similar results were obtained with a different G4-opposing sequence, pG4^{BOT}-II (fig. S5, D to F), supporting a universal mechanism by which homologous RNA transcripts stimulate G-loop assembly and protect them from degradation. Although increasing the level of complementary transcript to a 1:5 G4 plasmid-to-transcript ratio further promoted G-loop assembly, G-loop disassembly was inhibited (Fig. 3, B and C). Titration showed that stoichiometric concentrations ranging from 0.4 to 4 transcript copies per G4 were required for efficient G-loop disassembly, whereas higher concentrations prevented disassembly (Fig. 3D). Unlike complementary RNA, complementary ssDNA did not support either G4 suppression or stabilization (fig. S5G). We conclude that G-loop assembly and disassembly are controlled by RNA availability and levels, thereby conferring plasticity to G4 regulation (fig. S5H).

G-loop disassembly shapes G4–R-loop landscapes

To explore the role of G-loops in the regulation of cellular G4 landscapes, we deleted *Dhx36*, *Fancj*, or both genes in mouse embryonic stem cells (mESCs) and mapped genomic G4 structures by quantitative CUT&Tag (*14*) (Fig. 4A and fig. S6). More than 97% of G4 CUT&Tag peaks colocalized with PQSs, validating the specificity of this method (fig. S7A). Individual loss of either DHX36 or FANCJ resulted in an increase of G4 signal at 12,109 and 693 sites, respectively (fig. S7B). Moreover, their combined loss resulted in the up-regulation at 21,045 G4s genome wide, accounting for 29% of the total 71,445 G4s detected. Up-regulated G4s overlapped among the mutant lines (fig. S7C), and their G4 signals were further elevated in the double mutant compared with the single mutants (Fig. 4, B and C, and

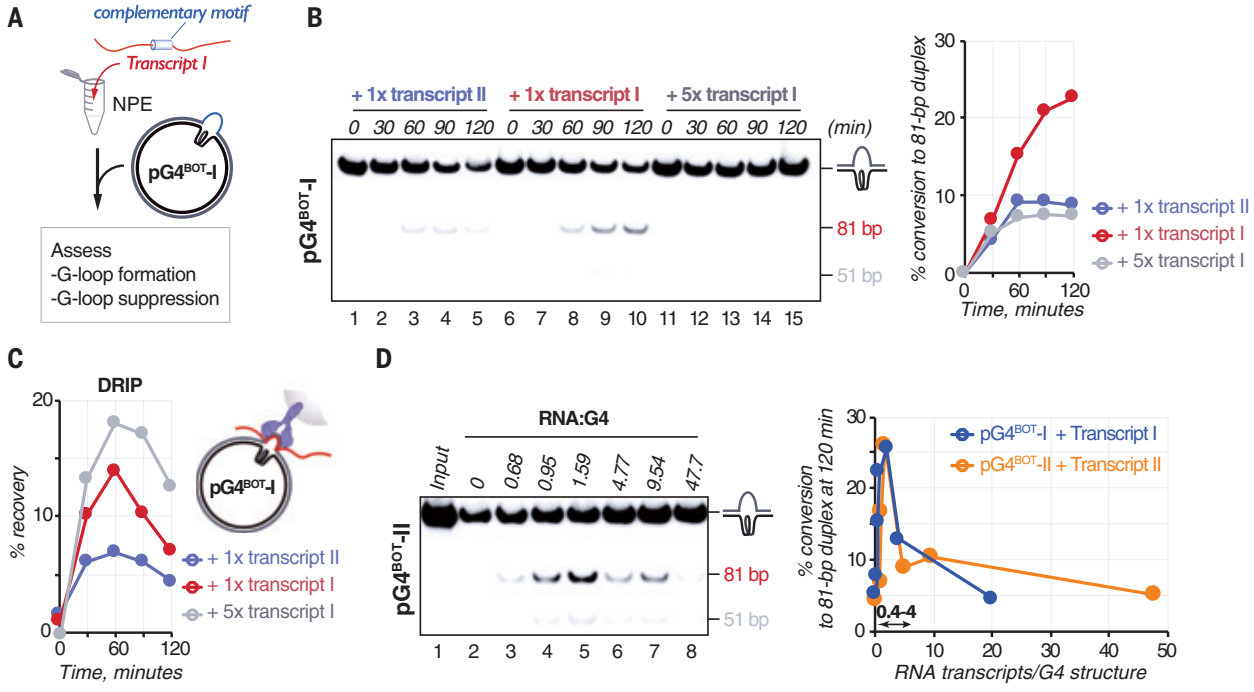


Fig. 3. RNA sequence and abundance regulate G-loops. (**A**) Experimental setup for transcript addition. A synthetic RNA transcript (transcript I) with a sequence complementary to the DNA opposite the G4 in pG4^{BOT}-I was used. DRIP and G4 conversion assays were performed in the presence of this substrate and RNA. (**B**) pG4^{BOT}-I was incubated in NPE supplemented with transcript I or a noncomplementary transcript (transcript II) at 10 nM (1×) or 50 nM (5×), and products were analyzed by G4 conversion assay (top). The 81-bp fragment was quantified, and the conversion percentage was calculated by comparing their intensity against the total intensity of all fragments at time point 0. Conversion was plotted against time (bottom). (**C**) pG4^{BOT}-I was incubated with pQuant in NPE supplemented with transcript I or transcript II at the indicated ratio, and products were analyzed by DRIP-qPCR with primers for the *G4* locus and the *pQuant* locus (for background signal measurement). After background subtraction, DRIP recovery percentages compared with input signals were plotted. (**D**) pG4^{BOT}-I and pG4^{BOT}-II were incubated in NPE supplemented with transcript I and transcript II, respectively, at the indicated ratio, and products were analyzed as in (B). The 81-bp fragment at time point 120 min was quantified, and the conversion percentage was calculated by comparing their intensity against the total intensity of all fragments at time point 0 (input). Conversion was plotted against the ratio of transcript to plasmid. G4 conversion was specifically stimulated at a 0.4:4 transcript-to-plasmid template ratio (double black arrow).

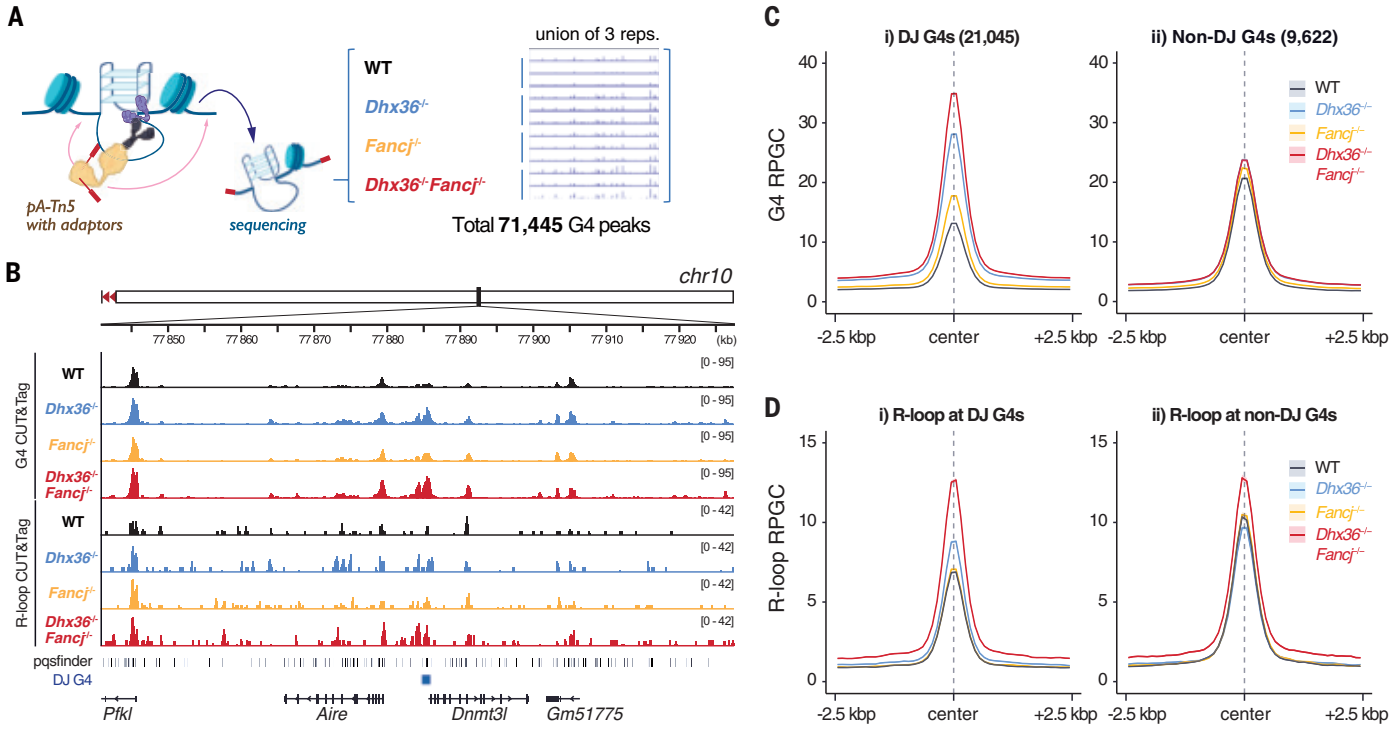


Fig. 4. G-loop disassembly shapes G-loop landscapes. (**A**) Schematic of quantitative G4 CUT&Tag. Permeabilized cells were treated with the FLAG-tagged BG4 antibody (blue) to capture G4s in chromatin. An anti-FLAG antibody (black) was subsequently added to tether the pA-Tn5 transposase bound to adaptors (red) for chromatin tagmentation. Genomic DNA was extracted after tagmentation and mixed with tagmented spike-in DNA. Tagmented fragments were amplified and sequenced. The number of total high-confidence G4 CUT&Tag peaks identified across technical triplicates in WT, *Dhx36*$^{-/-}$, *Fancj*$^{-/-}$, and *Dhx36*$^{-/-}$*Fancj*$^{-/-}$ mESCs is indicated. (**B**) Genome browser view at a representative locus showing RPGC-normalized (1× Genome Coverage) G4 and R-loop CUT&Tag peaks from triplicate (*y*-axis scale, right top) PQSs determined by pqsfinder and DJ G4s in the indicated mESCs. G4s significantly up-regulated in *Dhx36*$^{-/-}$*Fancj*$^{-/-}$ cells [$\log_2$ fold change ($\log_2$FC) >1, false discovery rate (FDR) ≤1%] were defined as DJ G4s and were significantly unchanged ($\log_2$FC <1, FDR ≤1%) as non-DJ G4s. (**C**) G4 distribution around 21,045 DJ G4 (i) and 9622 non-DJ (ii) G4 sites in the indicated mESCs. Average profiles of RPGC-normalized G4 CUT&Tag peaks within ±2.5 kilobase pairs (kbp) from their center are plotted. In WT cells, 9622 non-DJ G4 peaks shared a significantly higher signal ($P < 2.2 \times 10^{-16}$ calculated by the Wilcoxon test; Cohen's $d = 0.59$, medium) compared with the 21,045 DJ G4 peaks. Conversely, in *Dhx36*$^{-/-}$*Fancj*$^{-/-}$ cells, the 21,045 DJ G4 peaks shared a significantly higher signal ($P < 2.2 \times 10^{-16}$ calculated by the Wilcoxon test; Cohen's $d = 0.58$, medium) compared with the 9622 non-DJ G4 peaks. (**D**) R-loop CUT&Tag was performed in the indicated mESCs, and R-loop distributions around 21,045 DJ G4s (i) and 9622 non-DJ G4s (ii) were plotted with average profiles of RPGC-normalized R-loop CUT&Tag peaks within ±2.5 kbp from each G4 center.

fig. S7D), indicating partially redundant G4 suppression by DHX36 and FANCJ in mESCs. Very few G4s formed at de novo sites upon depletion of DHX36 and FANCJ; rather, the signal at existing G4 sites increased (Fig. 4B and fig. S7B). Therefore, the DHX36 and FANCJ helicases, rather than preventing de novo G4 formation, play an important role in limiting the abundance of intrinsic G4 structures throughout the genome.

Both the up-regulated 21,045 G4s (DJ G4s) and the unchanged 9622 G4s (non-DJ G4s) in double mutant cells compared with WT cells predominantly resided in promoters and enhancers and frequently overlapped with actively transcribed RNA (*3*, *34*) (fig. S7, E to G). DJ G4s were specifically found at G4 sites with low G4 signal in WT cells but were excluded from sites with high WT G4 signal, which often overlapped with non-DJ G4 sites (Fig. 4, B and C). This indicates that stably formed G4s, for example, those at active promotors, may not be substrates for DHX36-FANCJ–mediated disassembly. Consistent with the characteristic DJ G4 landscape, quantitative R-loop CUT&Tag detected a lower RNA-DNA hybrid signal at DJ G4s than at non-DJ G4s in WT cells (Fig. 4D). Loss of DHX36 and FANCJ resulted in robust accumulation of RNA-DNA hybrids at DJ G4s (Fig. 4D). Moreover, transcriptome sequencing revealed that loss of both helicases led to robust up-regulation of transcripts throughout the genome (fig. S8, A to D, and data S1 and S2). These data suggest that a defect in the dynamic G-loop assembly-disassembly cycle leads to G-loop accumulation and transcriptional dysregulation.

To further validate this mechanism, we examined another player in G-loop disassembly, XPF-ERCC1, using G4 and R-loop CUT&Tag in ERCC1-deficient mESCs (*35*) (fig. S9A). Because our biochemical data suggested that XPF-ERCC1 acts downstream of G4 unwinding (Fig. 2I), we hypothesized that its depletion would not affect G4 abundance but rather would increase R-loops due to the lack of G-loop disassembly. As expected, the loss of *Ercc1* resulted in the accumulation of R-loops at DJ G4 sites with at most a mild G4 up-regulation (fig. S9, B to D). We also addressed the recruitment of recombination factors to G4 sites by comparing BRCA2 chromatin immunoprecipitation sequencing data from WT mESCs (*36*) with our G4 sites, which revealed that 82% of BRCA2 peaks resided at the G4 sites (figs. S7A and S10). Therefore, G4-forming regions are a predominant BRCA2 substrate in unperturbed cells, supporting a genome-wide role for BRCA2 in G-loop assembly. Finally, consistent with G-loop disassembly requiring low concentrations of invading RNA in *X. laevis* egg extract, we found that the transcripts targeting to DJ G4 sites, particularly the long noncoding RNAs, were expressed at considerably lower levels compared with those to non-DJ G4s (fig. S8E). We thus conclude that the G-loop assembly-disassembly cycle through coordinated actions by the HR proteins, DHX36, FANCJ, and XPF-ERCC1 is a global mechanism used to repress G4s.

Two-tier G4 suppression ensures genome stability

Persistent G4 structures have been shown to cause DNA damage (*12*, *17*). Consistently, loss of DHX36 and FANCJ increased 53BP1-positive foci, which is indicative of DNA double-strand breaks (DSBs) (fig. S11A). Although DSBs detected in both 5-ethynyl-2′-deoxyuridine (EdU)–positive and EdU-negative cells, their accumulation was more pronounced in EdU-positive cells (Fig. 5A), indicating major damage induction in the S phase. DSBs are at least partly generated through transcription in the G_1 phase, because transcription inhibition reduced 53BP1 foci in EdU-negative cells (fig. S11B).

To further investigate the effect of *Dhx36* and *Fancj* deletion on DNA replication, we used single-cell EdU sequencing (scEdU-seq) (*37*). Two pulses of EdU labeling, in combination with single-cell sequencing, enabled us to detect fork-stalling events and genome-wide replication timing throughout the S phase (Fig. 5B). We found DJ G4s to be enriched in early replicating regions (Fig. 5C, top). FANCJ loss caused an increase in fork stalling, specifically at the early S phase. Although single DHX36 loss had little effect on DNA replication, loss of both helicases further exacerbated fork stalling at the early S phase (Fig. 5C), corresponding to reduced EdU incorporation (fig. S11, C and D). These results were not due to a change in replication timing (fig. S11, E and F) or to G4s stochastically formed at the replication fork, because replication stalling was confined to the early S phase; therefore, they instead show that G-loops transmitted from the G_1 phase led to replication fork stalling.

In addition to their role in G-loop–mediated G4 resolution, DHX36 and FANCJ also promote G4 resolution when replication forks collide with a G4 structure in *X. laevis* egg extract (*33*). Loss of this mechanism caused persistent replication stalling, which could result in enhanced γH2AX foci in mitotic chromosomes in the double-knockout cells (Fig. 5, D and E). Furthermore, these cells showed replication stress–induced genome instability, as represented by enhanced micronuclei and 53BP1 bodies and intrinsic growth defects (fig. S12). Stabilization of G4 structures by pyridostatin (*12*) phenocopied the DNA damage induction and genome instability observed in the double mutant, confirming that these adverse phenotypes were caused by persistent G4s throughout the genome (Fig. 5 and figs. S11 and S12). Loss of *Dhx36* alone, which accumulates G4 structures in mESCs (Fig. 4) but has only a mild effect on replication-coupled G4 resolution in egg extract (*33*), did not result in the accumulation of DSBs (Fig. 5, A and E, and fig. S12, A to C). Thus, the persistence of G4s upon S-phase entry is likely a major cause of replication-associated DSBs only if

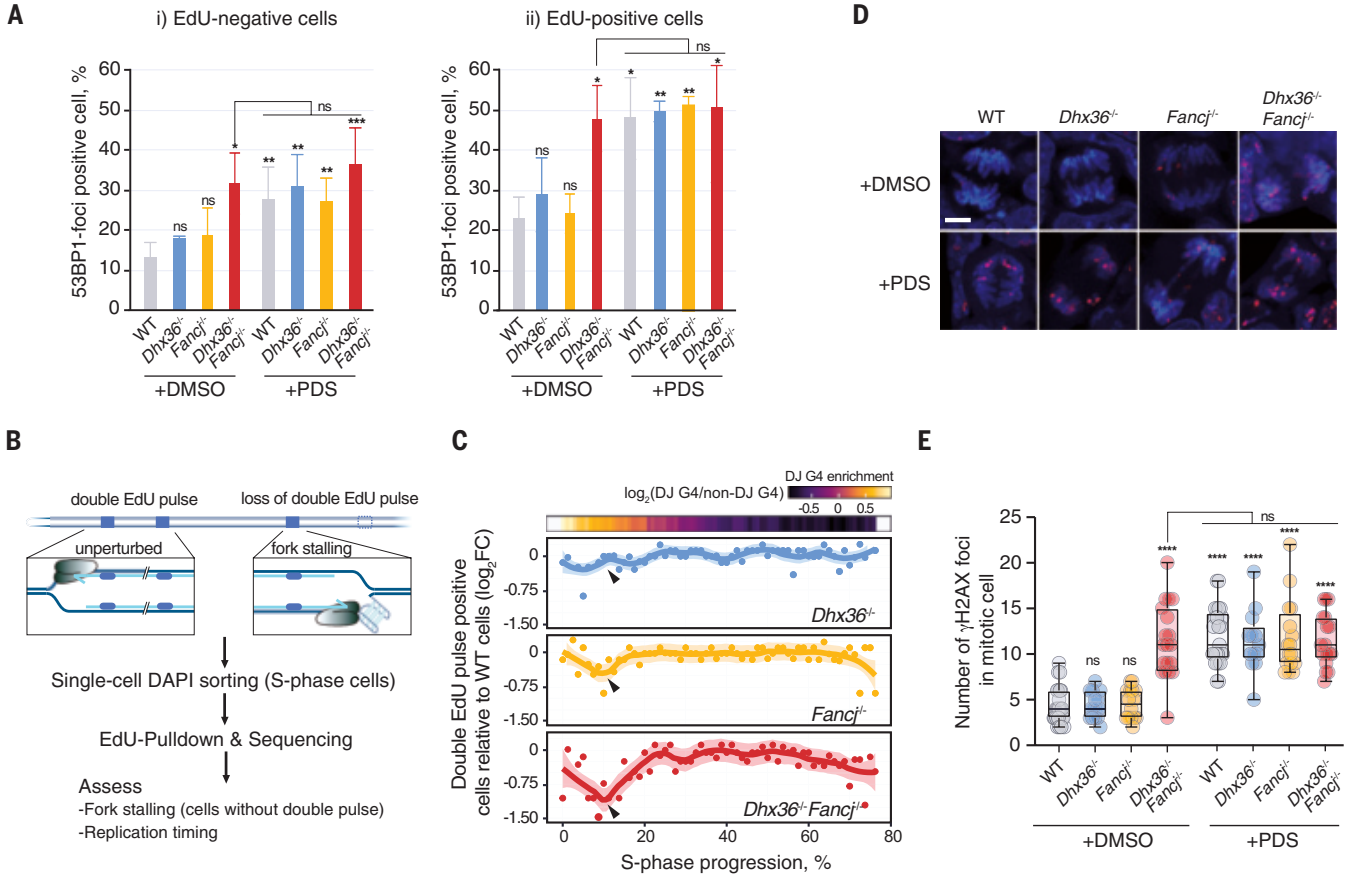


Fig. 5. Two-tier mechanism of G4 resolution. (**A**) 53BP1 foci-positive cells were counted and displayed as the mean percentage of total EdU-negative cells (i) or EdU-positive cells (ii) with SDs ($n = 3$, with >40 cells for each experiment). WT cells treated with DMSO, or where indicated, *Dhx36*$^{-/-}$*Fancj*$^{-/-}$ cells treated with DMSO, were compared for a *t* test. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$; ns, not significant. (**B**) Schematic of scEdU-seq. Two pulses of EdU labeling in combination with single-cell sequencing detected fork-stalling events. 4′,6-Diamidino-2-phenylindole (DAPI) intensity was used to determine genome-wide replication timing throughout the S phase. (**C**) Cells labeled with a double-EdU pulse were analyzed by scEdU-seq. The relative population of cells with the double EdU pulses compared with WT cells was plotted over S-phase progression in the indicated mESCs. Reduction of double EdU pulses indicates fork stalling. Heatmap of DJ G4 enrichment relative to non-DJ G4s is shown (top). (**D**) mESCs were treated with DMSO or pyridostatin (PDS) and visualized by immunostaining with a γH2AX antibody, along with DAPI counterstaining. Representative mitotic chromosomes are shown. Scale bar, 5 μm. (**E**) Box plots of the γH2AX focus number in mitotic cells described in (D) ($n = 2$, with >20 mitotic cells for each experiment). Center line indicates the median; box limits, 75 and 25 percentiles; and whiskers, min and max values. ****$P < 0.0001$.

replication-coupled G4 resolution is impaired. G4 resolution through G-loop assembly-disassembly throughout the cell cycle represents the first tier of protection against G4-induced DNA damage, but if this fails and G4s persist into S phase, then replication-coupled G4 unwinding provides an essential backup to ensure genome integrity.

Discussion

Our data establish a stepwise process that controls G4 dynamics through the assembly of a G-loop by the invasion of RNA transcripts across from the G4, followed by G4 unwinding coupled with G-loop disassembly (Fig. 2I). Interfering with this G-loop cycle results in severe dysregulation of the transcriptome, extensive genome instability, and proliferation defects, highlighting its vital role in cellular homeostasis and survival. This mechanism involves several unanticipated aspects. First, the G4 structure is recognized as a "DNA lesion," thereby enabling activation of the downstream DNA repair proteins in the absence of DNA damage. Second, it uses RNA molecules in trans that are not physically linked to ongoing transcription for RNA-DNA hybrid formation. Although such RNA trans action has been thought to occur at confined loci in vertebrates (*38*), our findings define it as a prevalent genome-wide mechanism that plays a twofold role in the G-loop assembly-disassembly cycle and the protection from nucleolytic attacks when G4 structures form. Third, this mechanism results in renewal of the non-G4 strand, which allows for erasing potentially mutagenic modifications on the displaced strand (*39*–*42*), thereby safeguarding the integrity of regulatory DNA motifs. Fourth, the mechanism is controlled by the concentration of invading RNAs, allowing for both G4 stabilization and suppression, and this defines RNA levels as an important regulatory parameter of G4 landscapes (fig. S5H). Finally, we observed that other DNA secondary structure motifs also promote RNA-DNA hybrid formation, but whether these are processed through a similar mechanism is a subject for future research.

Many of the proteins involved in this G-loop cycle, including RAD51, BRCA2, FANCJ, and XPF-ERCC1, also promote canonical DNA repair pathways and are directly linked to human diseases characterized by developmental abnormalities and predisposition to cancer. Although these pathways safeguard genome stability against exogenous and endogenous DNA damage, the G-loop mechanism specifically protects the genome against the accumulation of endogenous DNA structures. Given the abundance of G4 structures and DNA damage accumulation observed here, defects in G4 resolution can cause DNA damage to an extent that exceeds most other forms of endogenous DNA damage, thus posing a severe threat to cellular homeostasis. This highlights an essential, DNA lesion-independent role for DNA repair proteins in genome maintenance.

REFERENCES AND NOTES

1. D. Varshney, J. Spiegel, K. Zyner, D. Tannahill, S. Balasubramanian, *Nat. Rev. Mol. Cell Biol.* **21**, 459–474 (2020).
2. V. S. Chambers *et al.*, *Nat. Biotechnol.* **33**, 877–881 (2015).
3. R. Hänsel-Hertsch *et al.*, *Nat. Genet.* **48**, 1267–1272 (2016).
4. W. W. I. Hui, A. Simeone, K. G. Zyner, D. Tannahill, S. Balasubramanian, *Sci. Rep.* **11**, 23641 (2021).
5. R. Hänsel-Hertsch *et al.*, *Nat. Genet.* **52**, 878–883 (2020).
6. K. G. Zyner *et al.*, *Nat. Commun.* **13**, 142 (2022).
7. N. Maizels, *EMBO Rep.* **16**, 910–922 (2015).
8. D. Chakravarti, K. A. LaBella, R. A. DePinho, *Cell* **184**, 306–322 (2021).
9. M. L. Duquette, P. Handa, J. A. Vincent, A. F. Taylor, N. Maizels, *Genes Dev.* **18**, 1618–1629 (2004).
10. W. T. C. Lee *et al.*, *Nat. Commun.* **12**, 2525 (2021).
11. I. Cheung, M. Schertzer, A. Rose, P. M. Lansdorp, *Nat. Genet.* **31**, 405–409 (2002).
12. R. Rodriguez *et al.*, *Nat. Chem. Biol.* **8**, 301–310 (2012).
13. K. Paeschke *et al.*, *Nature* **497**, 458–462 (2013).
14. J. Lyu, R. Shao, P. Y. Kwong Yung, S. J. Elsässer, *Nucleic Acids Res.* **50**, e13 (2022).
15. P. Wulfridge *et al.*, *Mol. Cell* **83**, 3064–3079.e5 (2023).
16. M. L. Duquette, M. D. Huber, N. Maizels, *Cancer Res.* **67**, 2586–2594 (2007).
17. A. De Magis *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 816–825 (2019).
18. W. S. Hoogenboom, D. Klein Douwel, P. Knipscheer, *Dev. Biol.* **428**, 300–309 (2017).
19. G. Miglietta, M. Russo, G. Capranico, *Nucleic Acids Res.* **48**, 11942–11957 (2020).
20. J. Walter, L. Sun, J. Newport, *Mol. Cell* **1**, 519–529 (1998).
21. M. P. Crossley *et al.*, *Nature* **613**, 187–194 (2023).
22. P. L. Tran, J. L. Mergny, P. Alberti, *Nucleic Acids Res.* **39**, 3282–3294 (2011).
23. S. Carpenter *et al.*, *Science* **341**, 789–792 (2013).
24. M. Feretzaki *et al.*, *Nature* **587**, 303–308 (2020).
25. S. Redon, I. Zemp, J. Lingner, *Nucleic Acids Res.* **41**, 9117–9128 (2013).
26. I. Hazan, J. Monin, B. A. M. Bouwman, N. Crosetto, R. I. Aqeilan, *Cell Rep.* **29**, 560–572.e4 (2019).
27. J. J. Gruber *et al.*, *Cell Rep.* **28**, 2182–2193.e6 (2019).
28. R. Buisson *et al.*, *Mol. Cell* **65**, 336–346 (2017).
29. C. Frattini *et al.*, *Mol. Cell* **81**, 1231–1245.e8 (2021).
30. A. Kumagai, J. Lee, H. Y. Yoo, W. G. Dunphy, *Cell* **124**, 943–955 (2006).
31. P. M. Dehé, P. H. L. Gaillard, *Nat. Rev. Mol. Cell Biol.* **18**, 315–330 (2017).
32. M. Jones *et al.*, *Nat. Commun.* **11**, 1120 (2020).
33. K. Sato, N. Martin-Pintado, H. Post, M. Altelaar, P. Knipscheer, *Sci. Adv.* **7**, eabf8653 (2021).
34. M. Tastemel *et al.*, *Stem Cell Res.* **25**, 250–255 (2017).
35. N. Bona, G. P. Crossan, *Nat. Struct. Mol. Biol.* **30**, 1434–1445 (2023).
36. Y. Wang *et al.*, *Nat. Commun.* **13**, 3493 (2022).
37. J. van den Berg *et al.*, *Nat. Methods* **21**, 1175–1184 (2024).
38. L. Statello, C. J. Guo, L. L. Chen, M. Huarte, *Nat. Rev. Mol. Cell Biol.* **22**, 96–118 (2021).
39. L. B. Alexandrov *et al.*, *Nature* **500**, 415–421 (2013).
40. P. Revy *et al.*, *Cell* **102**, 565–575 (2000).
41. M. Muramatsu *et al.*, *Cell* **102**, 553–563 (2000).
42. S. Ito *et al.*, *Science* **333**, 1300–1303 (2011).
43. S. J. Elsasser, J. Lyu, C. Navarro Luzón, K. Sato, P. Knipscheer, "G4 and Rloop qCUT&Tag - RNA transcripts regulate G-quadruplex landscapes through G-loop formation" (Zenodo, 2025); https://doi.org/10.5281/zenodo.15127429.
44. J. van den Berg, "Gloop-scEdUseq: Zenodo Gloop release" (Zenodo, 2025); https://doi.org/10.5281/zenodo.15121012.

ACKNOWLEDGMENTS

We thank M. van Vugt and M. Takata for feedback on the manuscript; the Hubrecht animal caretakers for animal support; the Hubrecht flow cytometry core facility and the Utrecht Sequencing Facility for single-cell sorting and sequencing, which are subsidized by the University Medical Center Utrecht, the Hubrecht Institute, Utrecht University, and the Netherlands X-omics Initiative (NWO project 184.034.019); T. Nyman, E. Strandback, and the members of the Knipscheer laboratory for assistance and helpful discussions; S. Noordermeer for initial help with the DRIP-qPCR protocol; G. P. Crossan, J. Garaycoechea, and J. Smink for providing the *Ercc1*$^{-/-}$ mESCs; and Protein Production Sweden (PPS) for providing BG4 antibody. **Funding:** This work was supported by the Kanae Foundation for the Promotion of Medical Science (K.S.); the Japanese Biochemical Society (K.S.); the European Research Council (ERC Consolidator Grant ERC-CoG 101003210 to P.K., ERC Advanced grant ERC-AdG 101053581-scTranslatomics to A.v.O., and ERC Starting Grant ERC-StG 851564 to F.M.); The Netherlands Organisation for Scientific Research (NWO) through the Gravitation program CancerGenomiCs.nl (P.K.); The Oncode Institute, which is partly financed by the Dutch Cancer Society (KWF) (A.v.O. and P.K.); Vetenskapsrådet Sweden (grant 2020–04313, 2023-02380 to S.J.E.); KI StratRegen Junior Project Grant (S.J.E.); The Ming Wai Lau Center for Reparative Medicine, Sweden (S.J.E.); Stiftelsen för Strategisk Forskning (FFL7 to S.J.E.); The Knut och Alice Wallenbergs Stiftelse, Sweden (S.J.E.); and the Karolinska Institutet KI Research Foundation (grant 2022-01826 to C.N.). P.P.S. is funded by the Swedish Research Council as national research infrastructure. The National Academic Infrastructure for Supercomputing in Sweden (NAISS) is funded by the Swedish Research Council through grant agreement no. 2022-06725. The Supercomputing Cluster UPPMAX was used under projects NAISS 2023/6-19, NAISS 2023/22-84, SNIC 2022/6-14, and NAISS 2024/22-108. **Author contributions:** Conceptualization: K.S., P.K.; Data curation: K.S., J.L., J.v.d.B., A.v.O., M.T., S.J.E., P.K.; Formal analysis: K.S., J.L., J.v.d.B., V.M.C., C.N.L., J.D., S.J.E.; Funding acquisition: K.S., C.N.L., F.M., A.v.O., S.J.E., P.K.; Investigation: K.S., J.L., J.v.d.B., D.B., J.S., C.E.J., M.A., J.D., A.H.; Methodology: K.S., J.L., J.v.d.B., V.M.C., C.N.L., S.J.E., J.S., C.E.J., A.H.; Resources: J.S., C.E.J., M.T.; Software: J.L., J.v.d.B., V.M.C., C.N.L., S.J.E.; Validation: K.S., J.L., J.v.d.B., D.B., J.S., C.E.J., M.A., J.D., A.H.; Visualization: K.S., J.L., J.v.d.B., V.M.C., C.N.L., S.J.E., J.S., C.E.J.; Project administration: K.S., P.K.; Supervision: F.M., A.v.O., M.T., S.J.E., P.K.; Writing – original draft: K.S., P.K., J.L., J.v.d.B., J.S., F.M., M.T., S.J.E.; Writing – review & editing: K.S., P.K., J.L., S.J.E. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. All material generated in this study (including but not limited to oligonucleotides, plasmids, purified proteins, antibodies, and cell lines) are available on request. CUT&Tag data, RNA-seq data, and scEdU-seq data generated for this study have been deposited at the Gene Expression Omnibus under GSE268153 (scEdU-seq) and GSE269084 (CUT&Tag and RNA-seq). Custom code used in this study has been deposited on Zenodo (*43*,*44*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adr0493
Materials and Methods; Figs. S1 to S17; Tables S1 to S5; References (*45*–*117*); Data S1 to S2; MDAR Reproducibility Checklist

Submitted 12 June 2024; resubmitted 28 February 2025; accepted 19 April 2025

10.1126/science.adr0493

How burnout snuck up on me

Md Doulotuzzaman Xames

When I started my Ph.D., I believed I had found my true purpose. My research focused on how digital tools could help prevent burnout among clinicians—something I'm passionate about. I spent my days immersed in research about workload, stress, fatigue, and the fragile tipping points that push people beyond their limits. As my workload grew, I started to feel the pressure. But even as I worked later and later into the night, I kept reassuring myself: I was only studying burnout. It wasn't something that could happen to me.

Unlike many in my cohort, I settled on my dissertation topic early in the first semester, thanks to my prior research experience and a quick alignment with my adviser's interests. I hit the ground running, and at first, the work energized me. I loved feeling I was part of a broader research community and knowing my work could one day help people on the front lines of health care. But the challenge of balancing my research with teaching responsibilities and the required coursework was intense, and before I knew it, my preliminary exam—a big hurdle to continuing my studies—was right around the corner.

At the time, I didn't notice how I was letting work take over my life, bit by bit. Skipping lunch to finish "just one more" section of a manuscript. Working weekends because "I'm already behind." Feeling my chest tighten when I opened my email each morning, dreading any new additions to my growing to-do list. At first, I called it normal stress. Then, a rough patch. Eventually, I stopped calling it anything at all.

One evening near the end of my second year, desperate for reassurance, I took a burnout "self-test." To my surprise, I scored high on several classic symptoms. Emotional exhaustion? Check. Feeling numb and disconnected from my work? Check. Losing the sense of personal accomplishment? Check. I stared at the results, feeling exposed.

Still, I resisted the idea for weeks until I finally reached a breaking point. It arrived quietly one evening as I was staring at a paragraph I had rewritten 10 times. No matter how much I worked, the gap between what I wanted to write and what I could deliver only seemed to widen. I closed my laptop and thought, for the first time, "Maybe I can't do this anymore."

The next day, I had my regular meeting with my adviser. As we wrapped up our discussion, he paused, looking at me for a moment longer than usual. "Xames, you should take a break!" he said lightly, but with real concern. He had sensed what I hadn't yet fully admitted to myself.

That comment unlocked something inside me. For the first time, I allowed myself to admit that I was not OK. From my research on burnout, I knew the risk factors—long hours, poor boundaries, chronic stress. But I had completely ignored them creeping into my own life.

In the weeks that followed, I did something that felt both terrifying and necessary. I scaled back. I started to set real boundaries—no more writing emails after dinner, no more glorifying 60-hour workweeks. I went back to hobbies that had nothing to do with my dissertation. These are the kinds of restorative activities the research recommends.

It wasn't an instant fix. Some days, the old voices still whispered: You should be working harder. You're falling behind. But slowly, I learned to answer them differently: I am a person first, and a researcher second.

Ironically, or maybe inevitably, my work improved. Ideas came more freely when I wasn't drowning in anxiety. Writing felt less like extracting teeth and more like creating something real again. I was no longer studying burnout from the safe distance of an observer—I had lived it. It was a reminder that behind the abstract models and metrics are real people.

Today, my research remains centered on burnout, but my focus has shifted to also include recovery, sustainability, and compassion. The work feels deeper, messier, and more honest—and it no longer consumes my life. Most of all, what my experience taught me is that even when work feels urgent and important, so is your well-being. That lesson didn't come from a study. It came from the long, slow, humbling process of realizing that I am human—and that's not a flaw. □

Md Doulotuzzaman Xames is a Ph.D. candidate at Virginia Polytechnic Institute and State University.

ILLUSTRATION: ROBERT NEUBECKER