Human evolution revealed in the UK Biobank *p. 21*

A fluid model for crowd motion *pp. 27 & 46*

In situ sequencing during a Lassa outbreak *pp. 30 & 74*

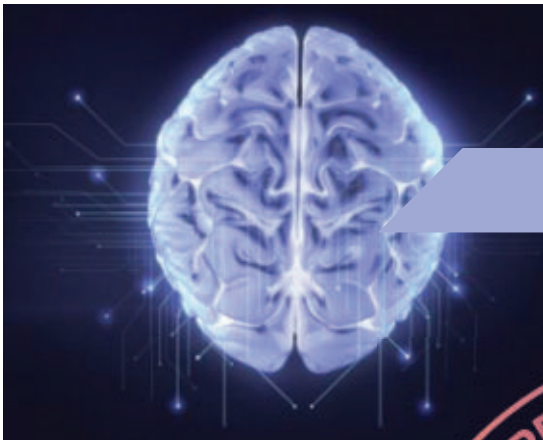
Science

$15
4 JANUARY 2019
sciencemag.org

AAAS

WALKING TALL

A giant synapsid from the Triassic with erect limbs *p. 78*



Apply for our exciting research Prize!

SCIENCE-PINS PRIZE FOR NEUROMODULATION

SCIENCE-PINS

$25, 000 Grand Prize!
Get published in *Science*!

The *Science*-PINS Prize is a highly competitive international prize that honors scientists for their excellent contributions to neuromodulation research. For purposes of the Prize, neuromodulation is any form of alteration of nerve activity through the delivery of physical (electrical, magnetic, or optical) stimulation or chemical agents to targeted sites of the nervous system.

For full details, judging criteria and eligibility requirements, visit:

www.sciencemag.org/prizes/pins

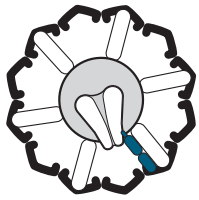
Submission Deadline: March 15, 2019

Science AAAS

PINS 品驰

Science Translational Medicine AAAS

CONTENTS



31 & 44
Internalizing receptors to forget

4 JANUARY 2019 • VOLUME 363 • ISSUE 6422

NEWS

IN BRIEF

IN DEPTH

FEATURES

INSIGHTS

PERSPECTIVES

POLICY FORUM

BOOKS *ET AL.*

CONTENTS

33 & 61
Ultracold plasma

4 JANUARY 2019 • VOLUME 363 • ISSUE 6422

ON THE COVER

Artist's rendering of an early morning in Silesia, Poland, during the Late Triassic. The elephant-sized dicynodont *Lisowicia bojani*, a distant cousin of present-day mammals, plods slowly through the woodland as gracile theropod dinosaurs search for prey. In the foreground, the mammalian ancestor *Hallautherium* moves cautiously along a branch. *Lisowicia*, the largest known dicynodont, had upright limbs and walked with an erect gait, similar to modern mammals. See page 78.
Illustration: Julius Csotonyi

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $1971; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery): $98. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling; 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 is granted by AAAS to libraries and others who use Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

Science
AAAS
1200 New York Avenue, NW, Washington, DC 20005
Clarendon House, Clarendon Road, Cambridge, UK CB2 8FH

SUBSCRIPTION SERVICES For change of address, missing issues, new orders and renewals, and payment questions: 866-434-AAAS (2227) or 202-326-6417, FAX 202-842-1065. Mailing addresses: AAAS, P.O. Box 96178, Washington, DC 20090-6178 or AAAS Member Services, 1200 New York Avenue, NW, Washington, DC 20005

INSTITUTIONAL SITE LICENSES 202-326-6730 **REPRINTS:** Author Inquiries 800-635-7181 **COMMERCIAL INQUIRIES** 803-359-4578 **PERMISSIONS** 202-326-6765, permissions@aaas.org **AAAS Member Central Support** 866-434-2227 www.aaas.org/membercentral.

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

INFORMATION FOR AUTHORS See **www.sciencemag.org/authors/science-information-authors**

EDITORIAL

Examining author gender data

Editor-in-Chief,
Science Journals.
jberg@aaas.org

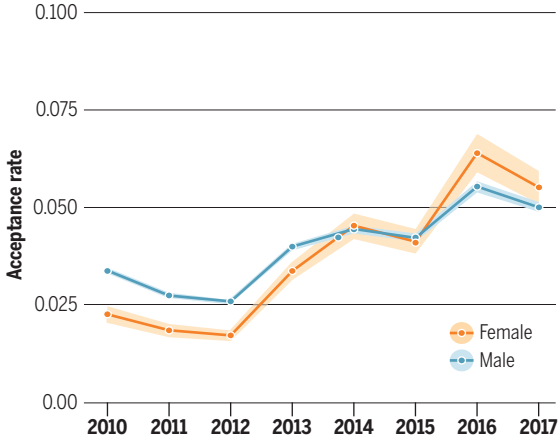
I previously reported results of a study on author gender in *Science* (science.sciencemag.org/content/355/6323/329) that was based on the examination of a random sample of approximately 2600 authors for which gender was inferred by painstaking analysis of websites and similar sources. Unfortunately, this approach does not scale well to large samples necessary for many analyses. We have since initiated systematic efforts to examine gender distributions of key populations of authors and reviewers for the *Science* family of journals using additional data and tools and plan to use this information to guide policy development and other appropriate steps to address any gender disparities.

Since our initial study, we have collected gender and other demographic information voluntarily from approximately 5000 individuals (authors and reviewers). To extend this dataset further, we've used first name–based gender inference software that we validated and calibrated with accurate datasets, as described in the accompanying *Sciencehound* post (https://blogs.sciencemag.org/sciencehound/2019/01/03/new-tools-for-gender-analysis). The automated tool provides inferred genders for more than 70% of the authors of both published and rejected *Science* papers submitted from 2010 to 2017. Based on comparisons with the individual-provided dataset, these inferred genders are more than 93% accurate on an individual basis. Furthermore, the inferred gender information can be extended to populations with gender distributions that appear to be more than 98% accurate.

The fractions of male and female authors for Reports submitted to *Science* are nearly constant over the 8-year period examined. Twenty-five ± 1% of the first authors are female, while 18 ± 1% of the corresponding authors are female, consistent with the results from our earlier analysis. These figures reflect the weighted averages across the different fields covered by *Science*. Separating submissions by field reveals that 30% of first authors of submissions in the life sciences are female, compared with 16% in the physical sciences and 22% in other fields. The values for corresponding authors are 19, 12, and 20%, respectively.

The gender distributions for papers that were published can be compared with those for the overall submissions pool. The acceptance rates for Reports were not significantly different for female as compared to male first authors for papers submitted in 2016 and 2017, although significant differences were observed favoring male authors from 2011 to 2015. Further work is in progress to determine if these disparities are due to gender biases in reviews, editorial decisions, or other factors such as institutional differences and preferences. Significant differences in Report acceptance rates by gender of corresponding authors were observed in 2012, 2014, and 2015. We can also examine other article types. For example, the fraction of female corresponding authors for Perspectives (driven largely by editorial invitations to Perspective authors and, indirectly, peer reviewers) grew steadily from 19 to 26% from 2010 to 2017. The *Science* news team's analysis of data regarding its use of women and men as sources and in quotes in its stories reveals an increase in the fraction of quoted female sources from approximately 20% to more than 30% over the course of 2018.

Acceptance rates for Reports by gender of first authors



With these data and tools in place, we are now well positioned for further analyses and actions that address gender disparities. We plan to examine the gender distribution of our peer reviewers. Social science studies indicate that women and men tend to have similar gender biases, based on their perceptions of the gender distribution of the population that they are examining. Nonetheless, ensuring that gender distributions for reviewers approximate those for authors is good practice, and peer reviewers do get access to exciting scientific results and are often invited to write Perspectives. We plan to share these and other analyses and encourage others to perform and share similar examinations.

–Jeremy Berg

GRAPH: J. BERG/*SCIENCE*; (TOP RIGHT) TERRY CLARK

10.1126/science.aaw4633

NEWS

IN BRIEF

Edited by **Jeffrey Brainard**

AREAS TO WATCH

What's coming up in 2019

Scientists in Europe and the United States face an uncertain political landscape in the new year, which could affect funding and collaborations. The threat is most acute in the United Kingdom, which plans to exit the European Union in March but has not settled on the terms of its departure. Some big research findings could share the headlines, however, including the first clear images of the supermassive black hole at the heart of our galaxy, from astronomers in an international collaboration called the Event Horizon Telescope. *Science*'s news staff forecasts other areas of research and policy likely to make news this year.

All eyes on polar ice

CLIMATE SCIENCE | If you want to understand Earth's warming future, look to the poles. This year, scientists in two international projects will heed that call. In September, researchers will position a German icebreaker, the *RV Polarstern*, to freeze in Arctic sea ice for a year's stay. The ship will serve as the central hub for the €120 million Multidisciplinary drifting Observatory for the Study of Arctic Climate, hosting researchers from 17 countries. They'll study how polar clouds, ocean dynamics, and first-year ice formation contribute to the Arctic's shift to ice-free summers. Then, near year's end, researchers from the United States and United Kingdom will fan out across the remote Thwaites Glacier, the part of the Antarctic ice sheet most at risk of collapsing into the ocean and driving up sea levels, in the first full season of a $50 million, 5-year effort. They'll probe the ice's structure and the water and land beneath it, using everything from seismometers to instrument-carrying seals. Both missions will benefit from revitalized satellite coverage, as two satellites launched last year, the Ice, Cloud, and Land Elevation Satellite-2 and the Gravity Recovery and Climate Experiment Follow-on, which measure ice height and mass, respectively, begin to beam science data back home.

A science whisperer for Trump

SCIENCE POLICY | For 2 years, President Donald Trump has been making decisions involving science and innovation without input from a White House science adviser. Meteorologist Kelvin Droegemeier, whom Trump nominated in late July 2018 to fill that void, was awaiting final Senate approval at press time. The question is what his arrival will mean for the administration's handling of an array of technical challenges, from regulation of human embryo engineering and self-driving cars to combatting cyberterrorism and fostering a more tech-savvy workforce. Some science-soaked issues may already have been settled, such as leaving the Paris climate accord and forsaking the Iran nuclear deal. But many others remain unresolved,

PHOTO: ALFRED WEGENER INSTITUTE/STEFAN HENDRICKS

The *RV Polarstern*, shown here on a 2013 polar research cruise, will spend a winter frozen in Arctic sea ice.

PHOTO: CARLOS BARRIA/REUTERS/NEWSCOM

including how to deal with Chinese espionage at U.S. universities without stifling global scientific cooperation.

New rights for alleged harassers

#METOO | This year, the U.S. Department of Education may finalize controversial proposed rules that would reduce universities' liability for policing sexual harassment and sexual assault and give more rights to the accused. The regulations, proposed in November 2018, would change how institutions investigate such allegations under the landmark 1972 law known as Title IX. They wouldn't be responsible for investigating most off-campus incidents of harassment or assault, and the standard of evidence for confirming allegations of on-campus misconduct could rise. The definition of sexual harassment would be narrowed from "unwelcome conduct of a sexual nature" to "unwelcome conduct on the basis of sex that is so severe, pervasive, and objectively offensive that it effectively denies a person equal access" to education. And defendants' lawyers will be able to cross-examine accusers. The department is accepting comments on the proposals until 28 January.

Seeking new physics in the muon

PARTICLE PHYSICS | By studying the magnetism of a particle called the muon, physicists hope to find results this year that could point to new particles or forces, something they have craved for decades. Scientists at Fermi National Accelerator Laboratory (Fermilab) in Batavia, Illinois, are examining whether the muon—a heavier and shorter-lived cousin of the electron—is more magnetic than theory predicts. The Muon g-2 experiment found a hint of such an excess when it ran at Brookhaven National Laboratory in Upton, New York, from 1997 to 2001. Physicists moved the experiment's 15-meter-wide electromagnet to Fermilab in 2013, upgraded the apparatus, and started to record data in January 2018. A first result could be one of the biggest stories in particle physics this year, with the world's biggest atom smasher, Europe's Large Hadron Collider, idled for 2 years of upgrades.

A fine-grained look inside cells

BIOPHYSICS | In cell biology, higher resolution means more gets revealed. Now, scientists are ready to use new combinations of tools and techniques to provide close-up looks at components inside cells in unprecedented detail, and in 3D. Already, researchers can analyze DNA, proteins, RNA, and epigenetic marks in single cells. This year, multidisciplinary teams plan to combine those methods with advances in cryoelectron tomography, labeling techniques to trace molecules, and other types of microscopy to see subcellular structures and processes. For example, a multifaceted technique for imaging and staining DNA could shed new light on how chromosomes fold. And the blended methods could yield clearer pictures at the molecular level of how cells divide and change shape, and how gene activity affects structure and function.

Solar dimming gets a test

CLIMATE SCIENCE | A geoengineering technique to curb global warming by temporarily dimming the sun's rays could get its first, modest field experiment this year. In solar geoengineering, vast amounts of reflective aerosol particles would be sprayed into the high atmosphere, mimicking the cooling effects of volcanic eruptions. The Stratospheric Controlled Perturbation Experiment, led by climate scientists at Harvard University, will test the idea in a small, controlled way. If its plans are approved by an advisory board, the team will loft a balloon into the stratosphere, where it will release 100 grams of reflective particles—probably calcium carbonate, the chalky ingredient in antacid tablets. Flying back through the plume, the balloon will observe its cooling effect. Solar-radiation management, as it's known, is controversial. It does not reduce the built-up carbon dioxide that drives climate change and ocean acidification, and there's no accepted international governance.

Divided we stand?

SCIENCE POLICY | You'll need a Ouija board to predict how U.S. science will fare this year under a divided government, with Democrats now in control of the House of Representatives while Republicans retain

With control of the U.S. House of Representatives, Democrats will likely examine environmental policies.

the Senate with President Donald Trump in the White House. There are the known flashpoints—Democrats challenging the Trump administration on its environment and energy policies, for example. Spending cuts will be on the table as lawmakers face tight budget caps mandated by a 2011 law. Then there are the what-ifs, including whether the Supreme Court will throw out a citizenship question on the 2020 census and lawmakers can suspend partisan bickering long enough to pass an infrastructure package that would boost U.S. innovation. A few science-savvy new members of Congress hope to lend a hand.

New GM mosquitoes take off

BIOTECHNOLOGY | The first release of genetically modified (GM) mosquitoes in Africa is set to happen in Burkina Faso this year, an initial step in a planned "gene drive" strategy against malaria. It will be the first release of GM mosquitoes of the genus *Anopheles*, which transmits the parasite responsible for the disease. The gene drive approach, under development at the nonprofit consortium Target Malaria, would spread mutations through the wild

population that knock out key fertility genes or reduce the proportion of female insects, which transmit disease. But the first GM *Anopheles* mosquitoes released won't bear such mutations and aren't intended to cut down the population. Researchers will let out fewer than 10,000 genetically sterilized males to observe how they survive and disperse in the wild and to help introduce the concept of GM mosquitoes to regulators and community members.

Nations size up biodiversity

CONSERVATION | Three years in the making, a $2.4 million assessment of Earth's biodiversity and ecosystems will be published in May. By evaluating trends over 50 years in indicators such as species extinctions and extent of marine protected areas, it will chart progress toward international goals on biodiversity conservation—and, in many places, how far short the world is falling. Experts from 50 nations have participated in a review of scientific literature and government data conducted under the auspices of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. The report, the first since a similar effort in 2005, will forecast the future of species on the planet under business-as-usual and other scenarios. The new assessment is intended to inform the next generation of biodiversity targets, due in 2020.

The next planetary mission

SPACE SCIENCE | In July, NASA will chart its next major step in planetary science when it selects the next billion-dollar mission under its New Frontiers program. The agency will choose between two finalists. Dragonfly would send a semiautonomous quad-copter to fly across the surface of Titan, the saturnian moon sculpted by rivers of liquid methane. The copter would search for clues of chemical reactions that could lead to life. The Comet Astrobiology Exploration Sample Return mission would return gases and ice from the nucleus of the comet 67P/Churyumov-Gerasimenko. Such samples, likely unaltered for billions of years, could provide a window into the role comets played in delivering water and organic compounds to Earth in its early history.

A push to return museum holdings

RESEARCH ETHICS | Researchers are beginning new efforts to return bones and cultural artifacts collected for study and as museum specimens to the peoples from whom they were obtained, often without consent. Expect renewed debate on this issue, as after centuries of exploitative collecting, some researchers use new methods to collaborate with those communities, and also expand efforts to return objects of art. A study from Australia published last month showed ancient DNA can be used to reliably link the remains of Aboriginal ancestors to living descendants. Some warn, however, that widespread adoption of similar methods could be used to coerce communities into genetic testing. In France, a government-commissioned report recommended in November 2018 that over the next 5 years, French museums work with colleagues in Africa to repatriate tens of thousands of cultural artifacts looted during colonial rule if their countries of origin ask for them.

A global assessment will examine endangered species, which include the ploughshare tortoise in Madagascar.

PHOTO: RYAN BOLTON/SHUTTERSTOCK

Disease crisis looms for swine

LIVESTOCK AGRICULTURE | Pig farmers—and perhaps some bacon lovers—will anxiously scan the headlines this year for news of African swine fever (ASF). Harmless to humans, the viral disease is highly infectious and lethal among pigs, causing serious economic damage through culls and trade bans. ASF made major jumps in Europe last year, turning up for the first time in pigs and wild boar in Bulgaria and in boar in Belgium and Hungary. The virus can jump from boar, which are difficult to manage, to swine. Germany, Denmark, and other major pork producers are on high alert. Most worrisome was the first detection of the virus in China, a long-dreaded development in the country with the world's largest pig population. China has recorded more than 80 outbreaks since August 2018, including in boar. Authorities have clamped down on the transport of pigs, culled more than 630,000,and last month reportedly banned pig farming where wild boar are present. Despite these efforts, the virus could still explode in China and elsewhere in Asia.

China eyes bioethics overhaul

BIOETHICS | China is likely to tighten its rules for genetic engineering of humans, including the creation of heritable traits, in the wake of an uproar over such work in 2018. A Chinese scientist named He Jiankui announced in November 2018 that he modified a gene in embryos that led to twin baby girls. The modification is meant to protect them and their descendants from HIV infection, a feat widely condemned in China and worldwide as unethical, unjustified, and possibly harmful to the babies. Most countries ban or outlaw such experiments. In China, however, what is apparently the most relevant regulation was enacted in 2003 and never updated to cover advances in gene editing. Since the announcement, numerous Chinese researchers, ethicists, and officials have called for an overhaul of the country's bioethics laws and regulations, although no agency or institution has been named to lead the effort. Another question for this year is whether He will face sanctions.

ILLUSTRATION: DAVIDE BONAZZI/SALZMAN ART

PUBLISHING

The world debates open-access mandates

Spurred by European funders behind Plan S, many countries consider similar moves

By **Tania Rabesandratana**

How far will Plan S spread?

Since the September 2018 launch of the Europe-backed program to mandate immediate open access (OA) to scientific literature, 16 funders in 13 countries have signed on. That's still far shy of Plan S's ambition: to convince the world's major research funders to require immediate OA to all published papers stemming from their grants. Whether it will reach that goal depends in part on details that remain to be settled, including a cap on the author charges that funders will pay for OA publication (*Science*, 30 November 2018, p. 983). But the plan has gained momentum: In December 2018, China stunned many by expressing strong support for Plan S (*Science*, 14 December 2018, p. 1218). This month, a national funding agency in Africa is expected to join, possibly followed by a second U.S. funder. Others around the world are considering whether to sign on.

Plan S, scheduled to take effect on 1 January 2020, has drawn support from many scientists, who welcome a shake-up of a publishing system that can generate large profits while keeping taxpayer-funded research results behind paywalls. But publishers (including AAAS, which publishes *Science*) are concerned, and some scientists worry that Plan S could restrict their choices.

If Plan S fails to grow, it could remain a divisive mandate that applies to only a small percentage of the world's scientific papers. (Delta Think, a consulting company in Philadelphia, Pennsylvania, estimates that the first 15 funders to back Plan S accounted for 3.5% of the global research articles in 2017.) To transform publishing, the plan needs global buy-in. The more funders join, the more articles will be published in OA journals that comply with its requirements, pushing publishers to flip their journals from paywall-protected subscriptions to OA, says librarian Jeffrey MacKie-Mason, the chief digital scholarship officer at the University of California, Berkeley.

> ***"[Plan S] is perhaps our best chance to transform the publishing industry soon."***
> **Jeffrey MacKie-Mason**, University of California, Berkeley

Robert-Jan Smits, the European Commission's OA envoy in Brussels, who is one of the architects of Plan S, says publishers have stalled by emphasizing the need for broad participation. "The big publishers told me: 'Listen, we can only flip our journals [to OA] if this is signed by everyone. So first go on a trip around the world and come back in 20 years. Then we can talk again,'" Smits recalls. "Some people try to do anything to keep the status quo."

OA mandates are nothing new: In Europe, 74 research funders require that papers be made free at some point, up from 12 in 2005, according to the Registry of Open Access Repository Mandates and Policies. But existing policies typically allow a delay of 6 or 12 months after initial publication, during which papers can remain behind a publisher paywall.

Plan S requires immediate OA; it also insists that authors retain copyright and that hybrid journals, which charge subscrip-

tions but also offer a paid OA option, sign "transformative agreements" to switch to fully OA.

Some European funders think Plan S goes too far. "We and many German [organizations] think that we should not be as prescriptive as Plan S is," says Wilhelm Krull, secretary general of the Volkswagen Foundation, a private research funder in Hannover, Germany. The country is Europe's top producer of scientific papers, ahead of the United Kingdom and France, whose main funding agencies have signed on to Plan S. Germany's biggest federal funding agency, DFG, said it supports Plan S's goals but prefers to let researchers drive the change. Other funders, including the Estonian Research Council, say the timeline is too tight, and they will reconsider joining when Plan S's impact is clearer.

Other European funders are weighing pros and cons. Spain's science ministry says it is analyzing the potential repercussions of Plan S on the country's science and finances, and on researchers' careers. FNRS, the fund for scientific research in Belgium's Wallonia-Brussels region, is waiting for Plan S to announce its cap on article-processing charges (APCs), the fees for publishing in OA journals, which the coalition's funders have pledged to pay. "We're not ready to commit if the costs are too high," says Véronique Halloin, secretary-general of FNRS, whose existing OA mandate caps APC reimbursement at €500—which Halloin admits is on the low side.

Many await the European Commission's policy: Although its grants represent a small percentage of research funding in Europe, its OA rules can influence national mandates. The commission's research chief, Carlos Moedas, supports Plan S, and its 7-year funding program Horizon Europe, which will begin in 2021, contains general statements of support for OA. Plan S's rules will go into the program's model contract for grants, Smits says.

Smits has found unexpected support from China, which now produces more scientific papers than any other country. Last month, China's largest government research funder and two national science libraries issued strong statements backing Plan S's goals. "China needs to contribute to international open access [and] open its research results to its own people," says Zhang Xiaolin of ShanghaiTech University in China, who chairs the Strategic Planning Committee of the Chinese National Science and Technology Library. Even if Chinese organizations do not join Plan S formally, similar OA policies in China would have a "huge, perhaps decisive impact on the publishing industry," MacKie-Mason says.

For now, North America is not following suit. The Bill & Melinda Gates Foundation was the first Plan S participant outside Europe, and another private funder may follow. But U.S. federal agencies are sticking to policies developed after a 2013 White House order to make peer-reviewed papers on work they funded freely available within 12 months of publication (*Science*, 10 April 2015, p. 167). "We don't anticipate making any changes to our model," said Brian Hitson of the U.S. Department of Energy in Oak Ridge, Tennessee, who directs that agency's public access policy.

Nor are the three main federal research funders in Canada ready to change their joint 2015 OA policy. Plan S is "a bold and aggressive approach, which is why we want to make sure we've done our homework to ensure it would have the best effect on Canadian science," says Kevin Fitzgibbons, executive director of corporate planning and policy at Canada's Natural Sciences and Engineering Research Council in Ottawa.

Outside Europe and North America, funders gave *Science* mixed responses about Plan S. India, the third biggest producer of scientific papers in the world, will "very likely" join Plan S, says Krishnaswamy VijayRaghavan in New Delhi, principal scientific adviser to India's government. But the Russian Science Foundation is not planning to join. South Africa's National Research Foundation says it "supports Plan S in principle," but wants to consult stakeholders before signing on. Jun Adachi of the National Institute of Informatics in Tokyo, an adviser to the Japan Alliance of University Library Consortia for E-Resources, says that despite interest from funders and libraries, OA has yet to gain much traction in his country.

South America has a strong tradition of OA repositories and fee-free publishing, often with government subsidies. Bianca Amaro, president of LA Referencia, a Santiago-based Latin American network of repositories, says Plan S takes a more "systemic view" than previous policies, and she values its pledge to monitor APCs and their impact—a worry for lower-income countries. "We'll see how Europe handles this," she says.

Of course, MacKie-Mason says, not every funding agency will agree that Plan S is the best way to universal OA. "But some will agree it's good enough and perhaps our best chance to transform the publishing industry soon," he says. It comes in the wake of often incremental OA initiatives in the past 15 years, and some disagreement about the best route to OA.

"In the OA movement, it seems to a lot of people that you have to choose a road: green or gold or diamond," says Colleen Campbell, director of the OA2020 initiative at the Max Planck Digital Library in Munich, Germany, referring to various styles of OA. "Publishers are sitting back laughing at us while we argue about different shades" instead of focusing on a shared goal of complete, immediate OA. Because of its bold, stringent requirements, she and others think Plan S can galvanize advocates to align their efforts to shake up the publishing system.

The Plan S team predicts steady growth in the coming months. Funders will discuss Plan S in São Paulo, Brazil, at the May meeting of the Global Research Council, an informal group of funding agencies. Although Smits will leave the European Commission in March, the Plan S coalition is seeking a replacement who can keep the momentum going.

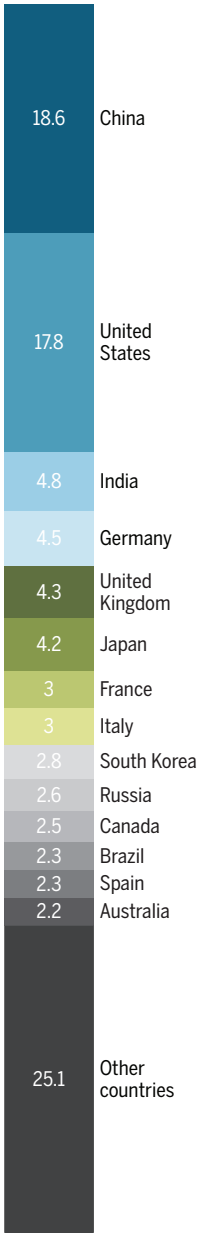
"The combined weight of Europe and China is probably enough to move the system," says astrophysicist Luke Drury, of the Dublin Institute for Advanced Studies and the lead author of a cautiously supportive response to Plan S by All European Academies, a federation of European academies of sciences and humanities.

If Plan S does succeed in bringing about a fairer publishing system, he says, a transition to worldwide OA is sure to follow. "Somebody has to take the lead, and I'm pleased that it looks like it's coming from Europe." ■

With reporting by Jeffrey Brainard, Sanjay Kumar, Dennis Normile, and Brian Owens.

Paper players

Percentages of the world's 2016 science articles by country



CREDITS: (GRAPHIC) N. DESAI/*SCIENCE*; (DATA) NATIONAL SCIENCE BOARD, *SCIENCE & ENGINEERING INDICATORS 2018*

In some wet tropical mountains, carbon dioxide is captured and flushed out of the atmosphere.

PHOTO: ROBERT HARDING/ALAMY STOCK PHOTO

CLIMATE

Tropical uplift may set Earth's thermostat

Indonesia's mountains could be cause of current glacial age

By **Paul Voosen**

Hate the cold? Blame Indonesia. It may sound odd, given the contributions to global warming from the country's 270 million people, rampant deforestation, and frequent carbon dioxide (CO_2)-belching volcanic eruptions. But over much longer times, Indonesia is sucking CO_2 out of the atmosphere.

Many mountains in Indonesia and neighboring Papua New Guinea consist of ancient volcanic rocks from the ocean floor that were caught in a colossal tectonic collision between a chain of island volcanoes and a continent, and thrust high. Lashed by tropical rains, these rocks hungrily react with CO_2 and sequester it in minerals. That is why, with only 2% of the world's land area, Indonesia accounts for 10% of its long-term CO_2 absorption. Its mountains could explain why ice sheets have persisted, waxing and waning, for several million years (although they are now threatened by global warming).

Now, researchers have extended that theory, finding that such tropical mountain-building collisions coincide with nearly all of the half-dozen or so significant glacial periods in the past 500 million years. "These types of environments, through time, are what sets the global climate," said Francis Macdonald, a geologist at the University of California, Santa Barbara, when he presented the work last month at a meeting of the American Geophysical Union in Washington, D.C. If Earth's climate has a master switch, he suggests, the rise of mountains like Indonesia's could be it.

Most geologists agree that long-term changes in the planet's temperature are governed by shifts in CO_2, and that plate tectonics somehow drives those shifts as it remakes the planet's surface. But for several decades, researchers have debated exactly what turns the CO_2 knob. Many have focused on the volcanoes that rise where plates dive beneath one another. By spewing carbon from Earth's interior, they could turn up the thermostat. Others have emphasized rock weathering, which depends on mountain building driven by plate tectonics. When the mountains contain seafloor rocks rich in calcium and magnesium, they react with CO_2 dissolved in rainwater to form limestone, which is eventually buried on the ocean floor. Both processes matter; "the issue is which one is changing the most," says Cin-Ty Lee, a volcanologist at Rice University in Houston, Texas.

Having the right rocks to drive the CO_2-chewing reaction is not sufficient. Climate matters, too. For example, the Siberian Traps, a region that saw devastating volcanic eruptions 252 million years ago, are rich in such rocks but absorb little, says Dennis Kent, a geologist at Rutgers University in New Brunswick, New Jersey. "It's too damn cold," he says. Saudi Arabia has the heat and the rocks but lacks another ingredient. "It's hotter than Hades but it doesn't rain." Indonesia's location in the rainy tropics is just right. "That is probably what's keeping us centered in an ice age," Kent adds.

Over the past few years, Macdonald and his collaborators have searched for other times when tectonics and climate could have conspired to open an Indonesia-size CO_2 drain. They found that glacial conditions 90 million and 50 million years ago lined up neatly with the collisions of a chain of island volcanoes in the now-vanished Neo-Tethys Ocean with the African and Asian continents. A similar collision some 460 million years ago formed the Appalachians, but it was thought to have taken place in the subtropics, where a drier climate does not favor weathering. By reanalyzing ancient magnetic fields in rocks formed in the collision, Macdonald's team found the mountains actually rose deep in the tropics. And their uplift matched a 2-million-year-long glaciation. "They're developing a pretty compelling story that this was a climate driver in Earth's past," says Lee Kump, a paleoclimatologist at Pennsylvania State University in University Park.

But those cases could be exceptions. So the team compiled a database of every tectonic "suture"—the linear features left by tectonic collisions—known to contain ophiolites, those bits of volcanic sea floor, over the past half-billion years. Based on magnetism in each suture's rocks and a model of continental drift, they mapped their ancient latitudes to see which formed in the topics, and when. "We were surprised that this is not as complicated as we thought," Macdonald said.

The team compared the results to records of past glaciations and found a strong correlation. They also looked for declines in volcanism, which might have cooled the climate. But their influence was much weaker, Macdonald said.

Kimberly Lau, a geochemist at the University of Wyoming in Laramie, calls the work "exciting in idea and novel in execution." Lee, however, would like to see direct evidence from ancient sediments that the collisions drove up rock weathering. "They have to go to the sink and study those," he says. And a recent study challenges the mountain thermostat idea with evidence for the importance of volcanoes. The study used ages from thousands of zircons, durable crystals that can indicate volcanic activity, to show that upticks in volcanic emissions were the dominant force driving the planet's warm periods. It's likely both teams have at least one hand on the truth, adds Lee, who contributed to the zircon paper.

The beauty of his team's model, Macdonald said at the end of his talk, is that it explains not just why glacial times start, but also why they stop. A hothouse Earth appears to be the planet's default state, prevailing for three-fourths of the past 500 million years. An Indonesia-style collision may push the global climate into a glacial period, but only for a while. Mountains erode and continents drift. And the planet warms again. ■

NEUROSCIENCE

Computers turn neural signals into speech

Fed data from invasive brain recordings, algorithms reconstruct heard and spoken sounds

By **Kelly Servick**

For many people who are paralyzed and unable to speak, signals of what they'd like to say hide in their brains. No one has been able to decipher those signals directly. But three research teams recently made progress in turning data from electrodes surgically placed on the brain into computer-generated speech. Using computational models known as neural networks, they reconstructed words and sentences that were, in some cases, intelligible to human listeners.

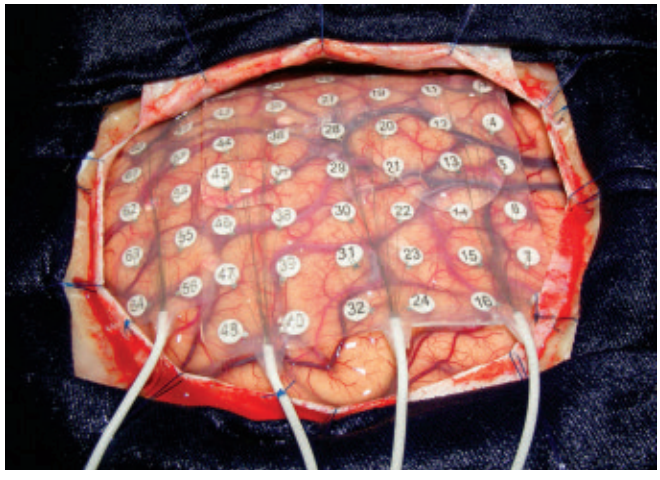
None of the efforts, described in papers in recent months on the preprint server bioRxiv, managed to re-create speech that people had merely imagined. Instead, the researchers monitored parts of the brain as people either read aloud, silently mouthed speech, or listened to recordings. But showing the reconstructed speech is understandable is "definitely exciting," says Stephanie Martin, a neural engineer at the University of Geneva in Switzerland who was not involved in the new projects.

People who have lost the ability to speak after a stroke or disease can use their eyes or make other small movements to control a cursor or select on-screen letters. (Cosmologist Stephen Hawking tensed his cheek to trigger a switch mounted on his glasses.) But if a brain-computer interface could re-create their speech directly, they might regain much more: control over tone and inflection, for example, or the ability to interject in a fast-moving conversation.

The hurdles are high. "We are trying to work out the pattern of … neurons that turn on and off at different time points, and infer the speech sound," says Nima Mesgarani, a computer scientist at Columbia University. "The mapping from one to the other is not very straightforward." How these signals translate to speech sounds varies from person to person, so computer models must be "trained" on each individual. And the models do best with extremely precise data, which requires opening the skull.

Researchers can do such invasive recording only in rare cases. One is during the removal of a brain tumor, when electrical readouts from the exposed brain help surgeons locate and avoid key speech and motor areas. Another is when a person with epilepsy is implanted with electrodes for several days to pinpoint the origin of seizures before surgical treatment. "We have, at maximum, 20 minutes, maybe 30," for data collection, Martin says. "We're really, really limited."

The groups behind the new papers made the most of precious data by feeding the information into neural networks, which process complex patterns by passing information through layers of computational "nodes." The networks learn by adjusting connections between nodes. In the experiments, networks were exposed to recordings of speech that a person produced or heard and data on simultaneous brain activity.



Epilepsy patients with electrode implants have aided efforts to decipher speech.

Mesgarani's team relied on data from five people with epilepsy. Their network analyzed recordings from the auditory cortex (which is active during both speech and listening) as those patients heard recordings of stories and people naming digits from zero to nine. The computer then reconstructed spoken numbers from neural data alone; when the computer "spoke" the numbers, a group of listeners named them with 75% accuracy.

Another team, led by neuroscientists Miguel Angrick of the University of Bremen in Germany and Christian Herff at Maastricht University in the Netherlands, relied on data from six people undergoing brain tumor surgery. A microphone captured their voices as they read single-syllable words aloud. Meanwhile, electrodes recorded from the brain's speech planning areas and motor areas, which send commands to the vocal tract to articulate words. The network mapped electrode readouts to the audio recordings, and then reconstructed words from previously unseen brain data. According to a computerized scoring system, about 40% of the computer-generated words were understandable.

Finally, neurosurgeon Edward Chang and his team at the University of California, San Francisco, reconstructed entire sentences from brain activity captured from speech and motor areas while three epilepsy patients read aloud. In an online test, 166 people heard one of the sentences and had to select it from among 10 written choices. Some sentences were correctly identified more than 80% of the time. The researchers also pushed the model further: They used it to re-create sentences from data recorded while people silently mouthed words. That's an important result, Herff says—"one step closer to the speech prosthesis that we all have in mind."

However, "What we're really waiting for is how [these methods] are going to do when the patients can't speak," says Stephanie Riès, a neuroscientist at San Diego State University in California who studies language production. The brain signals when a person silently "speaks" or "hears" their voice in their head aren't identical to signals of speech or hearing. Without external sound to match to brain activity, it may be hard for a computer even to sort out where inner speech starts and ends.

Decoding imagined speech will require "a huge jump," says Gerwin Schalk, a neuroengineer at the National Center for Adaptive Neurotechnologies at the New York State Department of Health in Albany. "It's really unclear how to do that at all."

One approach, Herff says, might be to give feedback to the user of the brain-computer interface: If they can hear the computer's speech interpretation in real time, they may be able to adjust their thoughts to get the result they want. With enough training of both users and neural networks, brain and computer might meet in the middle. ■

PHOTO: WENHT/ISTOCK.COM

PHOTO: RON STEINER/ALAMY STOCK PHOTO

ECOLOGY

Do plants favor their kin?

Once considered outlandish, the idea that plants help their relatives is taking root

By **Elizabeth Pennisi**

For people, and many other animals, family matters. Consider how many jobs go to relatives. Or how an ant will ruthlessly attack intruder ants but rescue injured, closely related nestmates. There are good evolutionary reasons to aid relatives, after all. Now, it seems, family feelings may stir in plants as well.

A Canadian biologist planted the seed of the idea more than a decade ago, but many plant biologists regarded it as heretical—plants lack the nervous systems that enable animals to recognize kin, so how can they know their relatives? But with a series of recent findings, the notion that plants really do care for their most genetically close peers—in a quiet, plant-y way—is taking root. Some species constrain how far their roots spread, others change how many flowers they produce, and a few tilt or shift their leaves to minimize shading of neighboring plants, favoring related individuals.

"We need to recognize that plants not only sense whether it's light or dark or if they've been touched, but also whom they are interacting with," says Susan Dudley, a plant evolutionary ecologist at McMaster University in Hamilton, Canada, whose early plant kin recognition studies sparked the interest of many scientists.

Beyond broadening views of plant behavior, the new work may have a practical side. In September 2018, a team in China reported that rice planted with kin grows better, a finding that suggested family ties can be exploited to improve crop yields. "It seems anytime anyone looks for it, they find a kin effect," says André Kessler, a chemical ecologist at Cornell University.

From termites to people, kin-specific behaviors have evolved over and over in animals, showing there is a strong advantage to helping relatives pass on shared genes. Dudley reasoned that the same evolutionary forces should apply to plants. Not long after researchers proved that plants can distinguish "self" from "nonself" roots, she tested whether they could also pick out and favor kin. She grew American searocket (*Cakile edentula*), a succulent found on North American beaches, in pots with relatives or with unrelated plants from the same population. With strangers, the searocket greatly expanded its underground root system, but with relatives, it held these competitive urges in check, presumably leaving more room for kin roots get nutrients and water. The claim, published in 2007, shocked colleagues. A few sharply criticized the work, citing flawed statistics and bad study design.

Since then, however, other researchers have confirmed her findings. Recently, working with *Moricandia moricandioides*, a Spanish herb, Rubén Torices and his colleagues at the University of Lausanne in Switzerland and the Spanish National Research Council demonstrated cooperation in flowering. After growing 770 seedlings in pots either alone or with three or six neighbors of varying relatedness, the team found

Sunflowers are among a number of plants that appear to recognize and help kin.

the plants grown with kin put out more flowers, making them more alluring to pollinators. The floral displays were especially big in plants in the most crowded pots of relatives, Torices and his colleagues reported on 22 May 2018 in *Nature Communications*.

Torices, now at King Juan Carlos University in Madrid, calls the kin effects "altruistic" because each individual plant gives up some of its ultimate seedmaking potential to expend more energy making flowers. In the end, he suspects, more seeds are fertilized overall in the closely related pots.

Doubts linger. Is a plant identifying genetic kin, or simply recognizing that its neighbor is more or less similar to itself? "I do not think that there has been convincing evidence for kin recognition in plants yet," says Hélène Fréville, a population biologist studying crops at the Montepellier outpost of the French National Institute for Agricultural Research.

Sagebrush bushes (*Artemisia tridentata*) have provided some strong clues, however. When injured by herbivores, these plants release volatile chemicals that stimulate neighboring sagebrush to make chemicals toxic to their shared enemies. Ecologist Richard Karban at the University of California, Davis, wondered whether kin were preferentially warned. His group had already found that sagebrush plants roughly fall into two "chemotypes," which mainly emit either camphor or another organic compound called thujone when their leaves are damaged. The team showed that the chemotypes are heritable, making them a potential kin recognition signal. In 2014, the researchers reported that when volatiles from a plant of one chemotype were applied to the same type of plant, those plants mounted stronger antiherbivore defenses and had much less

insect damage than when the volatiles were applied to a plant of the other chemotype—a hint of a kin effect.

The mustard *Arabidopsis thaliana* has provided another clue. About 8 years ago, Jorge Casal, a plant biologist at the University of Buenos Aires, noticed that *Arabidopsis* plants growing next to relatives shift the arrangement of their leaves to reduce shading of their neighbors, but don't do that when the neighbors are unrelated. How they sense the presence of relatives was a mystery, however.

The plants do have light sensors, and in 2015, Casal's team discovered that the strength of reflected light striking nearby leaves signaled relatedness and triggered the rearrangements. Relatives tend to sprout leaves at the same height, bouncing more light onto each other's leaves. By shifting leaves to reduce how much they shade each other, the relatives cumulatively grow more vigorously and produce more seeds, his team found. "There is no other case of kin recognition in plants where the cue, the receptors, and the fitness consequences have been established," Casal says.

Since then, he has shown that when sunflower kin are planted close together, they, too, arrange themselves to stay out of one another's way. The sunflowers incline their shoots alternately toward one side of the row or the other, Casal and his colleagues reported in 2017 in the *Proceedings of the National Academy of Sciences*. Taking advantage of the effect, they planted 10 to 14 related plants per square meter—an unheard of density for commercial growers—and got up to 47% more oil from plants that were allowed to lean away from each other than plants forced to grow straight up.

Chui-Hua Kong, a chemical ecologist at the China Agricultural University in Beijing, is exploiting a similar effect to boost rice yields. His lab studies rice varieties that give off weed-killing chemicals in their roots. Right now, they don't have high enough yields to replace commonly grown varieties that require herbicides. But in 3-year-long field tests, kin-recognizing versions of these self-protective rice varieties produced a 5% increase in yield when grown with kin, rather than unrelated plants, Kong and colleagues reported in late September 2018 in *New Phytologist*. To test the approach on a larger scale, he and his colleagues are planting "kin" seedlings of the weed-killing strain together in paddy fields in South China.

Brian Pickles, an ecologist at the University of Reading in the United Kingdom, proposes that kin recognition could even help forests regenerate. By tracing flows of nutrients and chemical signals between trees connected by underground fungi, he showed that the firs preferentially feed their kin and warn them about insect attacks. The finding suggested a family of firs would grow faster than unrelated trees.

To some biologists, the emerging picture of communicating, cooperating plants is still based on thin evidence. Laurent Keller, an evolutionary biologist at the University of Lausanne who has shown that some apparent signs of kin recognition in *Arabidopsis* can instead stem from innate differences among the plants, calls for more rigor in studies. "People have started to realize that it is important to think carefully about the design of the experiment to rule out other potential explanations," he says.

Keller is keeping an open mind and predicts stronger evidence of plant kin recognition will emerge. Karban is already convinced. "We are learning that plants are capable of so much more sophisticated behavior than we had thought," he says. "It's really cool stuff." ■

Initially disbelieved, Susan Dudley's work on plant kin recognition is winning over more biologists.

PLANETARY SCIENCE

Asteroid mission faces 'breathtaking' touchdown

As first data roll in from Hayabusa2, engineers plan descent to rocky surface

By **Dennis Normile**, *in Yonago, Japan*

Japan's Hayabusa mission made history in 2010 for bringing back to Earth the first samples ever collected on an asteroid. But the 7-year, 4-billion-kilometer odyssey was marked by degraded solar panels, innumerable mechanical failures, and a fuel explosion that knocked the spacecraft into a tumble and cut communications with ground control for 2 months. When planning its encore, Hayabusa2, Japan's scientists and engineers were determined to avoid such drama. They made components more robust, enhanced communications capabilities, and thoroughly tested new technologies.

But the target asteroid, Ryugu, had fresh surprises in store. "By looking at the details of every asteroid ever studied, we had expected to find at least some wide flat area suitable for a landing," says Yuichi Tsuda, Hayabusa2's project manager at the Japan Aerospace Exploration Agency's Institute of Space and Astronautical Science (ISAS), which is headquartered in Sagamihara. Instead, when the spacecraft reached Ryugu in June 2018—at 290 million kilometers from Earth—it found a cragged, cratered, boulder-strewn surface that makes landing a daunting challenge. The first sampling touchdown, scheduled for October, was postponed until at least the end of this month, and at a symposium here on 21 and 22 December, ISAS engineers presented an audacious new plan to make a pinpoint landing between closely spaced boulders. "It's breathtaking," says Bruce Damer, an origins of life researcher at the University of California, Santa Cruz.

Yet most everything else has gone according to plan since Hayabusa2 was launched in December 2014. Its cameras and detectors have already provided clues to the asteroid's mass, density, and mineral and

PHOTO: TASMIN CHU

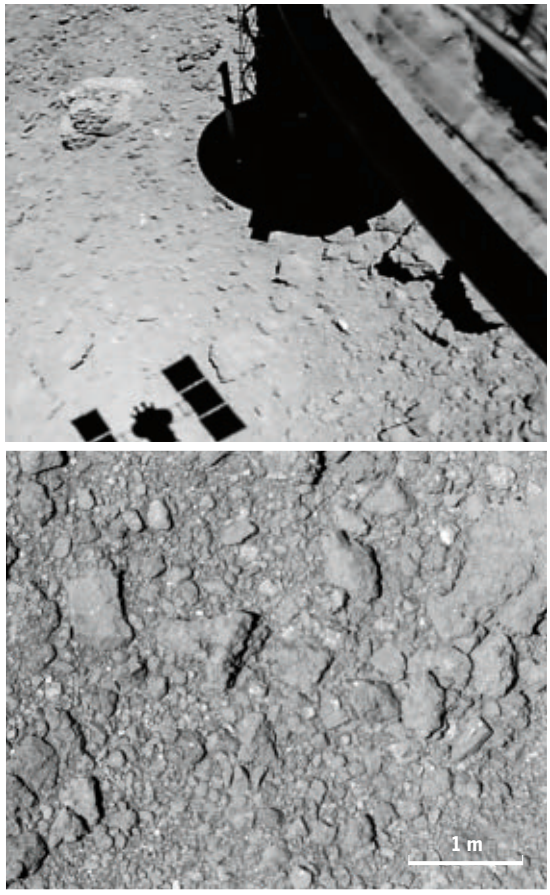
elemental composition, and three rovers dropped on the asteroid have examined the surface. At the symposium, ISAS researchers presented early results, including evidence of an abundance of organic material and hints that the asteroid's parent body once held water. Those findings "add to the evidence that asteroids rather than comets brought water and organic materials to Earth," says project scientist Seiichiro Watanabe of Nagoya University in Japan.

Ryugu is 1 kilometer across and 900 meters top to bottom, with a notable bulge around the equator, like a diamond. Visible light observations and computer modeling suggest it's a porous pile of rubble that likely agglomerated dust, rocks, and boulders after another asteroid or planetesimal slammed into its parent body during the early days of the solar system. Ryugu spins around its own axis once every 7.6 hours, but simulations suggest that during the early phase of its formation, it had a rotation period of only 3.5 hours. That probably produced the bulge, by causing surface landslides or pushing material outward from the core, Watanabe says. Analyzing surface material from the equator in an Earth-based laboratory could offer support for one of those scenarios, he adds. If the sample has been exposed to space weathering for a long time, it was likely moved there by landslides; if it is relatively fresh, it probably migrated from the asteroid's interior.

So far, Hayabusa2 has not detected water on or near Ryugu's surface. But its infrared spectrometer has found signs of hydroxide-bearing minerals that suggest water once existed either on the parent body or on the asteroid, says Mutsumi Komatsu, a planetary materials scientist at the Graduate University for Advanced Studies in Hayama, Japan. The asteroid's high porosity also suggests it once harbored significant amounts of water or ice and other volatile compounds that later escaped, Watanabe says. Asteroids such as Ryugu are rich in carbon as well, and they may have been responsible for bringing both water and carbon, life's key building block, to a rocky Earth early in its history. (Comets, by contrast, are just 3% to 5% carbon.)

Support for that theory, known as the late heavy bombardment, comes from another asteroid sample return mission now in progress. Early last month, NASA's OSIRIS-REx reached asteroid Bennu, which is shaped like a spinning top as well and, the U.S. space agency has reported, has water trapped in the soil. "We're lucky to be able to conduct comparative studies of these two asteroid brothers," Watanabe says.

Geologist Stephen Mojzsis of the University of Colorado in Boulder is not convinced such asteroids will prove to be the source of Earth's water; there are other theories, he says, including the possibility that a giant Jupiter-like gaseous planet migrated from the outer to the inner solar system, bringing water and other molecules with it around the time Earth was formed. Still, findings on Ryugu's shape and composition "scientifically, could be very important," he says.



Hayabusa2 imaged its shadow during a rehearsal descent (top). A close-up shows a surface strewn with boulders (bottom).

Some new details come from up-close looks at the asteroid's surface. On 21 September, Hayabusa2 dropped a pair of rovers the size of a birthday cake, named Minerva-II1A and -II1B, on Ryugu's northern hemisphere. Taking advantage of its low gravity to hop autonomously, they take pictures that have revealed "microscopic features of the surface," Tsuda says. And on 5 October, Hayabusa2 released a rover developed by the German and French space agencies that analyzed soil samples in situ and returned additional pictures.

The ultimate objective, to bring asteroid samples back to Earth, will allow lab studies that can reveal much more about the asteroid's age and content. ISAS engineers programmed the craft to perform autonomous landings, anticipating safe touchdown zones at least 100 meters in diameter. Instead, the biggest safe area within the first landing zone turned out to be just 12 meters wide.

That will complicate what was already a nail-biting operation. Prior to each landing, Hayabusa2 planned to drop a small sphere sheathed in a highly reflective material to be used as a target, to ensure the craft is moving in sync with the asteroid's rotation. Gravity then pulls the craft down gently until a collection horn extending from its underside makes contact with the asteroid; after a bulletlike projectile is fired into the surface, soil and rock fragments hopefully ricochet into a catcher within the horn. For safety, the craft has to steer clear of rocks larger than 70 centimeters.

During a rehearsal in late October, Hayabusa2 released a target marker above the 12-meter safe circle; unfortunately, it came to rest more than 10 meters outside the zone. But it is just 2.9 meters away from the edge of a second possible landing site that's 6 meters in diameter. Engineers now plan to have the craft first hover above the target marker and then move laterally to be above the center of one of the two sites. Because the navigation camera points straight down, the target marker will be outside the camera's field of view as Hayabusa2 descends, leaving the craft to navigate on its own.

"We are now in the process of selecting which landing site" to aim for, says Fuyuto Terui, who is in charge of mission guidance, navigation, and control. Aiming at the smaller zone means Hayabusa2 can keep the target marker in sight until the craft is close to the surface; the bigger zone gives more leeway for error, but the craft will lose its view of the marker earlier in the descent.

Assuming the craft survives the first landing, plans call for Hayabusa2 to blast a 2-meter-deep crater into Ryugu's surface at another site a few months later, by hitting it with a 2-kilogram, copper projectile. This is expected to expose subsurface material for observations by the craft's cameras and sensors; the spacecraft may collect some material from the crater as well, using the same horn device. There could be a third touchdown, elsewhere on the asteroid. If all goes well, Hayabusa2 will make it back to Earth with its treasures in 2020. ■

PHOTOS: (TOP TO BOTTOM) JAXA, UNIVERSITY OF TOKYO, KOCHI UNIVERSITY, RIKKYO UNIVERSITY, NAGOYA UNIVERSITY, CHIBA INSTITUTE OF TECHNOLOGY, MEIJI UNIVERSITY, UNIVERSITY OF AIZU, AIST

FEATURES

BIOLOGY IN THE BANK

How an open-access trove of data on Britons is unlocking the genetics of disease, behavior, and physical traits

By **Jocelyn Kaiser** *and* **Ann Gibbons**

PHOTO: NIGEL HILIER

In early 2017, epidemiologist Rory Collins at the University of Oxford in the United Kingdom and his team faced a test of their principles. They run the UK Biobank (UKB), a huge research project probing the health and genetics of 500,000 British people. They were planning their most sought-after data release yet: genetic profiles for all half-million participants. Three hundred research groups had signed up to download 8 terabytes of data—the equivalent of more than 5000 streamed movies. That's enough to tie up a home computer for weeks, threatening a key goal of the UKB: to give equal access to any qualified researcher in the world.

"We wanted to create a level playing field" so that someone at a big center with a supercomputer was at no more of an advantage than a postdoc in Scotland with a smaller computer and slower internet link, says Oxford's Naomi Allen, the project's chief epidemiologist. They came up with a plan: They gave researchers 3 weeks to download the encrypted files. Then, on 19 July 2017, they released a final encryption key, firing the starting gun for a scientific race.

Within a couple of days, one U.S. group had done quick analyses linking more than 120,000 genetic markers to more than 2000 diseases and traits, data it eventually put up on a blog. Only 60,000 markers had previously been tied to disease, says human geneticist Eric Lander, president and director of the Broad Institute in Cambridge, Massachusetts. "[They] doubled that in a week."

Within 2 weeks, others had begun to post draft manuscripts on the bioRxiv preprint site. By now, those data have spawned dozens of papers in journals or on bioRxiv, firming up how particular genes contribute to heart disease, diabetes, Alzheimer's, and other conditions, as well as genes' role in shaping personality, depression, birth weight, insomnia, and other traits. More controversially, data from the trove also pointed to DNA markers linked to education level and sexual orientation, stoking long-running controversies about the application of genetics to behavior in people.

When the Manchester-based UKB enrolled its first volunteer 13 years ago, some critics wondered whether it would be a waste of time and money. But by now, any skepticism is long gone. "It's now clear that it has been a massive success—largely because the big data they have are being made widely available," says Oxford developmental neuropsychologist Dorothy Bishop, a participant. Other biobanks are bigger or collect equally detailed health data. But the UKB has both large numbers of participants and high-quality clinical information. It "allows us to do research on a scale that we've never been able to do before," says Peter Visscher, a quantitative geneticist at the University of Queensland in Brisbane, Australia.

The crucial ingredient, however, may be open access. Researchers around the world can freely delve into the UKB data and rapidly build on one another's work, resulting in unexpected dividends in diverse fields,

UK Biobank Principal Investigator Rory Collins stands amid stored biospecimens from the project's half-million participants.

such as human evolution (see story, p. 21). In a crowdsourcing spirit rare in the hypercompetitive world of biomedical research, groups even post tools for using the data without first seeking credit by publishing in a journal.

"The U.K. is getting all of the world's best brains" to study its citizens, says Ewan Birney, director of the EMBL European Bioinformatics Institute in Hinxton, U.K., and a member of the UKB's steering committee. The U.K. focus is also the project's chief downside, as it explores just one slice of humanity: northern Europeans. It holds data for only about 20,000 people of African or Asian descent, for example. Yet as new papers appear every few days, researchers say the UKB remains a shining example of the power of curiosity unleashed. "It's the thing we always dreamed of," Lander says.

THE UKB WAS ANNOUNCED in the early 2000s as a classical epidemiological study—the kind used to associate risk factors such as diet and smoking with the development of disease over time. The model was the famous Framingham Heart Study, a long-term study that initially analyzed 5200 residents of Framingham, Massachusetts, seeking factors that influence heart disease. The UKB project, which has received $308 million in funding so far from the Wellcome Trust medical charity, the U.K. government, and disease foundations, "was going to be like Framingham, only 100 times bigger," says principal investigator Collins.

From 2006 to 2010, the UKB enrolled 500,000 people aged 40 to 69 through the United Kingdom's National Health Service. Mailed invitations were sent widely, including to people in poor and ethnically diverse areas of cities such as Birmingham. But in the end, participants were "anybody you could persuade," Collins says. Investigators sampled their blood and urine, surveyed their habits, and examined them for more than 2400 different traits or phenotypes, including data on their social lives, cognitive state, lifestyle, and physical health.

The blood samples yielded DNA for genomic analyses. Links to other U.K. databases added information such as cancer diagnoses, deaths, and hospitalizations. "If you're talking about common phenotypes, the Biobank shines," Lander says. "There's arm fat, smoking behavior, miserableness, neurotic behavior, time on your computer, eating behavior, drinking behavior."

Other biobanks have comparably rich health data, such as deCODE Genetics's detailed database on Iceland's population and biobanks run by U.S. health care providers. Some, such as the U.S. Million Veteran Program and the DNA testing company 23andMe, are bigger. But in most cases researchers can use these databases only by collaborating with their creators.

In contrast, the Wellcome Trust and U.K. Medical Research Council insisted that any researcher approved by the UKB board, anywhere in the world, be able to download anonymized data sets on all 500,000 participants. (Users pay a relatively modest fee of $2500 and agree to return their raw data, results, and code to the UKB after publishing. They also sign a legal agreement not to try to reidentify any participant.)

"It was a novel concept," says Collins, who says he's lost track of the times someone has asked him after a talk whether he's interested in collaborating. "I have to say, 'You just request the data.' To some extent people don't believe it."

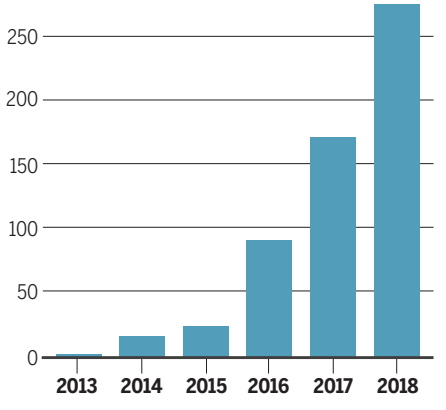
The aim is to maximize the scientific payoff: "By making data available to 100 people around the world, we can get a lot more research done than if I sit here and do one study a year with the data," he says.

In 2015, his team released the first batch of genetic data on a subset of 150,000 participants. Then came the July 2017 release of full genotyping data for all 500,000. Two months later, Benjamin Neale's group at the Broad Institute put up its blog doubling the number of markers linked to traits and disorders, as well as a web browser for looking up specific markers. "We viewed it as a service to the community," Neale says.

TODAY, about 7000 researchers have registered to use UKB data on 1400 projects, and nearly 600 papers have been published. Some studies simply link behaviors and disease, for example reporting that drinking more coffee can reduce mortality but that binge-watching TV is associated with more colon cancer. But most studies compare the genomes of people with some trait or disease with those without it, in order to home in on genes that influence that attribute; these projects are known as genome-wide association studies.

Engine of productivity

Published papers based on the UK Biobank's bounty of health and genetics data are piling up fast, in part because the data are freely available.



The result, every few days, is a new paper using UKB data to link particular gene variants to a disease or trait—arthritis, type 2 diabetes, depression, neuroticism, heart disease. "It's so easy for people who don't collect their own data," says statistical geneticist Danielle Posthuma of Vrije University in Amsterdam, who studies brain diseases. By combining data from the UKB and other collections, investigators can amass samples of a million people or more, amplifying the signal of gene variants with subtle effects. For some diseases, dozens or hundreds of genes appear to play a role. The genetic links are suggestive correlations; establishing cause and effect will take more genetics work and lab studies, which could reveal new disease pathways that might be drug targets.

CREDITS: (GRAPHIC) N. DESAI/*SCIENCE*; (DATA) UK BIOBANK

In the near term, the large sample sizes are boosting the power of "polygenic" risk scores, which calculate a person's disease risk by combining many genetic markers. For example, one study published in August 2018 in *Nature Genetics* drew on the July 2017 data to devise risk scores for five diseases, including breast cancer and heart disease. The authors, at Massachusetts General Hospital in Boston and the Broad Institute, found that a surprisingly high 8% of people of European descent have at least a threefold elevated risk for heart disease. And up to 6% have a threefold increase in risk for one of the four other diseases, suggesting they should be screened early and consider lifestyle changes or other measures that could improve their odds.

The most provocative studies have probed for genetic influences on human behavior. One, published in *Nature Genetics* in July 2018, drew on the UKB and 23andMe to pin down genetic contributions to a person's level of education. Together, 1300 genetic markers accounted for 11% of the variability among individuals, the researchers found. That's comparable to certain environmental influences in the UKB sample, such as family income, which predicted just 7% of the variance in educational attainment among participants; and mother's education level, which predicted 15%. Another study presented at a meeting last fall found four genetic markers that appear to have a strong influence on whether a person has had sex with someone of their own sex at least once (*Science*, 26 October 2018, p. 385).

Such studies are raising concerns that genetic tests could be used to screen embryos for desired traits or discriminate against individuals with certain genetic profiles. That would be a misuse of the findings, say the researchers who identified these links. They stress that the probabilities mean little on the individual level.

THE UKB'S UNUSUAL DESIGN does have some limitations. The big one: Ninety-four percent of participants are white. "It's really good if you're British or European," Lander says. But, "If you're an American without European ancestry or an African or Asian, you're going to be poorly serviced by the new polygenic risk scores." Nor will scores for traits such as educational attainment be meaningful in people with non-European ancestry.

The mailed invitation recruitment strategy didn't work as well as hoped, says Collins, who notes that young, low-income, white men are also scarce in the database. "We were aiming to get heterogeneity, but it's difficult."

Bishop blames the project's slant toward higher income, healthy, white people on a lack of incentives for participants—they don't get even a small payment or the promise of receiving their test results. The people attracted to the project were those with enough spare time to participate or "who [wanted] to help research," she says.

One problem is that many immigrants to the United Kingdom have little experience with the research world, says Naveed Sattar, an adviser to the UKB and a clinical researcher and epidemiologist at the University of Glasgow. "Most first generation Asians simply have no prior experience of what research is and that it may help their community and their children in the future," he says. Surveys have found that immigrants are often suspicious of participating in research—perhaps because of unethical past studies in some countries, or concern that genetic findings could be used to discriminate.

Engaging such groups is possible, says geneticist David Van Heel of Queen Mary University of London, who heads the Genes & Health study, which so far has enrolled 33,000 Britons of Bangladeshi and Pakistani ancestry. In his experience, South Asians in the United Kingdom are less likely to respond to mailed invitations. His project achieved success by approaching potential participants in person—sometimes in their native language—in "trusted" settings such as health clinics and community centers.

Collins and other geneticists hope other biobanks can help fill the gap. For example, the Wellcome Trust is now the main funder of the China Kadoorie Biobank, with data on 515,000 people from mainland China, belonging to 10 ethnic groups. In the United States, the All of Us biobank funded by the National Institutes of Health (NIH) aims to use community outreach to help enroll at least half of its 1 million participants from minority groups, and like the UKB, promises to make data freely available. The Human Heredity & Health in Africa initiative has 70,000 participants so far across the continent, with funding from NIH and the Wellcome Trust. "There are ways of fixing this up. But we've got a long ways to go," Birney says.

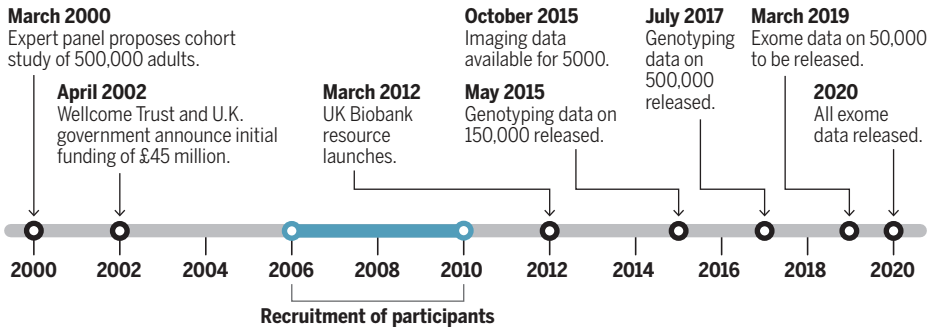
Meanwhile, the UKB's riches are growing. About half of the participants' primary care data, including clinical data and prescriptions, will become available next spring. The UKB has also done MRI scans of the brains, hearts, and abdomens of 25,000 participants, with plans to scan 100,000; researchers are examining and annotating the images.

Collins has been promoting the UKB's scientific treasure in Silicon Valley in California, where he hopes bioinformatics experts will dig in and come up with unexpected findings. The genetic data are ballooning, too: Several companies are now sequencing the exomes, or protein-coding regions, of all UKB participants, and the United Kingdom's public Sanger Institute is sequencing whole genomes from 50,000 volunteers. Unlike the genotyping data, which don't usually point to specific genes, the sequences will allow researchers who have found a genetic marker linked to a disease to quickly zero in on the causative gene and see the specific mutations at work.

Because of the $150 million cost of this sequencing work, the UKB had to compromise on open access: Companies have 9 to 12 months to use the exome data before they are made widely available. But Collins and his team, as well as geneticists around the world, are already gearing up for the wide release of the first batch of exome data on 50,000 participants. Again, they'll allow time for the download, then release a code. The starting gun in the next scientific race is set for March. ■

Long-term investment

Nearly 2 decades after U.K. funding organizations proposed a large, long-term health study, the database is paying off richly; its timeline is punctuated by massive, open-access data releases. Meanwhile, participants age and develop diseases, adding power and momentum to the project.

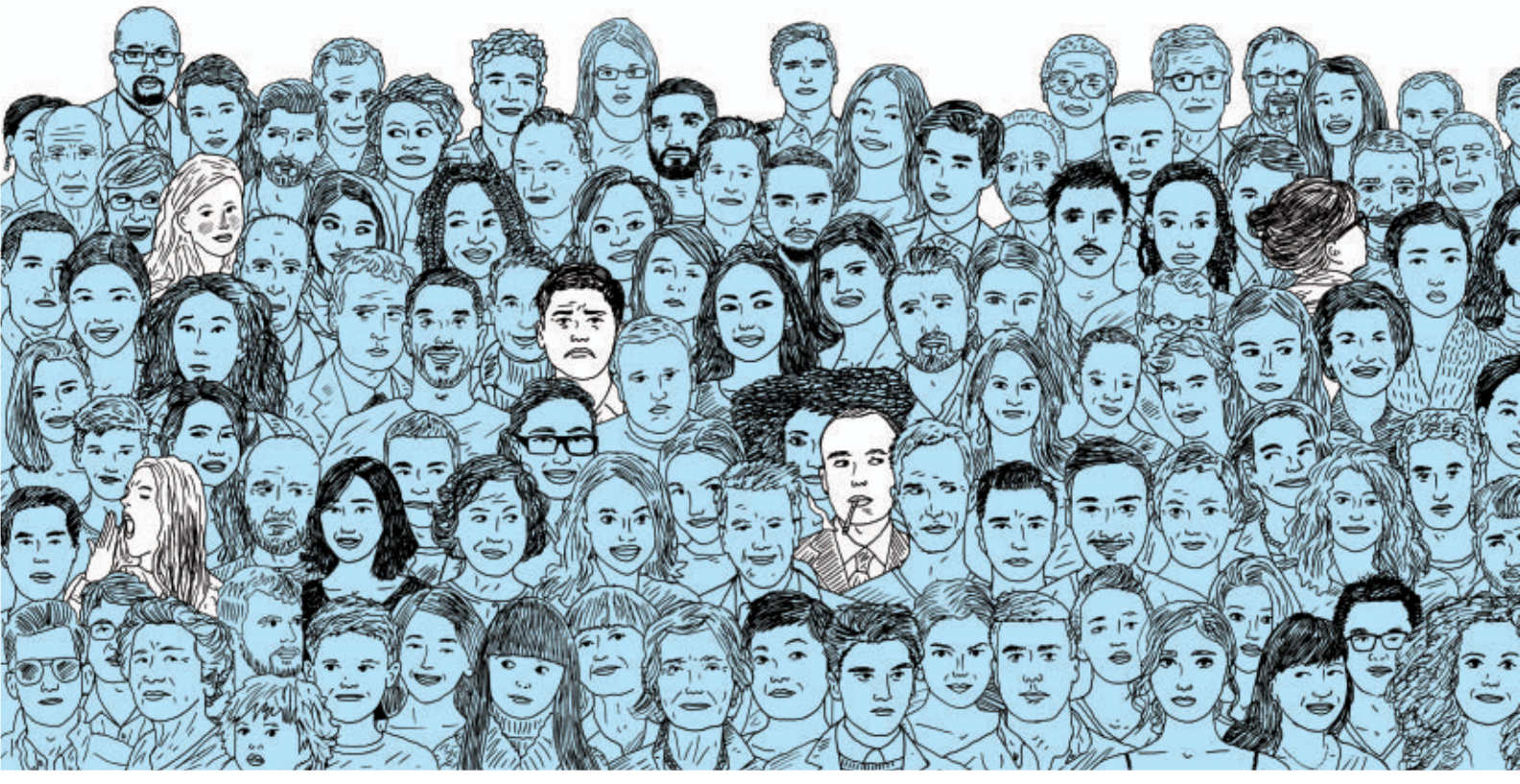


GRAPHIC: N. DESAI/*SCIENCE*

SPOTTING EVOLUTION AMONG US

The half-million people in the UK Biobank hold the genetic legacy of Neanderthals—and clues to how we are still evolving

By **Ann Gibbons**



Among participants in the UK Biobank are people whose Neanderthal DNA predisposes them to traits such as propensity to sunburn, staying up late, depression, smoking, and feeling lonely.

ILLUSTRATION: PETER ARKLE

Neanderthals are still among us, Janet Kelso realized 8 years ago. She had helped make the momentous discovery that Neanderthals repeatedly mated with the ancestors of modern humans—a finding that implies people outside of Africa still carry Neanderthal DNA today. Ever since then, Kelso has wondered exactly what modern humans got from those prehistoric liaisons—beyond babies. How do traces of the Neanderthal within shape the appearance, health, or personalities of living people?

For years, evolutionary biologists couldn't get their rubber-gloved hands on enough people's genomes to detect the relatively rare bits of Neanderthal DNA, much less to see whether or how our extinct cousins' genetic legacy might influence disease or physical traits.

But a few years ago, Kelso and her colleagues at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, turned to a new tool—the UK Biobank (UKB), a large database that holds genetic and health records for half a million British volunteers (see story, p. 18). The researchers analyzed data from 112,338 of those Britons—enough that "we could actually look and say: 'We see a Neanderthal version of the gene and we can measure its effect on phenotype in many people—how often they get sunburned, what color their hair is, and what color their eyes are,'" Kelso says. They found Neanderthal variants that boost the odds that a person smokes, is an evening person rather than a morning person, and is prone to sunburn and depression.

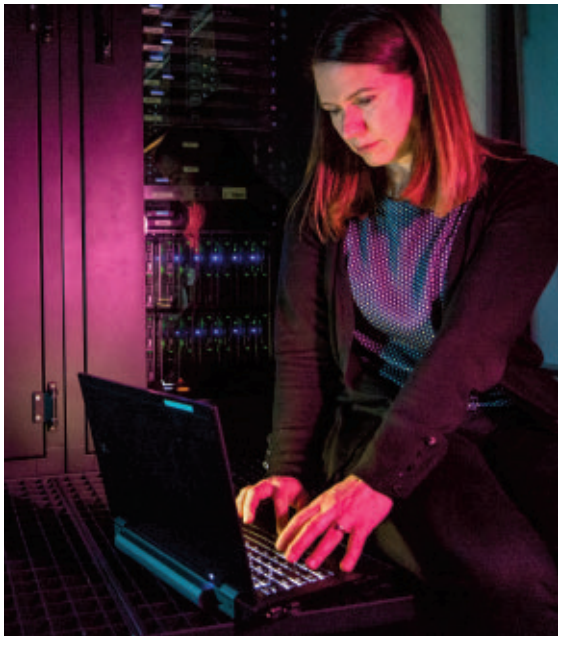
Kelso is one of many researchers who are turning troves of genetic and medical data on living people into windows on human evolution. In addition to unearthing archaic DNA, the studies are pinpointing genes that natural selection may now be winnowing out of the gene pool and other genes—for example those linked to fertility—that it may be favoring. Among the most fruitful of the data sources is the UKB, which makes its data accessible to researchers, no matter where they are and what their field. Its giant database is "a magical new resource that [will] help us answer a whole bunch of hard questions we're struggling with now because all of the data has been under lock and key," says population geneticist Jeremy Berg, a postdoc at Columbia University. "It is a step beyond other databases."

For the UKB architects, who designed it for biomedical research, the evolutionary discoveries are an unexpected bonus. "No one was thinking about Neanderthal traits when we designed the protocol," says molecular epidemiologist Rory Collins of the University of Oxford in the United Kingdom, who is principal investigator of the UKB. "The experiment [is] working well beyond people's expectations."

NEANDERTHALS SNEAKED INTO the UKB in 2013, when Harvard University population geneticist David Reich was in Oxford to give a talk. His host, Oxford geneticist Peter Donnelly, was overseeing the design of chips to identify genes of interest in blood samples like those in the UKB. Donnelly asked Reich whether he'd like to add Neanderthal variants to a custom chip used to genotype the UKB participants; that would allow Reich and others to fish for rare Neanderthal variants in half a million people. "David was very enthusiastic," Donnelly recalls.

Soon after, Reich and his postdoc, Sriram Sankararaman, emailed Donnelly a wish list of variants to add to the chip: 6000 relatively rare alleles likely to come from Neanderthals. Their calculations suggested the UKB was big enough to include enough carriers of these variants so researchers could probe the function of the genes. "Imagine 1% of the population has a Neanderthal variant," says Sankararaman, now a computational geneticist at the University of California (UC), Los Angeles. "If you're looking at half a million people, you're looking at enough copies of that variant in enough individuals [5000] so you can detect subtle effects."

At the same time, computational biologist Tony Capra at Vanderbilt University in Nashville had the same bright idea to search for Neanderthal DNA in a large database. He used proprietary electronic records of 28,000 Americans. His team was the first to publish, reporting Neanderthal DNA variants that raise the risk of depression, skin lesions, blood clots, and other disorders in people today (*Science*, 12 February 2016, p. 648). Inspired by Capra's study, Kelso jumped in, becoming the first to use UKB data to publish Neanderthal gene variants in living people. Her results suggest that although some Neanderthal gene variants may have been optimal for active lives outdoors in prehistoric Europe, they may be problematic for people now, who live mostly indoors in artificial light and get less exercise.



Janet Kelso fished for Neanderthal gene variants in the UK Biobank.

Groups led by Kelso and Sankararaman are now looking for links between Neanderthal DNA and traits in genotyped data from 500,000 people—the total UKB data set, which was released in July 2017. Already, they are learning that Neanderthal alleles help cause baldness and mental illness and boost certain immune functions, Sankararaman says. Meanwhile, another team has found variants that help explain why modern humans' heads are round, in contrast to the elongated, football-like shape of Neanderthal skulls (*Science*, 14 December 2018, p. 1229). Those researchers plan to combine forthcoming MRI brain scans of 100,000 UKB participants with genetic data to probe the genetic basis of brain differences between us and our extinct cousins.

Capra says when it comes to scanning and understanding DNA from Neanderthals, the UKB cohort offers even more analytical power than the medical databases he used, because it covers "a broader range of psychiatric and lifestyle traits." Those rich data have also made the UKB a hunting ground for clues to evolutionary changes that have shaped people's genomes in the past few generations—and may even be doing so today.

A FEW YEARS AGO, Molly Przeworski of Columbia University and Joe Pickrell of the New York Genome Center in New York City met for lunch near Columbia's campus. Talk turned to aging and Alzheimer's disease. Pickrell had been writing a blog, where he had discussed studies showing that between the ages of 70 and 85, carriers of the *ApoE4* allele, which boosts the risk of Alzheimer's and cardiovascular disease, died at about twice the rate of noncarriers. The pair wondered whether other gene variants affect survival so dramatically—and whether natural selection is weeding them out.

When it comes to natural selection in humans, most studies have only been able to detect dramatic cases thousands or millions of years ago in genes of known function. Now, Pickrell and Przeworski wondered whether they could detect genetic variants that affect survival today—and whether natural selection in recent generations has been weeding out harmful ones or favoring beneficial ones.

To do this, they realized they'd need data on DNA as well as on traits like participants' age at death. For statistical confidence, they'd need a giant sample size—at least 100,000—to detect how the frequency of common alleles varied in people of different ages. Databases like the UKB were the answer. "We suddenly realized that the some of these databases were large enough to let us study selection in contemporary humans," Przeworski says.

They soon got access to genetic and health data on 57,696 people in the Resource for Genetic Epidemiology Research on Aging database at Kaiser Permanente in Oakland, California, and 117,648 individuals in the UKB's 2015 data release. They sorted participants into 5-year age intervals, and looked at the frequency of many alleles, including *ApoE4*, in each age group, as well as how the variants correlated with 42 traits potentially associated with early death or long life, such as cardiovascular disease, cholesterol levels, asthma, age at puberty, and menopause.

Nearly all the variants they examined persisted at the same frequency even into old age, suggesting they had no large effect on survival. That implies natural selection has efficiently weeded out harmful variants, even if they act only in old age—

PHOTO: RONNY BARR/MAX PLANCK INSTITUTE FOR EVOLUTIONARY ANTHROPOLOGY

perhaps, Przeworski speculates, because the variants curb older men's fecundity. Or perhaps the hypothesized benefit that healthy grandmothers confer on grandchildren was at work.

The researchers did find two genes that suddenly became rare at older ages, suggesting they were harmful. One was *ApoE4*: As expected, fewer carriers—especially women—lived past age 80. Also, fewer men with a variant of the *CHRN3* gene that makes it harder to quit smoking survived past the age of 75 than did men without the variant.

The researchers concluded that natural selection has not yet had time to eliminate these two alleles, perhaps because changes in the environment and human behavior only recently made them deadly (*Science*, 20 May 2016, p. 876). For example, the *CHRN3* allele wouldn't have affected survival until many men were smoking. And women who were more active in the past might have been less vulnerable to the cardiovascular diseases caused by *ApoE4*, Przeworski speculates.

The researchers spotted another intriguing pattern. Genetic variants that lead to early puberty also became rarer in older age groups. Natural selection may have preserved those variants even though they shorten life span because they also boosted fertility.

A LONG LIFE, though, is much less important to evolution than fertility. When it comes to the game of evolution, in fact, the person who has the most kids wins by passing on the most genes. With the advent of birth control, people in industrial societies have more control than ever over their own fertility—but new studies zeroing in on the genes underlying fertility show the forces of selection may still be at work.

Multiple studies have suggested that when food sources became more reliable in industrialized societies, women began to mature faster, weigh more, give birth to their first child earlier, and enter menopause later—all traits possibly linked to having more babies. But researchers have been unable to tie those trends to underlying genes to get direct evidence of natural selection. Quantitative geneticist Peter Visscher and his colleagues at the University of Queensland in Brisbane, Australia, realized they could use the UKB to see firsthand which gene variants underlie those traits in people today, and whether they are really linked to fertility.

They searched the UKB's full cohort for people who had the most babies to see what traits they share, and what genes correlate with those traits. They documented the number of live births for women over age 45 and men over age 55. Then, they analyzed traits in women and men that might have influenced fertility, such as age of first birth, age of menopause, height, weight, body mass, blood pressure, and education. They found 23 traits in women and 21 in men linked to having more children. Not surprisingly, mothers who gave birth early and had late menopause—and therefore had a longer reproductive span—were more fertile. So were women who were heavier and shorter, perhaps because shorter bodies are more energy efficient, leaving a bigger reserve for pregnancy and nursing.

Visscher and his colleagues then set out to identify the genetic basis of these fertility-linked traits. They analyzed data from 157,807 of the women and 115,902 of the men. As predicted, they found that the most fertile women had higher frequencies of alleles that tend to make them shorter and heavier. In men, greater fertility was associated with more alleles that contribute to a higher body mass index and hand-grip strength. That suggests men with genes that make them taller and bulkier have more kids than sedentary types, whether because of female choice, some health-related reason, or the men's own preference.

"No one was thinking about Neanderthal traits when we designed the protocol."

Rory Collins, University of Oxford

Not all traits linked to fertility are physical or likely to have a big genetic component: Among women who had their first child later in life, those who had more education and did better on an intelligence test had more babies. This may be because better-educated couples tend to be wealthier and can afford more children.

But the fact that genes linked to traits thought to increase fertility are indeed more common in fertile people backs up the idea of recent selection on our genomes, even as both the environment and humans' preferences for mates and families are changing. "The UK Biobank allows us to show that natural selection not only took place in the past, but it's still ongoing," Visscher says.

TEASING OUT natural selection from other factors shaping genes can be tricky, however, especially when multiple genes work together to influence complex traits, such as height. About 5000 gene variants simultaneously influence a person's height, some boosting it, some reducing it, says Jian Yang, a statistical geneticist at the University of Queensland. The UKB's huge database allows researchers to find new variants and explore their impact and origins.

Using other databases, researchers had found that the number of genes that contribute to tallness in Europeans increased on a cline from south to north. Many researchers, including Berg, had concluded that northern Europeans had inherited those genes from an ancient migration—that of the Yamnaya herders who migrated from the Eurasian steppe to central Europe about 4000 years ago. Berg and others suggested natural selection had favored tallness in the Yamnaya or their ancestors, and ancient DNA reveals that the Yamnaya were tall.

But now, with UKB data, population geneticist Graham Coop of UC Davis and his colleagues, including Berg, are challenging that finding. In a bioRxiv preprint posted in June 2018, they analyzed genetic and height data on 500,000 people from the 2017 UKB data release. With so many people from similar backgrounds, the researchers could identify more height alleles, as well as note differences in diet, disease, and the environment. They found that northerners had no more tall variants than southerners.

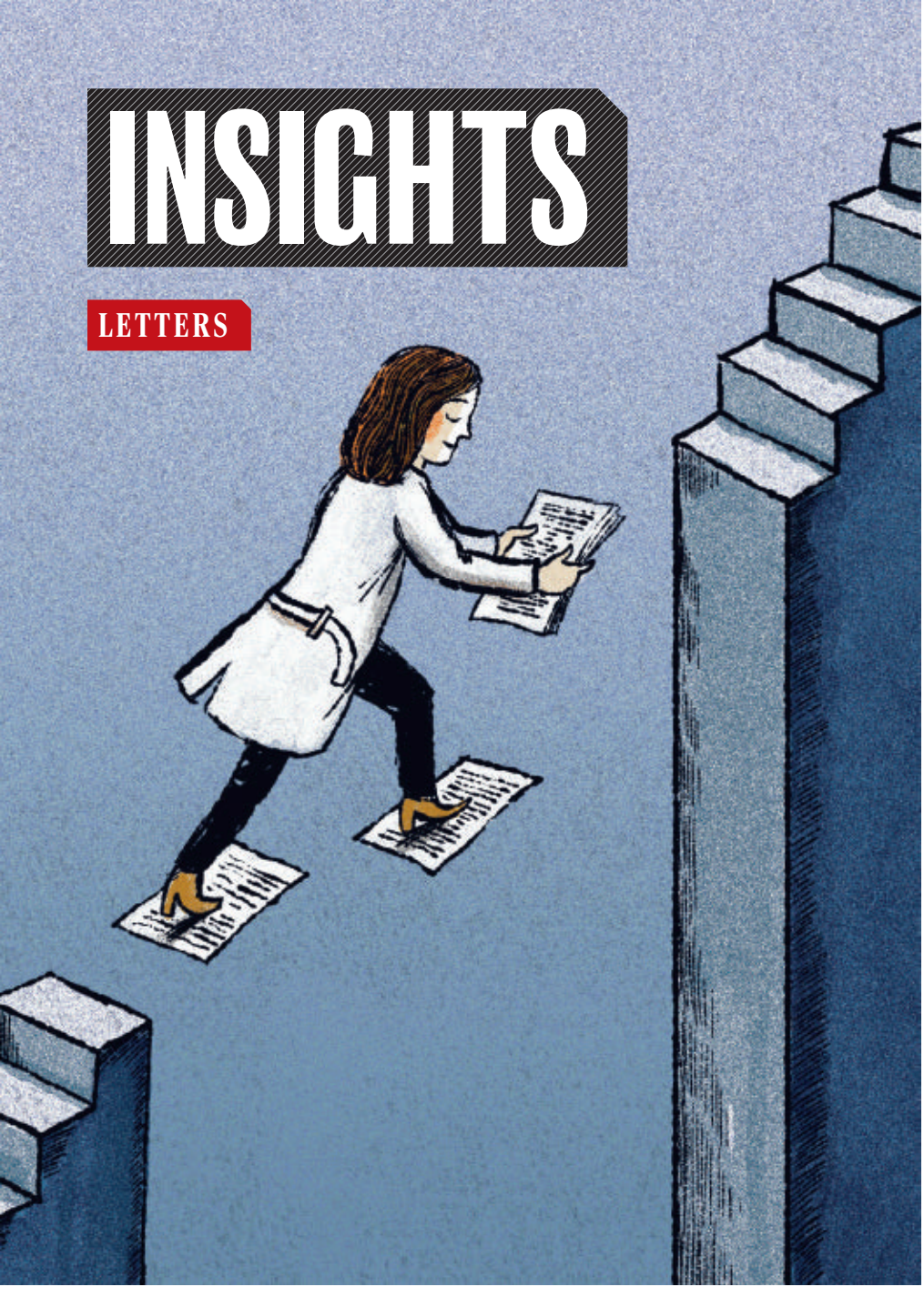
"It's true people in northern Europe are taller on average, but there is no evidence this has anything to do with natural selection," Berg says. He speculates that northerners' height might be an environmental effect, perhaps from a diet richer in protein, or from fewer childhood or prenatal illnesses.

Although UKB data cast doubt on natural selection's role in that case, they do suggest that evolution has favored genes for shortness in pygmy populations on the island of Flores in Indonesia. Visscher and colleagues scanned the DNA of Flores people for genes the UKB had linked to short stature. They found that Flores pygmies carry more such gene variants than their closest relatives in New Guinea and East Asia, suggesting evolution favored genes for shortness on the island (*Science*, 3 August 2018, p. 439). All these studies have generated "huge buzz among evolutionary biologists about how biobanks can provide very deep information about the genetics of different populations and their evolution," Kelso says.

She hopes to work with researchers designing databases in Africa and Asia to identify archaic DNA in those populations. Thanks to the success of the Neanderthal work, many researchers are eager for data from Melanesians, because they have inherited traces of DNA from Denisovans—the mysterious cousins of Neanderthals who lived in Siberia more than 50,000 years ago. "That would be amazing, to get Denisovan DNA from more living people [in biobanks]. That's our dream," Kelso says. ■

INSIGHTS

LETTERS



NEXTGEN VOICES

Challenging transitions

We asked young scientists these questions: **Have you ever encountered a particularly stark difference between an old and new position in your education or career? What was the difference between the positions, and what advice would you give to someone making a similar transition?** Here, respondents share the challenges they faced when they took on new responsibilities and roles, changed fields, or moved to new places. To others in similar situations, they advise: Be confident, prepared, and patient; communicate; and always ask for help when needed. —**Jennifer Sills**

Be prepared

The greatest challenge during my transition from a teaching assistant to an assistant professor in a large public university was the teaching responsibility. Whereas teaching assistants focus only on the subject matter and a small group of students, a professor must select textbooks, prepare syllabi, coordinate laboratory experiments, teach large classes, handle teaching assistants, manage the course website, and accommodate athletes' schedules and students with disabilities. By taking courses on teaching, I developed skills in communication, evaluation and assessment, education psychology, academic advising, and student accommodations. For me, preparation was the key to a smooth transition.

Niluka D. Wasalathanthri
Department of Chemistry, University of Connecticut, Storrs, CT 06269, USA.
Email: niluka.wasalathanthri@uconn.edu

Postdocs are expected to perform as experienced researchers. This can be challenging given the dependence of Ph.D. students on their supervisors for scientific ideas and experimental designs. To excel as a postdoc, I advise thoroughly reviewing the literature. Read the relevant papers completely, especially the methods sections. This will greatly enhance your ability to design your own experiment.

Syed Shan-e-Ali Zaidi
Plant Genetics Lab, Gembloux Agro-Bio Tech, University of Liège, Gembloux, 5030 Namur, Belgium.
Email: shan.e.ali@outlook.com

After juggling my experiments and managing a small lab of 10 members as a graduate student, I now focus solely on meeting the needs of a 75-person lab. Although I work far fewer hours now, I must work much faster in a shorter time frame. My excellent mentors and my management experience eased my transition. To succeed as a lab manager, I advise others to get as much regulatory, personnel management, and ordering experience as possible.

Elena Mahrt
Center for the Genetics of Host Defense, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA.
Email: elena.mahrt@utsouthwestern.edu

As a scientist, I could focus on research, but my transition to a professor role came with new responsibilities. Instead of simply reading a publication to plan new experiments, I now read with an eye toward how to explain the concepts to a student. A scientist might manage a group of 5 to 10 researchers, whereas a teacher

ILLUSTRATION: ROBERT NEUBECKER

manages students ranging from undergraduates to postdoctoral fellows. As a scientist, I could remain silent and work out problems internally. As a teacher, I have to talk constantly, yet remain calm.

Sudhakar Srivastava
Institute of Environment and Sustainable Development, Banaras Hindu University, Varanasi, Uttar Pradesh 221005, India.
Email: sudhakar.srivastava@gmail.com

The training of a physician focuses on the familiarity with medical knowledge and clinical guidelines, whereas solid statistics and a programming background are required to become a data scientist. To make the transition, I joined a Ph.D. program after medical school; spent 3 years taking classes on statistical inference, machine learning, and computational biology; and participated in programming contests with undergraduates. As a result, I benefited from both the medical domain knowledge and the quantitative skills I learned in my journey.

Kun-Hsing Yu
Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA.
Email: kun-hsing_yu@hms.harvard.edu

Find community

When I transitioned from one field in biology to another, I had to acclimate to subtle differences between fields, such as strategies for collaboration and publication. Anyone who is planning to change fields should make sure that the new workplace has a good working environment. It would have been impossible for me to get started without the help of colleagues willing to teach me the nuances of my new field. In return, I taught them skills I had developed in my original field.

Karin S. L. Johansson
Institute of Technology, University of Tartu, 50411 Tartu, Estonia. Email: ksl.johansson@outlook.com

After finishing my Ph.D., I made a career transition from aquaculture to entomology. I had to abandon a decade of fish research experience and start studying insects from scratch. Within the entomology community in China, however, a new researcher has little chance to get a research grant, and my career stalled. When our institute decided to launch a new aquaculture research program, I returned to the aquaculture field without hesitation and reconnected and collaborated with colleagues and old friends in the aquaculture community. To those transitioning from one field to another, I recommend being prepared for setbacks and maintaining connections to your original field.

Fengbo Li
Zhejiang Academy of Agricultural Sciences, Hangzhou, Zhejiang 310021, China.
Email: fengboli@gmail.com

When I transitioned from a master's degree program in Colombia to a Ph.D. program in the United Kingdom, I encountered cultural shock and language barriers. To address these challenges, I acknowledged the differences, maintained an open dialogue with peers and supervisors to ensure accurate communication, and tried to view setbacks with perspective. I also surrounded myself with other multilingual people who could relate to the process of learning another language and its frustrations, and I looked for academic role models who weren't native English speakers. Rather than compare myself to others who had different backgrounds or experience, I took pride in my ability to overcome obstacles.

Maria Fernanda Torres Jimenez
Department of Biological and Environmental Sciences, University of Gothenburg, and Gothenburg Global Biodiversity Centre, Gothenburg, Sweden.
Email: maria.torres@bioenv.gu.se

Be patient

As a premedical student, I focused on obtaining high exam scores and a competitive GPA. When transitioning to medical school, I realized that I was now studying for my future patients, not an exam score. The expectation that I would retain all the information I learned for a lifetime of clinical practice was daunting. I believe it is important for students making this transition to prepare themselves not just for the sprint to the next exam but for the "marathon" that is medical training.

Cody Lo
University of British Columbia, Vancouver, BC V6T 1Z3, Canada. Email: codylo@alumni.ubc.ca

Transitioning from a dental school in India to a business school in the United States was tough. Looking back, I was not adequately prepared to appreciate the teaching pedagogy in a business school, which was in stark contrast to what I had been exposed to in dental school. Difficulties adapting to the surroundings, environment, and culture of a different country further compounded my problems during my first year of graduate studies. To those in a similar predicament, I would recommend being a good listener, setting achievable goals for each day, focusing on seemingly small activities, being detail oriented, and asking for help if you need any. You will be surprised at how much people are willing to help those coming from a different country.

Veerasathpurush Allareddy
College of Dentistry, University of Illinois at Chicago, Chicago, IL 60612, USA. Email: sath@uic.edu

When I transitioned from undergraduate to postgraduate (Ph.D.) studies, there was a substantial change in my effort-reward system. In high school and college, the harder I studied, the better results I would get. In contrast, as a Ph.D. candidate, substantial effort does not always mean a visible reward right away. I realized that outcomes do not depend on me alone but also on supervisors, available resources, and even luck. My advice is to try to find small, perhaps unexpected, rewards from that effort and to be patient because a long-term important reward is likely still to come.

Carmen Romero-Molina
Department of Biochemistry and Molecular Biology, University of Seville, 41012 Seville, Spain.
Email: carmin533@hotmail.com

Trust yourself

I transitioned from a position as a specialist in the pharmaceutical industry to a position as a Ph.D. student. I went from solving day-to-day tasks with short deadlines to doing research projects lasting for several years. Sometimes, when weeks and even months go by without results, I have to remind myself that I made the transition to challenge myself and develop my skills. Remembering the reason that I'm here helps me overcome frustration and setbacks.

Signe Mosegaard
Research Unit for Molecular Medicine, Aarhus University, DK-8200 Aarhus N, Denmark.
Email: signe.mosegaard@clin.au.dk

I left a research assistant position in a 20-member lab to do my Ph.D. as the sole member of a new lab. Our productivity depended on my effort, and I understood both the responsibility and the opportunity the position entailed. I advise others to start

ILLUSTRATION: ROBERT NEUBECKER

a Ph.D. with a secure and honest vision of what they want to achieve. Otherwise, all the distractions in the world will not be comfort enough during the difficult or unexpected moments.

Steven M. Heaton
Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, VIC 3800, Australia. Email: steven.heaton@monash.edu

Medical school was like drinking from a fire hose. There was so much to learn; the more efficient I was, the better. When I transitioned to graduate school during my M.D./Ph.D. training, the rules for success were less clear. My research mentor often tells me to be creative, but there is no textbook for creativity. Testing the boundaries of science requires experiments or techniques that you have never done before, and when you try new things, you often fail. The key, I believe, is to work on questions that truly interest you. Genuine scientific curiosity can form the foundation for sustained perseverance. Asking "Why?" can turn failed experiments into new opportunities. It turns out that graduate school is also like drinking from a fire hose, but the fun part is you get to choose what you drink.

Jonathan Joon-Young Park
Department of Genetics, Yale University School of Medicine, New Haven, CT 06520, USA. Email: jonathan.park@yale.edu

For academics, transitioning to parenthood can be daunting. During the first years of my child's life, I experienced academia with a new perspective. We have traveled to conferences and field work all over the world. With the help of patient colleagues, supportive family and friends, conference day care, and a strong will, I have tried my best to be an active member of my scientific community. The more we participate, the more we serve as role models, push societal acceptance of equality, and improve conditions for future academic mothers.

Christine D. Bacon
Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden. Email: christinedbacon@gmail.com

When I started my Ph.D. project, I depended on guidance from my supervisor. When I transitioned to postdoctoral work, I had to independently navigate my research schedule, including both long-term and short-term goals. It was up to me to stay engaged, focus on my goals, and change direction when appropriate. Self-navigation driven by intrinsic motivation helped me find success as a postdoctoral researcher.

Sha Yu
School of Biological Sciences, Seoul National University, Seoul 8826, South Korea. Email: shayu@snu.ac.kr

When I first transitioned from Ph.D. to postdoc, the most challenging difference was the change in expectations. As a Ph.D. student, I benefited from my supervisor's helping hand and the understanding that my colleagues would tolerate mistakes. As a postdoc, I had to face the expectation that I could produce high-impact results with minimal supervision.

Emre Ozan Polat
ICFO–The Institute of Photonic Sciences, Castelldefels, 8860 Barcelona, Spain. Email: emre-ozan.polat@icfo.es

Communicate effectively

When I left my position as a graduate research assistant in academia to begin work as a scientist in industry, I learned that industry requires complex communication skills. I have to engage and communicate with internal and external stakeholders with substantially different levels of expertise, expectations, and backgrounds on a daily basis. Excellent oral and written communication skills are essential to ensure that messages are delivered clearly, precisely, and efficiently. Understanding the organizational landscape also plays a crucial role in effective communication, and the academic platform does not provide that type of complex environment. Thus, I believe internships and industrial co-op positions are the best opportunities for postgraduate students who would like to get true exposure to the industrial atmosphere and to improve their soft skills.

Dhanuka Wasalathanthri
Sanofi US, Fiskdale, MA 01518, USA. Email: dhanuka02@gmail.com

When I was a graduate student, it was natural and helpful to vent my negative feelings about incomplete tasks to someone in our group. After transitioning to a position as an adviser, I realized that I had to express my frustration more constructively. Expressing pessimism to the graduate students can affect them, sometimes more than we expect. Although I still share my concerns, I now try to lay out the problems and possible solutions without emphasizing my feelings, and I encourage the students to work together to solve problems.

Wei Wang
Fujian Institute of Research on the Structure of Matter, Chinese Academy of Sciences, Fuzhou, Fujian 350002, China. Email: wangwei@fjirsm.ac.cn

During my Ph.D., I started a medical device company. The transition from scientist to entrepreneur was challenging because the business world was completely new to me. Young scientists in a similar position, conflicted between a comfortable academic trajectory and the unknown startup world, should not hesitate to reach out to those more experienced. You will be surprised at the insights a quick phone call, email, or coffee chat can generate. In graduate school, while you're learning about the scientific process of hypothesis design and testing, explore what it takes to translate new ideas into market-ready products. It is never too early to learn how to become a good salesperson. The ability to pitch an idea effectively will help you communicate your science more successfully in any setting.

Divyansh Agarwal
Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. Email: divyansh@upenn.edu

10.1126/science.aaw2641

ILLUSTRATION: ROBERT NEUBECKER

PERSPECTIVES

CROWD DYNAMICS

Flowing crowds

Modeling human crowds as a fluid allows prediction of group behavior

By **Nicholas T. Ouellette**

Collective behavior of social animals, particularly coordinated group movements, is one of the most striking phenomena in the natural world, as anyone who has been enthralled by flocks of starlings or schools of sardines can attest. Research in this broad, interdisciplinary field has recently exploded, with motivations ranging from understanding the biological basis of social behavior (*1*) to developing algorithms for designing and controlling swarms of robots (*2*). There is good reason to think that the behavior of human crowds is quite similar to these animal groups and that studying humans might help elucidate the origins of crowd panic and other dangerous instabilities that can lead to injury or loss of life (*3*). All these goals require modeling, both as a check on our understanding and as a predictive tool for analyzing new situations. On p. 46 of this issue, Bain and Bartolo (*4*) describe a powerful new way to model human crowds. Instead of focusing on individuals, they build a continuum "hydrodynamic" model of the crowd as a whole and then constrain it with observational data collected from marathon runners. This approach circumvents many of the sometimes-questionable assumptions that have previously been made and provides a roadmap for constructing an empirically grounded theory of collective behavior.

The dominant paradigm for describing collective behavior is agent-based modeling. Each individual in a group is treated as an "agent" that follows a set of rules to determine its behavior. Most commonly, these rules include instructions to not stray too far from the group, to avoid collisions, and, for directed motion, to move in the same direction as nearby agents (*5*). Agent-based models have succeeded in qualitatively reproducing patterns observed in real animal groups (*1*), providing strong evidence that local interaction alone is sufficient to drive the formation of group-level coherent behavior. However, pattern isn't everything, and just because a model's output qualitatively looks acceptable does not mean that the model is right. There is also reason to be skeptical of this approach because it requires a priori assumptions about animal behavior that are at least oversimplified if not incorrect. In recent years, researchers have attempted to constrain these assumed rules by measuring real animal groups (*6*–*9*). However, this approach is fraught. Data can be hard to come by for many reasons—including, in the case of humans, substantial ethical considerations. More fundamentally, extracting individual interaction rules from observations of group behavior is a complex, nonlinear inverse problem, so drawing reliable conclusions in a model-free way is often impossible.

Rather than thinking about a group as a composite of individual agents with their own rules, a group can instead be considered as an entity itself. The properties of the group certainly emerge from interactions between the individuals, but to model these properties, it is not necessary to know where they come from. In this sense, collective behavior can be treated analogously to how the mechanics of materials are modeled. To describe how water flows, one does not need to consider molecular interactions; rather, one can apply conservation laws for a macroscopic amount of water and constrain them with empirical observations (in the case of hydrodynamics, a linear constitutive law that relates stress and strain rate). Such an approach cannot capture the behavior of water molecules, but if the goal is to formulate a predictive theory of hydrodynamics, they are not necessary.

Bain and Bartolo have essentially followed this approach for human crowds. They begin with generic equations of motion for the

A human crowd supports a wave-like transmission of information when the crowd is modeled as a single entity rather than a composite of individuals.

Department of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305, USA. Email: nto@stanford.edu

continuum crowd density and velocity starting from mass and momentum conservation. To gather the necessary information to constrain these equations, they made a series of observations of a large number of people in a confined space—in this case, runners at the start of major marathons who are strongly constrained by geometry (along a street) and by the control of race officials. Runners are typically allowed to start the race in small groups according to their expected speed. The authors quantitatively analyzed videos of runners at the start of races and found that each starting event (when officials let runners start to advance toward the starting line) triggered an upstream-propagating wave of density and velocity perturbations through the crowd. In other words, the crowd supported a wave-like transmission of information (in this case, the start of the race). The properties of this wave provided the information needed to constrain the equations of motion, which in turn enabled predictions about the dynamics of other crowds without resorting to any assumptions about human behavior. That's a very important outcome, including for those interested in modeling potentially dangerous situations such as crowd panic.

The approach of Bain and Bartolo opens many avenues for future work for collective behavior researchers more generally. For example, it should inspire studies that pinpoint a group response to perturbations—such as the traveling waves launched by the starting events of a marathon race—to constrain continuum models. Some studies along these lines have already been done, such as characterizing the response of starling flocks to predators (*10*), of ants to mechanical stresses (*11*), and of midge swarms to sensory cues (*12*). More work is necessary to incorporate these findings into dynamical continuum models that avoid the need for a priori assumptions about animal behavior. Ultimately, such models may even be an effective way to determine the local interactions themselves because any agent-based model must approach the continuum model as a limiting case. ■

REFERENCES

1. J. K. Parrish, L. Edelstein-Keshet, *Science* **284**, 99 (1999).
2. M. Rubenstein, A. Cornejo, R. Nagpal, *Science* **345**, 795 (2014).
3. D. Helbing, I. Farkas, T. Vicsek, *Nature* **407**, 487 (2000).
4. N. Bain, D. Bartolo, *Science* **363**, 46 (2019).
5. C. W. Reynolds, *Comput. Graph.* **21**, 25 (1987).
6. Y. Katz, K. Tunstrøm, C. C. Ioannou, C. Huepe, I. D. Couzin, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 18720 (2011).
7. J. E. Herbert-Read *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 18726 (2011).
8. W. Bialek *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 4786 (2012).
9. J. G. Puckett, D. H. Kelley, N. T. Ouellette, *Sci. Rep.* **4**, 4766 (2014).
10. A. Procaccini *et al.*, *Anim. Behav.* **82**, 759 (2011).
11. M. Tennenbaum, Z. Liu, D. Hu, A. Fernandez-Nieves, *Nat. Mater.* **15**, 54 (2016).
12. R. Ni, J. G. Puckett, E. R. Dufresne, N. T. Ouellette, *Phys. Rev. Lett.* **115**, 118104 (2015).

10.1126/science.aav9869

PHOTO: RHETT A. BUTLER/MONGABAY.COM

CONSERVATION

The sound of a tropical forest

Recording of forest soundscapes can help monitor animal biodiversity for conservation

By **Zuzana Burivalova[1], Edward T. Game[2,3], Rhett A. Butler[4]**

Conservation areas around the world aim to help conserve animal biodiversity, but it is often difficult to measure conservation success without detailed on-the-ground surveys. High-resolution satellite imagery can be used to verify whether or not deforestation has occurred in areas dedicated for conservation (*1*). Such remote sensing analyses can reveal forest loss and, in some cases, severe forest degradation, such as through fragmentation and intensive selective logging, especially if it includes the construction of roads or camps. However, conservation benefit is determined not only by forest loss but also by the level of degradation in those forests left standing. Bioacoustics—specifically the recording and analysis of entire soundscapes—is an emerging tool with great promise for effectively monitoring animal biodiversity in tropical forests under various conservation schemes (*2*, *3*).

Even forests that appear intact in satellite imagery can have low biodiversity conservation value because of effects such as canopy simplification, understory fires, invasion by exotic species, or overhunting. These forms of degradation are difficult to monitor remotely with satellite imagery, resulting in a common but faulty assumption that conserving forest cover is necessarily equivalent to conserving biodiversity. Continuing advances in spectral imagery and lidar (light detection and ranging) reveal progressively finer levels of forest change, but they still remain a proxy for animal biodiversity rather than a direct measure of it (*4*).

Repeated on-the-ground surveys can provide the required information to assess animal biodiversity. However, such surveys are expensive, cover limited ground, and may be affected by the biases of individual experts. One possible alternative is the use of bioacoustics, which can detect animals by their vocalizations. Depending on vegetation structure and the vocalizing species, acoustic recorders can detect animal calls and song from several hundred meters away (*5*). Autonomous sound-recording devices are now available from several companies as small units that are inconspicuous to humans. They can be programmed to record either continuously, if there is sufficient solar power or cellular network signal for direct transmission of data to cloud storage, or at given intervals, if battery power and data storage are limiting factors (*6*). Several multiyear recordings have now been completed (*7*).

Selected times of the day can convey a disproportionately large amount of information about the resident biodiversity; for example, mornings and evenings have been found to be particularly important for detecting differences between forests that are used in different ways by humans (*8*). With further developments in energy and data storage and transmission, continuous recording is likely to become the norm.

Relative to on-the-ground surveys, bioacoustics is inexpensive, making it more

[1]*Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ 08540, USA.* [2]*The Nature Conservancy, South Brisbane, QLD 4101, Australia.* [3]*School of Biological Sciences, University of Queensland, St. Lucia, QLD 4072, Australia.* [4]*Mongabay.com, Menlo Park, CA 94026, USA. Email: zuzanab@princeton.edu*

Hornbills, such as this rhinoceros hornbill in Bukit Tigapuluh National Park, Sumatra, Indonesia, have prominent vocalizations that can be identified in soundscapes.

feasible to repeat measurements over time. Also, the results are not influenced by individual researchers' biases or simply by the presence of observers in the field. The method offers the possibility to monitor multiple taxonomic groups at the same time (all vocalizing birds, mammals, insects, and amphibians), as opposed to, for example, camera traps. Finally, the data can be reanalyzed in the future with improved algorithms or to search for specific acoustic features. Analysis of human-made sounds can help to clarify how sounds from machinery (such as tractors, bulldozers, and chainsaws) affect habitat quality and to track illegal human activities, such as gunshots from poachers or chainsaws in illegal logging (*9*).

Acoustic data from soundscapes can be analyzed in many ways (*10*). Various indices can be calculated that characterize the soundscape for each time and frequency unit (*11*, *12*). Alternatively, individual species can be identified by experts, algorithms (*13*), or deep learning (*14*).

Soundscape analysis using indices appears most suitable to monitor the general state and recovery of forests, because it does not require site-specific species lists (*8*). Random forest models based on multiple acoustic indices can predict species richness with very high accuracy (*11*). However, further studies linking on-the-ground biodiversity surveys to soundscape indices are needed from a wide variety of forest types and human disturbances to determine whether such indices can be generalized. In areas where hunting is important, the recordings could also be used to determine the presence or absence of the hunted species (typically large mammals and birds) using individual species recognition algorithms.

Bioacoustics has particular potential in the context of industry sustainability certification and zero-deforestation commitments, both of which have become popular, widely publicized conservation strategies (*1*, *15*). Companies involved in such industries as palm oil, beef, soy, and pulp and paper production commit to not cause any deforestation through their industrial development. Typically, this means that any new plantation, ranch, or farm can only be developed in an area that is already deforested or heavily degraded. In some countries, such as Brazil, companies are legally obliged to protect parts of their concessions from deforestation. However, precise definitions of zero deforestation are often missing (*15*). The conservation benefit of such industry-protected forests should be determined not just by how much forest loss has been avoided, but also by the level of biological integrity of those forests left standing. Bioacoustics has the potential to provide this information (see the figure).

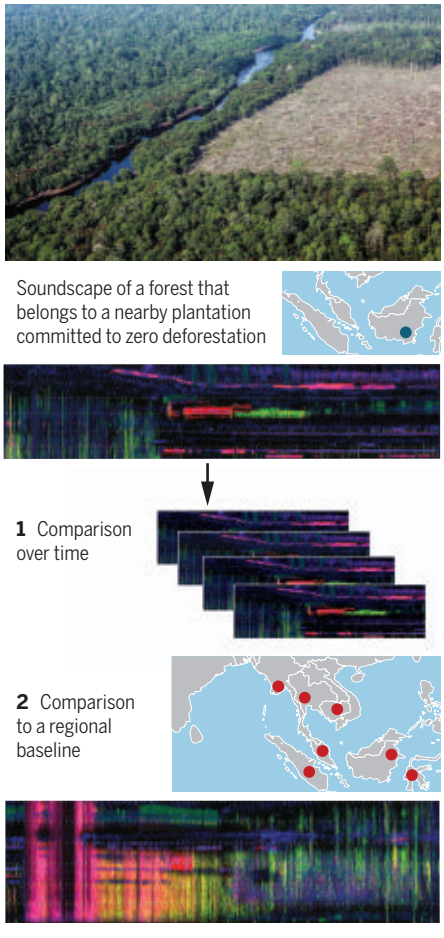
Advances in bioacoustics, as well as the robustness and affordability of sound-recording devices, make it possible for companies or independent consultants to deploy sound recorders in areas of forest maintained by a company under legal requirements, certification, or a zero-deforestation commitment. If the soundscape of a forest spared from conversion were becoming more impoverished and altered beyond the natural variation of the soundscape baseline, on-the-ground survey would be warranted. Slow, gradual changes in soundscape composition due to climate change might be beyond the direct control of the companies, but abrupt and quick change in soundscapes is more likely to be attributable to management. In these cases, other measures (such as prevention of hunting, reforesting edges or the degraded areas of the conserved zone with native species, or curbing fires) would be called for by auditors, who are typically involved in independent verification of a company's commitments.

Because of the enormous size of the acoustic datasets and the computational power required to analyze them, there is a need for a global organization to host a global acoustic platform, which would allow direct, on-the-fly analysis. The development of such a data hosting and analysis platform should be a priority, together with the collection of regional soundscape baselines by scientists. Such baselines would be especially useful for understanding and accounting for the natural seasonal and interannual variation of soundscapes, as well as for comparison of the industry-protected forest soundscapes with the closest available undisturbed sites.

Nongovernmental organizations and the conservation community need to be able to truly evaluate the effectiveness of conservation interventions. Many (but not all) companies want to be able to provide objective, consistent, and easy-to-share evidence documenting their conservation efforts, at a low cost. Environmentally aware consumers may feel more confident about purchasing from brands that can show the results of their conservation efforts, on top of their certification logo or zero-deforestation commitment. The scientific community will benefit from a huge tranche of data on ecological communities across the tropics. It is therefore in the interest of certification bodies to harness the developments in bioacoustics for better enforcement and effectiveness measurements of their schemes. ■

How soundscape monitoring can aid conservation

This diagram shows how bioacoustics monitoring could be implemented in a concession governed by a corporate conservation commitment or sustainability certification. Soundscape recordings would be compared to each other over time, as well as to regional baselines from the closest available intact forest landscapes.



GRAPHIC: N. CARY/*SCIENCE*; (PHOTO) RHETT A. BUTLER/MONGABAY; (SOUNDSCAPES) QUEENSLAND UNIVERSITY OF TECHNOLOGY

REFERENCES

1. H. K. Gibbs *et al.*, *Conserv. Lett.* **9**, 32 (2016).
2. J. Sueur, A. Farina, *Biosemiotics* **8**, 493 (2015).
3. B. Krause, A. Farina, *Biol. Conserv.* **195**, 245 (2016).
4. M. M. C. Bustamante *et al.*, *Glob. Change Biol.* **22**, 92 (2016).
5. K. Darras, P. Pütz, Fahrurrozi, K. Rembold, T. Tscharntke, *Biol. Conserv.* **201**, 29 (2016).
6. A. Rodriguez *et al.*, *Ecol. Inform.* **21**, 133 (2014).
7. S. H. Gage, A. C. Axel, *Ecol. Inform.* **21**, 100 (2014).
8. Z. Burivalova *et al.*, *Conserv. Biol.* **32**, 205 (2018).
9. C. Astaras, J. M. Linder, P. Wrege, R. D. Orume, D. W. Macdonald, *Front. Ecol. Environ.* **15**, 233 (2017).
10. J. L. Deichmann *et al.*, *Biotropica* **50**, 713 (2018).
11. R. T. Buxton *et al.*, *Conserv. Biol.* **32**, 1174 (2018).
12. L. M. Ferreira *et al.*, *J. Ecoacoust.* **2**, PVH6YZ (2018).
13. A. P. Hill *et al.*, *Methods Ecol. Evol.* **9**, 1199 (2018).
14. D. Stowell, Y. Stylianou, M. Wood, H. Pamuła, H. Glotin, *Methods Ecol. Evol.* 10.1111/2041-210X.13103 (2018).
15. S. Brown, D. Zarin, *Science* **342**, 805 (2013).

10.1126/science.aav1902

INFECTIOUS DISEASE

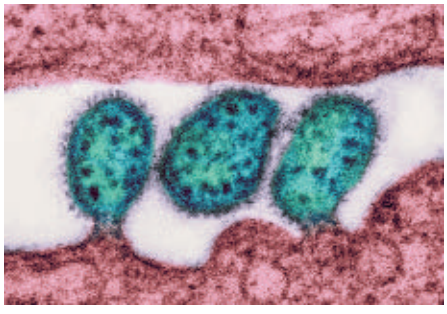
Understanding Lassa fever

Genomics study informs about Lassa fever epidemiology

By **Nahid Bhadelia**

Lassa fever is a viral hemorrhagic fever prevalent in West Africa that has been gaining international attention as an emerging infectious disease with the potential to cause epidemics (*1*). Confirmed and suspected cases of Lassa fever have been steadily rising in Nigeria over the past 3 years. Laboratory-confirmed cases have increased from 106 in 2016 to 143 in 2017 and had already reached 562 by November 2018 (*2*). Part of defining the scope of the problem is trying to assess whether this is a true increase in the number of people afflicted by the infection, due to either changes in the virus itself or the geographical spread of the vector (rodents of the *Mastomys* spp.), or a reflection of higher rates of detection and diagnosis secondary to the increased attention and interest of clinicians and laboratorians (*1*). Lassa fever outbreaks illustrate the issues associated with the response and management of emerging infectious diseases: How do you plan the public health, clinical, and community responses to a disease while you are still learning about the epidemiology, pathophysiology, and the ecological factors contributing to the spread of the pathogen? On page 74 of this issue, Kafetzopoulou *et al.* (*3*) present the results of a rapid genomic study of Lassa virus (LASV) from the cases of 2018, which have improved understanding of how the disease has been spreading in Nigeria and have led to informed and targeted disease-control strategies. The study also further describes the use of a new and compact genomic sequencing device, which may start playing a larger role in defining other emerging infectious disease outbreaks in real time.

The device generates exponentially longer reads of genetic material than traditional sequencing and offers some remarkable advantages over existing next-generation sequencing platforms (*4*). To date, sequencing the genomes of organisms has been like printing individual pages of a novel in the wrong order (many of which end with the same words and phrases) and trying to put the story together with guess work. The device allows researchers to instead print out whole chapters, making it easier to get a sense of the bigger picture. In the biological world, longer reads of genomic information also provide invaluable information about structural variations and epigenetic modifications between individual organisms. However, with longer reads comes greater error rates, which appear to improve with repeated reads of the same genetic material, as long as it is of high-enough quality and quantity (*5*). With its compact size, portability, and quick turnaround time, the device can be rapidly deployed in outbreak areas where laboratory capacity may not exist for genomic sequencing and when samples cannot be exported out of the country for analysis (*4*). Its uses include not only sequencing to evaluate viral evolution and chains of transmission but also, as the authors highlight, identification of multiple cocirculating viruses in patient samples. The technology has already been used in outbreaks with other pathogens, including Ebola virus (*6*).



LASV (blue) is spread by rat urine and droppings and infects humans through ingestion or inhalation.

Kafetzopoulou *et al.* used this technology to compare the phylogenetic differences between the strains of viruses from 120 confirmed LASV samples from Nigeria from the spring of 2018. With the increase in the number of total cases as well as clusters of cases in recent years, one of the concerns has been whether the virus has changed, allowing Lassa fever to transmit between humans more easily. By examining the level of genetic diversity between viruses in each of the different samples, alongside epidemiological information about the cases, the authors demonstrated that most of the viral genomes were different enough from each other that they had to have come from humans infected by different rodents, rather than from those infected through transmission from other humans (which would have greater shared phylogenetic homology). Another recent study from the current outbreak of Lassa fever in Nigeria has drawn similar conclusions using more traditional genetic sequencing technology (*7*).

Despite the critical epidemiological clues provided by both studies, it is helpful to keep in mind that the narrative of Lassa fever is driven by where the testing for the disease occurs. Owing to the high number of asymptomatic cases and the nonspecific symptoms mimicking myriads of infectious diseases in the region in patients who actually become sick, establishing the true burden of Lassa fever in Nigeria and other West African countries has been near impossible (*8*). The disease is mainly spread through contact with the urine and feces of multimammate rats in the household setting, a human-vector interface that happens mostly in poor communities and in rural areas. Unfortunately, these are also communities with considerably less access to health care, and, because diagnostic testing for Lassa fever is only available at a handful of reference laboratories, it is thought that many suspected patients are either never tested or only tested after a delay (*9*). Conversely, ~80% of suspected cases tested in 2018 were actually negative for Lassa fever (*2*).

Because of a lack of access to laboratory testing, a sizeable portion of the world's population, particularly in resource-limited areas, is simply treated on the basis of symptomatology for common infectious diseases such as malaria or cholera rather than tested on presentation to confirm the underlying diagnosis (*10*). This paradigm allows emerging infectious diseases to circulate in populations without initial detection. Hence, the real test of emerging microbial detection techniques will be how accurate, affordable, and amenable to widespread use they are and whether they can test both for common endemic infectious diseases as well as rarer pathogens of high concern, such as viral hemorrhagic fevers. To get the true sense of the disease burden and deaths from these emerging infectious diseases, to really solve the problem of Lassa fever, we still need diagnostic technologies closer to the point of care, and everywhere patients get sick. ■

Section of Infectious Diseases, Boston University School of Medicine and National Emerging Infectious Diseases Laboratories, Boston University, Boston, MA, USA.
Email: nbhadeli@bu.edu

REFERENCES

1. L. Roberts, *Science* **359**, 1201 (2018).
2. Nigeria Centre for Disease Control, "Weekly epidemiological report: Epi week 1, week 46" (2018).
3. L. E. Kafetzopoulou *et al.*, *Science* **363**, 74 (2019).
4. H. Lu *et al.*, *Genom. Proteom. Bioinf.* **14**, 265 (2016).
5. A. D. Tyler *et al.*, *Sci. Rep.* **8**, 10931 (2018).
6. J. Quick *et al.*, *Nature* **530**, 228 (2016).
7. K. J. Siddle *et al.*, *N. Engl. J. Med.* **379**, 1745 (2018).
8. C. Houlihan, R. Behrens, *BMJ* **358**, j2986 (2017).
9. R. S. Dhillon *et al.*, *Lancet Infect. Dis.* **18**, 601 (2018).
10. A. M. Caliendo *et al.*, *Clin. Infect. Dis.* **57** (suppl. 3), S139 (2013).

10.1126/science.aav8958

PHOTO: AMI IMAGES/SCIENCE SOURCE

NEUROSCIENCE

Weakening synapses to cull memories

Calcium sensor synaptotagmin-3 helps weaken synaptic strength and supports forgetting

By **Nataniel J. Mandelberg** *and* **Richard Tsien**

From correct answers on a school exam to a loved one's birthday, we have all forgotten things we wish we had not. The ability to forget, however, is a feature rather than a flaw of how our brains work. As the celebrated author Jorge Luis Borges wrote about a man incapable of forgetting, *Funes the Memorious* (*1*), "I suspect, however, that he was not very capable of thought. To think is to forget differences, generalize, make abstractions." Although Funes's example is literary, it contains a grain of truth. Neuroscientists have traditionally paid more attention to how the brain remembers than how it forgets, but there is increasing clarity about mechanisms and roles of forgetting (*2*, *3*). By forgetting, we prioritize and separate the useful from the irrelevant and more easily reorganize information to learn (*4*). On page 44 of this issue, Awasthi *et al.* (*5*) show that the Ca^{2+}-sensing protein synaptotagmin-3 (SYT3) is essential for synaptic weakening and link this molecular process to beneficial forgetting in mice.

A compelling association exists between memory and changes in neuron connectivity. Neurons are linked by synapses, structures in which boutons from the axon of the upstream (presynaptic) neuron communicate with spines on dendrites of the downstream (postsynaptic) neuron via neurotransmitter release. If the presynaptic neuron reliably drives the activity of the postsynaptic neuron, the synapse's strength, or weight, increases through long-term potentiation (LTP). However, if the activities of the neurons are poorly correlated, the connection weakens through long-term depression (LTD). Controlling LTP and LTD in rodents drastically affects memories they have formed: Previously learned fear conditioning in mice can be deactivated through optogenetically induced LTD and reactivated with LTP (*6*). LTP occurs when the activity of the presynaptic neuron causes a large influx of Ca^{2+} into the postsynaptic neuron. It manifests as an increased number of α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) type 2 subunit–containing glutamate (GluA2) receptors at the spine, making the postsynaptic neuron more responsive to input from the presynaptic cell. By contrast, LTD is driven by smaller Ca^{2+} events and reflects the removal of GluA2 receptors from the synapse through endocytosis, weakening the connection.

To clarify the mechanism of GluA2 receptor endocytosis, Awasthi *et al.* looked closely at members of the synaptotagmin family of Ca^{2+}-sensitive proteins. Synaptotagmins control the exocytosis of presynaptic vesicles (full of neurotransmitters) from the bouton, and postsynaptic SYT1 and SYT7 are required for glutamate receptor exocytosis in LTP (*7*). Awasthi *et al.* show that SYT3 has key qualifications to be the arbiter of LTD through its regulation of GluA2 receptors: SYT3 is abundant at postsynaptic regions, is endocytosed when neurons are stimulated, binds directly to GluA2 receptors, and controls their internalization. Awasthi *et al.* show that SYT3 has a functional impact on synaptic plasticity. Induction of LTP was unaffected by *Syt3* gene deletion in mice, whereas the decay of LTP and the induction of LTD, both reliant on GluA2 receptor endocytosis, were abolished.

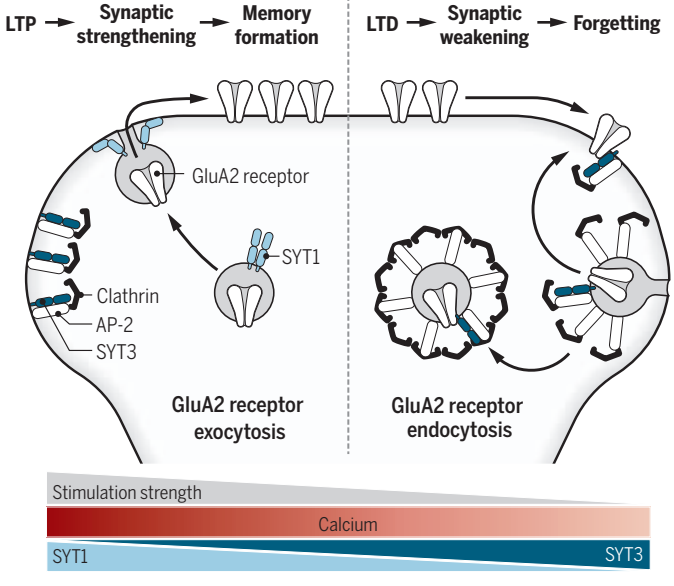
Notably, SYT3 binds Ca^{2+} at 5- to 20-fold lower concentrations (*8*) than does SYT1, which participates in postsynaptic recruitment of GluA2 receptors in LTP (*7*). Taken together, these findings (*5*, *7*, *8*) align with a model that attributes the different Ca^{2+} requirements of LTP and LTD to their structurally divergent Ca^{2+} sensors. Other players, such as calmodulin kinases and calcineurin, may also participate in the all-important molecular decision between synaptic strengthening and weakening. Work is needed to further integrate postsynaptic vesicle cycling and signaling biochemistry.

Awasthi *et al.* show that these circuit-level findings have in vivo relevance by training mice to find target locations. When *Syt3* was deleted, the mice learned the target location as well as the wild-type mice did but, when the target was moved, showed impaired forgetting of the initial location, reflected by an unwillingness to leave the target's original location. These results extend reports that GluA2 receptor endocytosis mediates memory loss (*9*) and that blocking this endocytosis preserves memories (*10*), by showing that this process is mediated by SYT3 in vivo.

Awasthi *et al.* go on to demonstrate that this inability to forget hinders behavioral flexibility, much as in Borges's story of Funes. The mice were again tasked with finding a hidden platform in a pool of water, but the location of the platform was changed every day. Mice in which *Syt3* was deleted persisted with former platform locations rather than seeking new ones, as if unable to distinguish between a past memory and a new, immediately relevant experience.

Switching between remembering and forgetting

In this model, different SYT isoforms with different Ca^{2+} affinities (*8*) determine whether GluA2 receptors should be added to synapses, which strengthens them (LTP), or removed, which weakens them (LTD). This contributes at the synaptic level to the decision in the brain of whether a memory should be encoded or forgotten.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

Langone Medical Center, New York University, New York, NY, USA. Email: richard.tsien@nyumc.org; nataniel.mandelberg@nyumc.org

The study of Awasthi *et al.* highlights the importance of vesicle cycling on both sides of the synapse. Clarification of the cell biology of how postsynaptic weights are weakened shifts the spotlight to questions about the spatiotemporal allocation and reallocation of such weights. We speculate that the internalized glutamate receptor vesicles are a synaptic resource too precious to waste and can be redistributed among nearby dendritic spines to strengthen nearby postsynapses (*11*), much as vesicles of neurotransmitters can be reallocated among presynaptic boutons along an axon (*12*). This study begins to show how neurons might use similar tools pre- and postsynaptically to channel resources to the most important synapses while culling synapses that no longer encode relevant information.

The work of Awasthi *et al.* has a close yet unexplored relationship to pathological processes in neuropsychiatric and neurodegenerative disorders. The exaggerated removal of glutamate receptors, including GluA2 receptors, is a feature of Alzheimer's disease (AD) (*13*) and potentially linked to the associated forgetting. This process involves protein interacting with C kinase-1 (PICK1), another mediator of GluA2 receptor endocytosis (*14*), but the role of SYT3 and the relationships between PICK1, SYT3, and other proteins involved in GluA2 receptor endocytosis remain unclear. This pathophysiological endocytosis could contribute to the memory loss experienced by AD patients, and we speculate that pharmacological interventions that restore normal GluA2 receptor endocytosis could help mitigate these defects. Furthermore, behavioral inflexibility is a hallmark of autism spectrum disorders (ASD) and might be assigned to deficits in forgetting, as supported by five fruitfly models of ASD risk genes (*2*). In another study, patients with ASD were asked to choose the location of a stimulus. Although they performed equally well as the control patients, the ASD patients showed extra reversion back to the original location even after the stimulus location changed (*15*). Elucidation of mechanisms of this inflexibility will benefit from the insights that Awasthi *et al.* have elegantly provided. ■

REFERENCES

1. J. L. Borges, *Labyrinths* (New Directions Publishing, 1964).
2. R. L. Davis, Y. Zhong, *Neuron* **95**, 490 (2017).
3. O. Hardt *et al.*, *Trends Cogn. Sci. (Regul. Ed.)* **17**, 111 (2013).
4. B. A. Richards, P. W. Frankland, *Neuron* **94**, 1071 (2017).
5. A. Awasthi *et al.*, *Science* **363**, eaav1483 (2019).
6. S. Nabavi *et al.*, *Nature* **511**, 348 (2014).
7. D. Wu *et al.*, *Nature* **544**, 316 (2017).
8. S. Sugita *et al.*, *EMBO J.* **21**, 270 (2002).
9. Z. Dong *et al.*, *J. Clin. Invest.* **125**, 234 (2015).
10. P. V. Migues *et al.*, *J. Neurosci.* **36**, 3481 (2016).
11. C. Mullins *et al.*, *Neuron* **89**, 1131 (2016).
12. K. Staras *et al.*, *Neuron* **66**, 37 (2010).
13. H. Hsieh *et al.*, *Neuron* **52**, 831 (2006).
14. S. Alfonso *et al.*, *Eur. J. Neurosci.* **39**, 1225 (2014).
15. A.-M. D'Cruz *et al.*, *Neuropsychology* **27**, 152 (2013).

10.1126/science.aaw1675

METABOLISM

Improving crop yield

Synthetic photorespiration bypass increases crop yield

By **Marion Eisenhut** *and* **Andreas P. M. Weber**

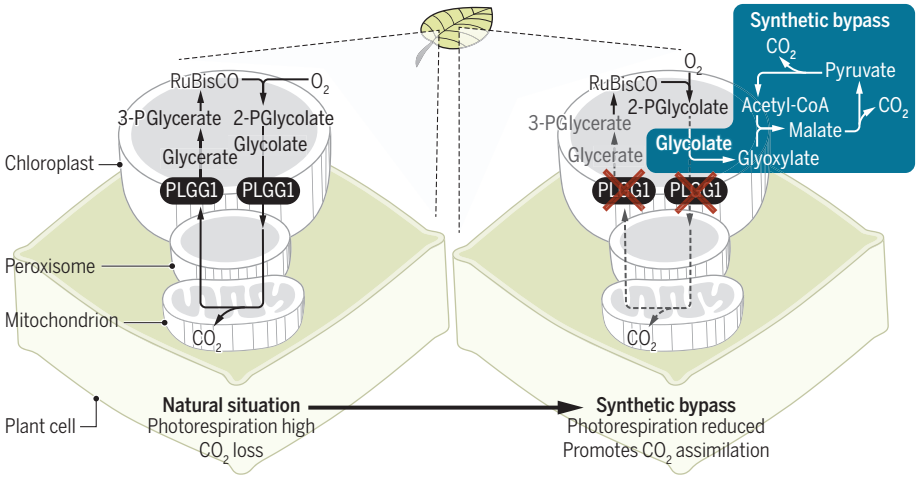
The enzyme ribulose 1,5-bisphosphate carboxylase-oxygenase (RuBisCO) is one of the most abundant proteins on Earth. During photosynthesis, it assimilates atmospheric CO_2 into biomass and hence is a major driver of the global carbon cycle. However, the enzyme is catalytically imperfect. It accepts not only CO_2 as a substrate, but also O_2, which leads to the formation of a toxic byproduct, 2-phosphoglycolate (2-PGlycolate) (*1*). The metabolic pathway photorespiration detoxifies 2-PGlycolate, and it is essential for performing photosynthesis in an O_2-containing atmosphere. Importantly, photorespiration causes a 20 to 50% yield penalty, depending on the environmental conditions and the type of photosynthesis employed (*2*). Multiple attempts have been undertaken to overcome this yield penalty and thereby increase biomass production in plants, with limited success to date. On page 45 of this issue, South *et al.* (*3*) present a synthetic pathway that fully detoxifies 2-PGlycolate inside plant chloroplasts. Transgenic tobacco plants expressing this pathway show strongly enhanced biomass production in field trials, suggesting that this could be used to improve crop yields.

Institute of Plant Biochemistry, Cluster of Excellence on Plant Science (CEPLAS), Heinrich Heine University Düsseldorf, Universitätsstrasse 1, 40225 Düsseldorf, Germany. Email: aweber@hhu.de

Photorespiration is an essential metabolic repair pathway in all organisms that perform oxygenic photosynthesis, from cyanobacteria, through algae, to land plants (*2*, *4*). Core photorespiratory metabolism comprises nine enzymatic steps that are distributed over chloroplast, peroxisome, and mitochondrion within a plant cell. It converts detrimental 2-PGlycolate into the Calvin-Benson cycle intermediate 3-PGlycerate and thereby returns 75% of otherwise unusable carbon to photosynthetic metabolism. However, during this salvage pathway, 25% of previously fixed CO_2 gets lost, and energy is consumed (see the figure). Hence, albeit essential, photorespiration is also considered a wasteful and inefficient process (*2*). Accordingly, photorespiration has been identified as a prime target for engineering to improve crop yields, and diverse strategies have been developed to improve photosynthetic efficiency by reducing photorespiration and/or enhancing the CO_2 fixation processes. Some of these attempts are inspired by naturally occurring CO_2-concentrating mechanisms present in, for example, cyanobacteria and algae. Others are based on implementing synthetic metabolic routes to redirect the canonical pathway of CO_2 assimilation and photorespiration (*5*).

Engineering wasteful photorespiration into a beneficial process

The fixation of O_2 by RuBisCO in chloroplasts leads to high rates of photorespiration and a concomitant loss of CO_2 from mitochondria. A synthetic bypass and the restricted activity of PLGG1 allow metabolism of glycolate with release of CO_2 inside of the chloroplasts, which promotes CO_2 fixation by RuBisCO and improves yield.



GRAPHIC: A. KITTERMAN/*SCIENCE*

South *et al.* revisited two previously established synthetic bypasses of photorespiration (*6*, *7*) and tested a newly designed pathway in genetically modified tobacco plants. These pathways aim to completely metabolize the photorespiratory metabolite glycolate, which is generated from 2-PGlycolate by phosphoglycolate phosphatase within the chloroplast. They release CO_2 close to RuBisCO (not in mitochondria, as in natural photorespiration) to increase the ratio of CO_2 to O_2 fixation. Alternative pathway (AP) 1 originates from the bacterium *Escherichia coli* and uses five enzymes that oxidize glycolate via glyoxylate and tartronic semialdehyde to glycerate (*6*). The second bypass, AP2, uses three enzymes that convert glycolate via glyoxylate and malate to acetyl–coenzyme A (CoA). AP2 also requires the expression of catalase for detoxification of hydrogen peroxide that results from conversion of glycolate to glyoxylate by glycolate oxidase (*7*). AP1 and AP2 were previously shown to increase biomass (*6*, *7*). AP3 was newly designed by South *et al.* In AP3, only two transgenes had to be introduced into the plant chloroplast: a glycolate dehydrogenase that converts glycolate into glyoxylate derived from the green alga *Chlamydomonas reinhardtii* was redirected to tobacco chloroplasts, and similar to AP2, a malate synthase was expressed to convert glyoxylate to malate and eventually to acetyl-CoA via the native chloroplast-resident nicotinamide adenine dinucleotide phosphate (NADP)–malic enzyme (see the figure). Using the green algal glycolate dehydrogenase instead of plant glycolate oxidase prevents production of hydrogen peroxide, and hence additional expression of catalase is unnecessary.

Two important differences from the original pathway designs (*6*, *7*) represent major advances. Besides introducing a synthetic bypass, South *et al.* also reduced the expression of PLASTIDIAL GLYCOLATE/GLYCERATE TRANSPORTER 1 (PLGG1) (*8*). This modification was suggested previously (*9*) to increase the potential of synthetic bypasses, because it restricts the export of glycolate from chloroplasts and hence promotes its consumption by the synthetic bypass. A larger portion of glycolate is decarboxylated within the chloroplast by the synthetically engineered bypass, leading to enhanced CO_2 fixation activity of RuBisCO. This comes with an impressive yield gain of more than 40%. Importantly, yield improvements positively correlated with the expression levels of the introduced enzymes, which highlights the importance of high and balanced expression of the transgenes. Typical annual yield gains in crop breeding are below 2%; hence, the synthetic pathway holds potential for a step change in yield improvement by genetic modification of crops. In contrast to earlier work (*6*, *7*), the pathways were introduced into the model crop tobacco, which was investigated not only in growth chambers and greenhouses, but also in field trials. Thus, the yield gains manifested in an agriculturally relevant scenario and not only in controlled environments.

Importantly, the synthetic pathways open new avenues for reevaluating long-standing hypotheses regarding the importance of photorespiration beyond detoxification of 2-PGlycolate. Photorespiration is considered indispensable for photosynthesis in an O_2-containing atmosphere, and mutants defective in photorespiration can only survive in a high-CO_2 atmosphere (*10*). Genetic suppressor screens on such mutants have been unsuccessful to date. The study of South *et al.* demonstrates that a photorespiratory phenotype (repression of *PLGG1*) can be suppressed by metabolic engineering. The true reason or reasons for the indispensability of photorespiratory metabolism are intensely debated and include the detoxification of 2-PGlycolate; carbon salvage; biosynthesis of the amino acids glycine and serine (*11*); generation of activated C1-units; and protection from photoinhibition and dissipation of excess excitation energy (*2*, *12*, *13*). The work of South *et al.* indicates that plant metabolism adapts to the synthetic pathways and compensates for reduced flux through the peroxisomal and mitochondrial parts of native photorespiration. This implies that 2-PGlycolate detoxification and carbon recycling are the critical functions of photorespiration.

Recently, the optimization of a mechanism that protects plants from excess light, nonphotochemical quenching (NPQ), which is dissipation of excess excitation energy as heat, afforded appreciable yield gains (*14*). It is important to test whether a combination of engineered photorespiration with optimization of NPQ will enable additive yield gains. Realizing the yield gains afforded by the synthetic bypass in crops will require genetic engineering because the required enzymes are not present in plant genomes and hence cannot be targeted by breeding or genome editing technologies. ■

REFERENCES

1. T. J. Erb, J. Zarzycki, *Curr. Opin. Biotechnol.* **49**, 100 (2018).
2. H. Bauwe *et al.*, *Trends Plant Sci.* **15**, 330 (2010).
3. P. F. South *et al.*, *Science* **363**, eaat9077 (2019).
4. M. Eisenhut *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17199 (2008).
5. A. Bar-Even, *Plant Sci.* **273**, 71 (2018).
6. R. Kebeish *et al.*, *Nat. Biotechnol.* **25**, 593 (2007).
7. A. Maier *et al.*, *Front. Plant Sci.* **3**, 12 (2012).
8. T. R. Pick *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 3185 (2013).
9. A. P. Weber, A. Bräutigam, *Curr. Opin. Biotechnol.* **24**, 256 (2013).
10. C. R. Somerville, *Plant Physiol.* **125**, 20 (2001).
11. R. M. Benstein *et al.*, *Plant Cell* **25**, 5011 (2013).
12. A. Kozaki, G. Takeba, *Nature* **384**, 557 (1996).
13. M. Eisenhut *et al.*, *Mol. Plant* **10**, 47 (2017).
14. J. Kromdijk *et al.*, *Science* **354**, 857 (2016).

10.1126/science.aav8979

ATOMIC PHYSICS

Really cool neutral plasmas

Properties of laser-cooled neutral plasmas can be used to model high–energy-density plasmas

By **Scott Bergeson**

Plasmas are supposed to be hot. Hydrogen nuclei undergo fusion in the Sun because plasma temperatures and pressures are so high. On page 61 of this issue, Langin *et al.* (*1*) report on a completely different kind of plasma by photoionizing a laser-cooled gas of strontium atoms. The ion temperature is a chilly 0.05 K, so thermal speed of the ions is equivalent to a person taking a brisk walk. Surprisingly, the properties of this low-density, low-temperature plasma provide clues about the workings of high–energy-density physics relevant for fusion power research.

A very simple description of a plasma is that it is an ionized gas. In equilibrium, ionization occurs when the temperatures are high enough and when charged particles in the plasma are moving fast enough that collisions tear electrons away from their parent atoms and ions. The Boltzmann equation is the main tool for modeling the plasma environment (*2*). With a handful of approximations and extensions, this and related equations successfully describe processes used to create integrated circuits, light neon signs, and generate colorful flames. This success is somewhat surprising because the collisions occur through Coulomb interactions, which are long-range interactions and lead to many-body effects, but the Boltzmann equation is based on two-body collisions in a low-density environment. However, effective collision cross sections that include many-body effects can be calculated (with help from Chapman, Enskog, Bogoliubov, and others), so these kinetic theories can often give very accurate results (*3*, *4*).

Department of Physics and Astronomy, Brigham Young University, Provo, UT 84602, USA. Email: scott.bergeson@byu.edu

The foundation of this treatment rests on a hierarchy of length scales. The Debye length λ_D is the distance over which electrons rearrange their positions so that there is zero electric field inside the plasma. This length must be shorter than the extent of the plasma but longer than the average distance between ions in the plasma, the so-called Wigner-Seitz radius a_{WS}. These constraints ensure that there are many particles in a λ_D-sized sphere so that Boltzmann's statistical assumptions about collisions will hold. The Debye length must also be orders of magnitude longer than the classic distance of closest approach r_0, which can be thought of as the minimum distance between two ions in a head-on collision.

Treatments like the Boltzmann equation are valid when $\lambda_D >> a_{WS} >> r_0$. This regime is that of "ideal" plasmas, and all plasmas that obey these conditions are similar. An equivalent way to express these foundational assumptions in plasma science uses characteristic energies instead of lengths. The average kinetic energy per particle is approximately $KE \approx k_B T$, where k_B is Boltzmann's constant and T is the plasma temperature. For singly charged ions, the average nearest-neighbor electrical potential energy per particle is $U = e^2/(4\pi\varepsilon_0 a_{WS})$, where e is the electron charge and ε_0 is the permittivity of free space. The ratio of these two energy scales is called the strong-coupling parameter $\Gamma \equiv U/KE = r_0/a_{WS}$, basically the cube root of density divided by temperature. The basic assumption of kinetic plasma theories is that $\Gamma << 1$, that is, conditions of low density and high temperature.

When the foregoing hierarchy of length or energy scales is not met, the plasma is said to be "non-ideal" or "strongly coupled" (*5*). In fusion-class plasmas, for example, the non-ideal limit is approached in the early stages of the plasma evolution during compression and early heating of the system, and $\Gamma \sim 1$. The traditional concept of a collision becomes problematic because $\lambda_D \approx a_{WS} \approx r_0$.

The plasmas created by Langin *et al.* are similar to high–energy-density plasmas because they have comparable values of Γ (*6*). Thermodynamic properties of plasmas can be expressed in terms of Γ, so all plasmas with a given value of Γ are thermodynamically similar. Thus, these ultracold neutral plasmas can help probe the frontier of fusion science (see the figure). Measurements of collision properties (momentum transfer, thermal relaxation, diffusion, collision cross sections, Coulomb logarithms, and related quantities) can be made in these low-temperature, low-density plasmas and then directly applied to computer models of plasmas with similar values of Γ.

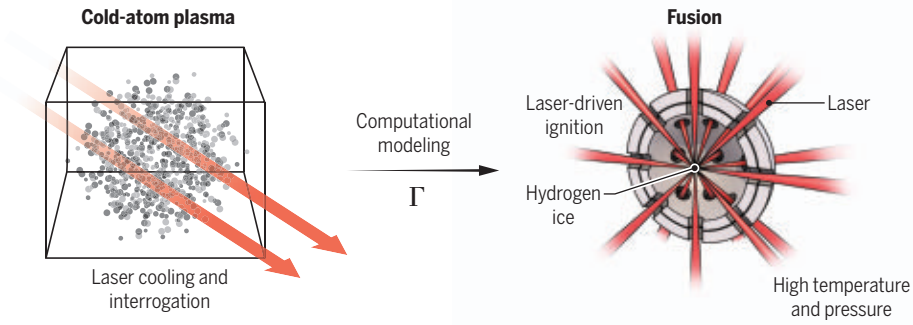
These plasmas do not constitute the first laser-cooled ions, which were reported by the groups of Dehmelt (*7*), Wineland (*8*), and others in the 1970s. Nor are these the coldest ions reported; the ion trapping and quantum information community achieve mean temperatures approaching the zero-point energy of the trap (*9*). Ultracold neutral plasmas have also been reported, and Killian and co-workers have contributed to the development of this field (*10*). These are, however, the coldest neutral plasmas yet reported, and neutrality means that the strong laboratory fields associated with ion trapping are absent.

The presence of those fields typically dominates the ion motion for trapped ions and obscures the underlying interesting plasma physics. Langin *et al.* have shown how to laser-cool ions in a neutral plasma, which overcomes a critical roadblock in the field of strongly coupled plasma physics. Although photoionized laser-cooled gases are initiated with essentially zero kinetic energy, the ions instantly experience strong accelerating forces from neighboring ions and heat up. This "disorder-induced heating" limits Γ to values near 2 (*11*), but laser-cooling the plasma ions makes it possible to manipulate the value of Γ. Thus, collision physics in this system can serve as a check on benchmark calculations. It also means these low-temperature plasmas can be used as simulators for high–energy-density plasmas.

In what seems like a paradox, the ultracold neutral plasmas of Langin *et al.* can help us understand collision parameters in high–energy-density plasma science. Under dense fusion plasma conditions, these parameters are nearly impossible to measure directly. Even with the best computer simulations, it is challenging to compute the values of these parameters with confidence. Extensions of kinetic theories into the strongly coupled regime, which are validated through modeling of plasmas, will facilitate computer modeling of more complex and technologically interesting plasmas. As laser-cooled plasmas become physically larger or reach longer confinement times, it may be possible to initiate and study classic plasma instabilities (*12*) or to initiate and characterize bump-on-tail distribution relaxations (*13*). These systems could also lead to higher-brightness focused-ion beam sources (*14*), which perhaps could be useful in ion implantation or x-ray source design. For very large values of $\Gamma = 172$, the ions will form a Coulomb crystal. Perhaps in that configuration, it will be possible to engineer massively entangled states useful for quantum computation or for high-precision metrology. ■

What cold plasmas can say about hot ones

The properties of neutral plasmas created by Langin *et al.* inform models of plasmas at higher temperature and pressure when they have similar values of the strong-coupling parameter Γ.



Making cold measurements
Spectroscopic studies of cold neutral plasmas allow measurement of many parameters that are difficult to obtain from high–energy-density plasmas.

Modeling hot problems
These parameters improve models of laser-driven nuclear fusion because the two plasmas have similar parameterized collision rates.

GRAPHIC: C. BICKEL/*SCIENCE*

REFERENCES

1. T. K. Langin, G. M. Gorman, T. C. Killian, *Science* **363**, 61 (2019).
2. P. L. Bhatnagar, E. P. Gross, M. Krook, *Phys. Rev.* **94**, 511 (1954).
3. S. D. Baalrud, J. Daligault, *Phys. Rev. Lett.* **110**, 235001 (2013).
4. L. G. Stanton, M. S. Murillo, *Phys. Rev. E* **93**, 043203 (2016).
5. R. P. Drake, *High-Energy-Density Physics* (Springer, ed. 2, Heidelberg, 2018).
6. T. K. Langin *et al.*, *Phys. Rev. E* **93**, 023201 (2016).
7. W. Neuhauser, M. Hohenstatt, P. Toschek, H. Dehmelt, *Phys. Rev. Lett.* **41**, 233 (1978).
8. D. J. Wineland, R. E. Drullinger, F. L. Walls, *Phys. Rev. Lett.* **40**, 1639 (1978).
9. T. Rosenband *et al.*, *Science* **319**, 1808 (2008).
10. T. C. Killian *et al.*, *Phys. Rev. Lett.* **83**, 4776 (1999).
11. Y. C. Chen *et al.*, *Phys. Rev. Lett.* **93**, 265003 (2004).
12. P. W. Terry, *Rev. Mod. Phys.* **72**, 109 (2000).
13. D. V. Dylov, J. W. Fleischer, *Phys. Rev. Lett.* **100**, 103903 (2008).
14. D. Murphy *et al.*, *Nat. Commun.* **5**, 4489 (2014).

10.1126/science.aau7988

POLICY FORUM

OCEAN GOVERNANCE

From voluntary commitments to ocean sustainability

A common pledge and review system is needed

By **Barbara Neumann** *and* **Sebastian Unger**

Voluntary commitments by states, governmental or nongovernmental organizations, and other actors, aiming to deliver outcome-oriented activities, have become a well-recognized mechanism in international sustainability policy (*1*–*3*). For ocean governance, the calling for and pledging of voluntary commitments could become a game changer, with two major international processes harnessing such voluntary contributions in recent years: the Our Ocean conferences, an annual high-level series initiated by U.S. Secretary of State John Kerry in 2014, and the United Nations (UN) Ocean Conference, which took place for the first time in June 2017. Such calls and commitments provide opportunities to raise awareness, promote engagement, and catalyze political will for action on the part of states as well as public and private sectors. However, without effective and transparent review systems, it is difficult to link pledged commitments to actual implementation. Quality control and ensuring that commitments are effective and impactful will be difficult to achieve. A uniform global process is required to register and assess commitments, including consistent reporting and monitoring systems with clear targets, baselines, and review systems.

Institute for Advanced Sustainability Studies (IASS), 14467 Potsdam, Germany. Email: sebastian.unger@iass-potsdam.de

A SURGE OF OCEAN COMMITMENTS

The UN Ocean Conference had encouraged state and nonstate actors to submit commitments to advance implementation of Sustainable Development Goal (SDG) 14 and associated targets (*4*). Part of a comprehensive framework of 17 interlinked goals under the UN's 2030 Agenda for Sustainable Development (*5*), SDG 14 calls on states and the global community to "conserve and sustainably use the oceans, seas and marine resources for sustainable development." It is underpinned by 10 specific targets addressing marine pollution, conservation, ocean acidification, fisheries, benefits for Small Island Developing States, small-scale fisheries, scientific knowledge and marine research, and international law.

More than 1300 voluntary commitments for ocean action, such as measures for combating marine pollution or strengthening capacity for marine research, were made at the UN Ocean Conference by governments, the UN system, civil society organizations, academia, the scientific community, and the private sector (*6*). Despite the open call for contributions, the majority of commitments registered were still made by governmental actors and civil-society organizations (*7*). But they also include innovative initiatives from the private sector and philanthropic organizations such as the Tuna 2020 Traceability Declaration or the Seafood Business for Ocean Stewardship (SeaBOS) platform, which seek change both through collective action and indirect appeals to countries or intergovernmental bodies to adopt ocean governance reforms. By November 2018, the number of contributions registered through the UN's registry of voluntary commitments (*8*), a web-based site that remains open for new registrations and updating on progress, had grown to 1478.

Commitments have been made to improve tracking of products from tuna (such as Atlantic bluefin tuna) from vessel to final buyer.

The pledging of voluntary commitments across government, civil society, and the private sector also stands at the heart of the Our Ocean conferences. Although not directly linked to the 2030 Agenda, the Our Ocean Conference series is complementing efforts of the UN process and has a strong topical relationship to SDG 14 (*9*). In total, 305 commitments for action were announced at the 2018 conference, covering six topical strands: marine protected areas, climate change, sustainable fisheries, marine pollution, sustainable blue economy, and maritime security (*10*). At the Our Ocean Conference 2017, hosted by the European Union in Malta, 437 announcements for "tangible and measurable commitments" had been made toward ocean health and sustainability (*9*, *11*), including a large number from the private sector. Although smaller in the number of commitments than the UN process, the Our Ocean conferences succeed particularly in mobilizing financial resources or pledges for creation of new marine protected areas.

A TRANSFORMATIVE TOOL?

Though not replacing state measures to implement legally binding agreements, we believe that voluntary commitments hold great additional potential for driving transformative change for the ocean. They mobilize actions and means for improving ocean health, support the creation of new partnerships across different sectors and actor groups, and facilitate learning processes and exchange of innovative practice. By lowering barriers to address complex cross-cutting problems, the nonbinding nature of voluntary commitments also helps to overcome established but problematic sectorial approaches in ocean governance. Voluntary commitments also create normative pressure (*1*, *12*) and increase expectations to play an active role in improving ocean health.

However, central oversight is needed to ensure that promises are kept. Without a transparent and rigorous pledge and review system for all ocean-related commitments, there is a risk of double-announcing in various forums or creating a flurry of low-impact or short-term activities that do not deliver progress on targets. Other critical challenges

associated with voluntary commitment processes are accountability, enforcement, effectiveness, and progress accounting (*2*, *13*).

The pledging processes under the UN Ocean Conference and the Our Ocean series seek to address these challenges to a certain extent through their individual registrations procedures that request the formulation of commitments along defined criteria. Our Ocean 2018, for the first time, published a report seeking to describe progress on commitments made under previous Our Ocean conferences (*14*), and the UN's registry of voluntary commitments invites pledging entities to provide updates on progress through their website (*8*). But the two registry systems are neither harmonized in terms of the data gathered nor in the standards for acceptance and registration of commitments. This lack of robust and consistent tracking and reporting processes, and missing links to existing environmental baseline data, impede assessments of transformative effects and overall progress toward goals.

A centralized registry, however, is needed to learn whether voluntary commitments produce desired outcomes on the ground, to identify trends, and to facilitate adjustments of policies. If kept separately, a competition between pledging systems with diverging objectives, different standards, and a general lack of monitoring and accountability may obscure their potential for improving ocean health and governance.

BUILDING BLOCKS FOR POST-2020

As 5 of 10 SDG 14 targets mature in 2020 and—possibly with the exception of the target to conserve at least 10% of coastal and marine areas (*7*)— will most likely not be achieved by then, the coming years will be critical for achieving the ocean goal. A credible post-2020 strategy is therefore needed to support the implementation of SDG 14, ideally harmonized with the post-2020 biodiversity framework currently developed under the Convention on Biological Diversity (CBD). One of the key building blocks for such a strategy could be a unified and comprehensive global registry for voluntary commitments. The existing pledging schemes and registries of voluntary commitments under the UN Ocean and the Our Ocean conferences would lend themselves as strong starting points to developing such a global registry and reporting mechanism.

The role of this new system would be to take stock of voluntary commitments; report on progress on implementation; provide transparency and independent verification; provide joint quality criteria for voluntary commitments; identify trends and highlight thematic and geographical gaps; and analyze distance and progress to SDG 14 and other ocean-related SDGs and targets.

To identify trends and to measure distance and progress to targets, this global registry should be linked to baseline data in existing databases and assessment processes on the state of the marine environment. For example, the World Database on Protected Areas, which is run by the UN Environment World Conservation Monitoring Centre (UNEP-WCMC) and the International Union for Conservation of Nature (IUCN), the UN's World Ocean Assessment, or the UN Food and Agriculture Organization's world fishery and aquaculture statistics could provide the necessary information.

In addition, independent scientific data and assessments such as the Ocean Health Index (*15*) or the MPAtlas of the Marine Conservation Institute could be taken into account. The UN Decade of Ocean Science for Sustainable Development (2021 to 2030), an upcoming global effort to boost international cooperation in ocean sciences, could help to further strengthen the knowledge base of the review process.

Under the proposed strategy and a central registry, pledging of voluntary commitments could take place throughout the year and be highlighted on an annual basis at the Our Ocean conferences or other high-level meetings of states and relevant actors such as the UN Environment Assembly. Comparability of commitments and assessment of impact and progress would be facilitated through common reporting formats. The registry would be evaluated and commitments assessed every 3 years at the UN Ocean Conference, providing an accountability moment for the global community. This would also aid the steering of calls for action into directions where topical or geographical gaps have been identified, and the aligning of actions on international, regional, and national scales. Bringing together these different types of data and information, such a common pledge and review system could determine whether the global community is on track to achieve the goals set for the ocean and help to orchestrate further action.

The registry should be hosted by an international body and be maintained and updated regularly in close cooperation with competent global and regional organizations, ensuring transparent access to data and information. This could be supplemented by independent reviews from scientific institutions and nongovernmental organizations. The registry would provide grounds for developing and applying indicators and analytical frameworks for monitoring and evaluating performance and impacts, and assist in sharing of good practices.

Regular assessments of the pledge and review process for voluntary ocean commitments should be reported to and assessed by the UN's High-level Political Forum on Sustainable Development and reflected in the Global Sustainable Development Report. The registry should also seek synergies with reporting systems for other goal-based policy frameworks such as the UN Paris Agreement on climate change.

Both the UN Ocean Conference process and the Our Ocean series will continue to collect voluntary commitments toward the next UN Ocean Conference planned for 2020 and the upcoming Our Ocean conferences in Norway (2019) and Palau (2020). And there have been first discussions of possible coordination of the two commitment systems, a promising prospect for developing an orchestrated post-2020 strategy for ocean sustainability with a uniform global pledge and review system for voluntary ocean commitments. ■

REFERENCES AND NOTES

1. B. Guy, *Econ. Soc. Rev. (Irel.)* **45**, 223 (2014).
2. P. Pattberg, O. Widerberg, *Ambio* **45**, 42 (2016).
3. M. Beisheim, A. Ellersiek, "Partnerships for the 2030 Agenda for Sustainable Development: Transformative, inclusive and accountable?" (SWP Research Paper, Berlin, 2017).
4. United Nations, "Voluntary Commitments for the implementation of Goal 14" (2017); https://sustainabledevelopment.un.org/content/documents/12816The%20Ocean%20Conference%20-%20Guidance%20on%20Voluntary%20Commitments%20final.pdf.
5. United Nations, "Transforming our world: The 2030 Agenda for Sustainable Development. UNGA Resolution A/RES/70/1" (Resolution adopted by the General Assembly on 25 September 2015 70/1, New York, 2015).
6. IISD, "Summary of The Ocean Conference: 5-9 June 2017" (Earth Negotiations Bulletin, vol. 32, 2017); http://enb.iisd.org/oceans/sdg14conference/enb/.
7. Division for Sustainable Development, M. Vierros, R. Buonomo, "In-depth analysis of Ocean Conference Voluntary Commitments to support and monitor their implementation. 14 Life Below Water" [Department of Economic and Social Affairs (DESA)], United Nations, New York, 2017); https://sustainabledevelopment.un.org/content/documents/17193OCVC_in_depth_analysis.pdf.
8. United Nations, "The Ocean Conference Registry of Voluntary Commitments" (2018); https://oceanconference.un.org/commitments/.
9. European Commission, "Our Ocean Conference 2017. Final report" (2017); www.ourocean2017.org/sites/default/files/ooc-2017-report.pdf.
10. Ministry of Marine Affairs and Fisheries, Republic of Indonesia, "Our Ocean Commitments" (2018); https://ourocean2018.org/?l=our-ocean-commitments.
11. European Commission, "Our Ocean 2017 commitments" (2017); www.ourocean2017.org/sites/default/files/ooc-2017-list-of-commitments_en.pdf.
12. M. Stafford-Smith *et al.*, *Sustain. Sci.* **12**, 911 (2017).
13. J. Foti, "Promises kept: Ensuring ambition and accountability through a Rio+ 20 'compendium of commitments'" (Working paper, WRI, 2012).
14. Ministry of Marine Affairs and Fisheries, Republic of Indonesia, "Progress of Our Ocean Conference Commitment" (2018); https://ourocean2018.org/assets/files/progress_commitment.pdf.
15. B. S. Halpern *et al.*, *PLOS ONE* **12**, e0178267 (2017).

ACKNOWLEDGMENTS

We thank participants of the 2017 Potsdam Ocean Governance Workshop for their contributions on the topic, and are grateful to the following experts for their input and reflections: M. Caldwell, J. Hammersland, D. Herr, M. Knigge, M. Kobayashi, A. Mondré, H. Schopmans, and T. Thiele. We thank L. von Pogrell, J. Pütz, and S. Heinecke for support in researching data. This work is supported by the German Federal Ministry of Education and Research (BMBF) through its Research for Sustainable Development program (FONA), and the Federal State of Brandenburg.

10.1126/science.aav5727

NATO's efforts stopped short of confronting the environmental degradation caused by warfare.

PHOTO: WIKIMEDIA COMMONS

HISTORY OF SCIENCE

A military alliance goes green

Seeking solutions to Cold War divisions, in the mid-20th century NATO embraced environmentalism

By **Dagomar Degroot**

War and preparation for war have long led militaries to exploit, transform, and degrade environments. How ironic, then, that at the height of the Cold War, the North Atlantic Treaty Organization (NATO)—the most powerful military alliance ever assembled—emerged as a leading proponent of environmentalism. Simone Turchetti's *Greening the Alliance* is the first book to explain the surprising rise, repeated revision, and possible decline of NATO's environmental research program.

Turchetti organizes his book chronologically. After insightful passages on sources and methods, he traces the diplomatic tensions and maneuverings that, in 1958, led representatives of the United States and Britain to cooperate in securing support for a new NATO Science Committee. Supporting environmental research, they hoped, would promote parallel diplomacy to repair deepening divisions among the alliance's 12 signatories. Yet because the committee sponsored research with obvious strategic applications, it actually exacerbated divisions among the allies.

Greening the Alliance
Simone Turchetti
University of Chicago Press, 2018. 263 pp.

In 1966, Turchetti argues, growing dissatisfaction with NATO's science program came to a head. The rise of environmentalism, the sinking of the tanker *SS Torrey Canyon*, and the political opportunism of Richard Nixon all led to an American push for a new kind of environmental research at NATO. Enter the "Committee on the Challenges of Modern Society" (CCMS), established in 1969 to fund research into toxicology, devastated ecosystems, and environmental monitoring. Research pioneered by the committee steadily advanced a brand of environmentalism that prized scientific rationalism rather than the radical counterculture of the grassroots environmentalist movement.

Yet, Turchetti explains that European allies repeatedly stalled CCMS initiatives, which in any case never confronted the environmental impact of military activities. When the CCMS exposed new divisions within NATO, American support withered, and the alliance's commitment to environmentalism seemed to fade. By that time, however, emerging weapons systems in the United States and the Soviet Union increasingly demanded constant environmental surveillance. This new reality encouraged another round of environmental research, with obvious strategic applications.

After the rise of Ronald Reagan and Margaret Thatcher led NATO to again de-emphasize environmental research, the collapse of the Soviet Union led the CCMS to support efforts to rehabilitate defunct Cold War military bases. Environmental research at NATO lives on but faces challenges amid the alliance's increasing focus on the "cybersphere."

In the end, Turchetti convincingly argues not only that militaries fostered the emergence of modern environmental sciences but also that they long set the agenda for major research programs in its disciplines. His analysis of the American role within NATO is especially compelling. Far from a Cold War hegemon, the United States in *Greening the Alliance* could only achieve its objectives with the cooperation of often-reluctant allies. Skeptics will note, however, that American officials repeatedly determined when, and how, NATO would pursue environmental research and policy.

The protagonists of *Greening the Alliance* are scientists in the upper echelons of Cold War military and diplomatic institutions. This focus makes for a richly detailed story of political maneuvering and high-minded ideals, yet it also deprives Turchetti's narrative of context that might have given it greater significance. Politicians and military officers are rarely mentioned by name, and ordinary people who may have shaped the course of scientific research, such as North Sea fishermen resistant to early oceanographic surveys, rarely receive much attention.

Key scientific controversies discussed in *Greening the Alliance*—the possibility of nuclear winter, for example—frequently appear with little explanation, and Turchetti opts not to compare NATO's environmental programs with similar and simultaneous efforts in other western institutions. (Neil Maher has recently revealed that NASA similarly struggled to implement an "environmental turn" in the 1970s, for example.) Moreover, because Turchetti rarely explains the scope and significance of NATO's environmental science program within the broader development of 20th-century science, it is difficult for the reader to know just how important the activities of the CCMS really were.

Written in workmanlike prose, *Greening the Alliance* is therefore a book that will primarily appeal to a relatively small group of historians and political scientists. Yet for those specialists, it succeeds in telling a new and critically important story. ■

The reviewer is at the Department of History, Georgetown University, Washington, DC 20057, USA. Email: dd865@georgetown.edu

10.1126/science.aav1863

ECONOMICS

Robots, telework, and the jobs of the future

Globalization and AI are primed to disrupt tomorrow's workplace, argues an economist

By **Jon Peha**

In the conference room of a popular San Francisco–based magazine, journalists and editors walk through the door, gathering for a staff meeting. One colleague, however, rolls in on two large motorized wheels.

"EmBot" is a human-sized robot that shows live video of Emily Dreyfus, a staff writer who lives 3000 miles away. But EmBot is more than a face on a screen. It can turn toward whomever is speaking or chase a colleague of Dreyfus's down the hall. EmBot is an example of the telepresence technology that Richard Baldwin, author of *The Globotics Upheaval*, believes will cost many workers in wealthy nations their jobs, allowing them to be replaced with "telemigrants" from abroad.

But even telemigrants face stiff competition in the emerging economy. EmBot has software-based cousins that use artificial intelligence (AI) to perform tasks we once thought required highly skilled humans, from searching for the legal precedents of a patent dispute to spotting cancer in a magnetic resonance image.

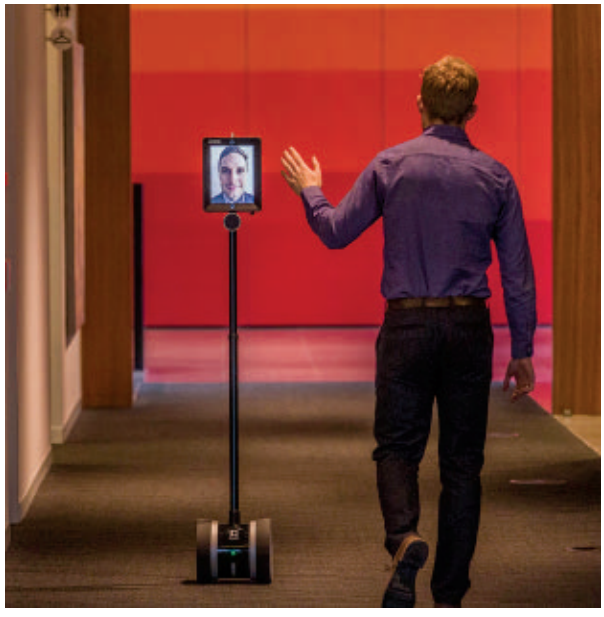
Baldwin's thesis is that globalization and AI robots constitute a "globotics" tidal wave that will shake the foundation of middle-class prosperity in wealthy nations. This, he argues, will lead to social upheaval, just as 19th-century steam engines and mechanical looms created workplace disruptions that brought workers to the streets in sometimes violent protest.

Although globotics will improve productivity and create new jobs, by some estimates cited in the book, it could also replace more than half of current jobs in a developed economy. The effects will be uneven.

Some skilled workers will prosper as they gain the ability to compete for jobs across the globe; others will become obsolete. An accountant whose primary competitive advantage is that she lives within easy driving distance of a commercial hub, for example, will now have to compete with accountants around the world. An accountant whose primary advantage is the ability to detect complex patterns in financial data will have to compete with machine-learning software that never sleeps or asks for a raise.

For people hoping to choose a profession that won't soon be replaced, Baldwin recommends those that require physical proximity and that take advantage of distinctly human qualities, such as creativity, social awareness, ethics, and empathy. AI robots cannot (yet) write articles for *Science*, for example, and they are terrible preschool teachers.

Disruptive change can lead to backlash.



A remote reporter for the *Toronto Star* greets a co-worker.

Baldwin argues that 2016 vote outcomes favoring Brexit in the United Kingdom and Donald Trump in the United States were backlashes to globotics. Although this was not a period of high unemployment—Trump was elected after 7 years of solid economic growth that drove the U.S. unemployment rate from 10 to 4.6%—it is possible that globotics could have contributed to low growth in wages.

Although the book offers valuable insights into the long-term impact that globalization and AI will have on workers, the case that globotics will bring upheaval is less convincing. Upheaval occurs when technology advances at a pace that is too fast for society to absorb. Baldwin portrays the progress of telepresence and AI as suddenly becoming rapid, but these technologies have been progressing for decades.

Thanks to the vast expansion of undersea fiberoptic cables in the 1990s, widespread adoption of the Internet, and continuous improvement in collaboration software, we already live with many effects of telepresence. Multinational companies routinely hire the most competitive workers anywhere in the world and use this technology to move jobs to workers rather than moving workers to jobs. Similarly, AI algorithms have already replaced humans in many endeavors, including monitoring everything from surveillance video to credit card purchases, just as websites have all but replaced travel agents.

So far, there has been workplace change but little true upheaval. Both telepresence and AI technology are still making impressive progress, but *The Globotics Upheaval* provides no way to judge whether the pace of advancement will eventually exceed what society can absorb.

Nonetheless, Baldwin presents a compelling view of the future of work and the serious challenges ahead while there is still time to prepare. He wisely argues that we must protect workers, without necessarily protecting specific jobs as they become outdated, and that we must do more to help those who've been displaced by technology reenter the workforce and offer such individuals a strong safety net along the way.

I would add education reform to the prescription: Moving forward, schools and universities should teach students to work with emerging technology rather than compete against it. ■

The Globotics Upheaval
Globalization, Robotics, and the Future of Work
Richard Baldwin
Oxford University Press, 2019. 300 pp.

The reviewer is at the Department of Engineering and Public Policy and the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA. Email: peha@cmu.edu

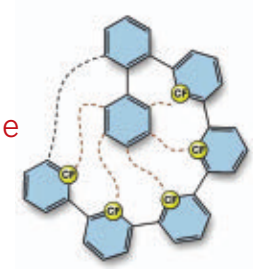
10.1126/science.aav6273

PHOTO: RICK MADONIK/*TORONTO STAR*/GETTY IMAGES

RESEARCH

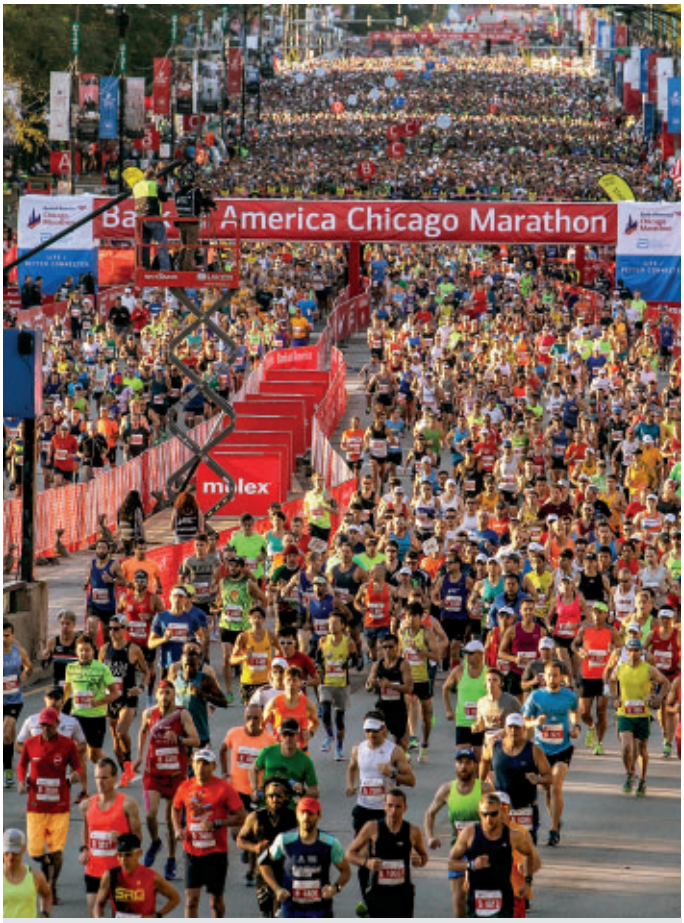
Fluorinated aryl groups couple to form nanographenes

Kolmer et al., *p. 57*



IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**



Runners moving down Columbus Drive at the Chicago Marathon, October 2017

CREDITS: (TOP TO BOTTOM) KOLMER *ET AL.*; DYLAN BUELL /STRINGER/GETTY IMAGES

CROWD DYNAMICS

A crowd that flows like water

The behavior of large numbers of insects, animals, and other flocks is often based on rules about individual interactions. Bain and Bartolo applied a fluid-like model to the behavior of marathon runners as they walked up to the start line of the Chicago Marathon (see the Perspective by Ouellette). They observed nondamping linear waves with the same speed for different starting corrals of runners and at different races around the world. Their model should apply both to this type of polarized crowd as well as to other groups, which may help guide crowd management. —BG

Science, this issue p. 46; see also p. 27

VIROLOGY

Mobile detection of Lassa virus

Lassa fever is a hemorrhagic viral disease endemic to West Africa. Usually, each year sees only a smattering of cases reported, but hospitalized patients risk a 15% chance of death. Responding to fears that a 10-fold surge in cases in Nigeria in 2018 signaled an incipient outbreak, Kafetzopoulou *et al*. performed metagenomic nanopore sequencing directly from samples from 120 patients (see the Perspective by Bhadelia). Results showed no strong evidence of a new strain emerging nor of person-to-person transmission; rather, rodent contamination was the main source. To prevent future escalation of this disease, we need to understand what triggers the irruption of rodents into human dwellings. —CA

Science, this issue p. 74; see also p. 30

PROTEIN TRANSLOCATION

Posttranslational translocon architecture

About a third of proteins are transported into endoplasmic reticulum by the universally conserved Sec61 protein-conducting channel. Itskanov and Park determined a cryo–electron microscopy structure of the Sec complex from yeast, which mediates posttranslational translocation of many secretory proteins across the endoplasmic reticulum membrane. The study reveals how Sec63 activates the Sec61 channel for substrate polypeptide insertion. The structure also explains the mutually exclusive binding of Sec63 and the ribosome to the channel. —SMH

Science, this issue p. 84

CHEMICAL PHYSICS

C_{60} at high resolution

It generally takes more energy for molecules to vibrate than to rotate. A vibrational absorption band thus encompasses many distinct concurrent rotational transitions, but these tend to blur together when the molecules have more than a few atoms. Changala *et al*. succeeded in cooling C_{60} fullerenes sufficiently to obtain rotational resolution within a C–C stretching band. Success hinged on careful optimization of argon buffer gas flow. Such quantum state–resolved features could aid characterization of fullerene-type compounds in exotic environments such as interstellar space. —JSY

Science, this issue p. 49

ATOMIC PHYSICS

Making a strongly coupled plasma

Plasmas—gases of ionized atoms and electrons—are naturally formed at high temperatures, such as those reached in the interiors of stars. Describing plasmas theoretically is tricky when they are in the strongly coupled regime; reaching that regime in the laboratory would provide a valuable benchmark for theory. To that end, Langin *et al*. worked with a cold plasma created out of atoms of strontium that were ionized by laser light (see the Perspective by Bergeson). They used lasers to cool the ions down to about 50

millikelvin, reaching the desired strongly coupled regime. —JS

Science, this issue p. 61; see also p. 33

PALEONTOLOGY

A proto-mammalian giant

Early terrestrial amniotes evolved into two groups: the sauropsids, which led to the bird and dinosaur lineages, and the synapsids, which led to mammals. Synapsids were diverse during the Permian but were greatly reduced after the end-Permian extinction (about 252 million years ago). The few groups that survived into the Triassic were mostly small and retained a sprawling gait. Sulej and Niedźwiedzki, however, describe a dicynodont from the Late Triassic of Poland that is as large as some coexisting dinosaurs and appears to have had an erect gait—like modern mammals. Thus, megaherbivores in the Triassic were not only dinosaurs. —SNV

Science, this issue p. 78

SOCIAL SCIENCES

Deadliest 100 days of the Holocaust

More than 25% of the approximately 6 million Jews murdered during the Holocaust were killed in one 100-day period in 1942. Stone used an unusual dataset of railway transportation records to show that during this period, the Nazis murdered more than 1.47 million Jews, a kill rate that is 10 times higher than previous estimates. Contradicting contemporary analyses of the Holocaust, the author shows that Operation Reinhard was exceptionally violent in its extreme kill rate, number, and proportion of the population murdered, even when compared to other 20th-century genocides. —PJB

Sci. Adv. 10.1126/sciadv.aau7292 (2018).

Transport records have been analyzed to estimate the number of Jews murdered by the Nazis in 1942.

PROKARYOTIC IMMUNITY

Additional, diverse CRISPR systems

CRISPR systems have been revolutionizing molecular biology. Mining the metagenomic database, Yan *et al*. systematically discovered additional subtypes of type V CRISPR-Cas systems. The additional Cas12 effectors displayed a range of activities, including target and collateral cleavage of single-stranded RNA and DNA, as well as double-stranded DNA nicking and cleavage. These diverse nuclease activities suggest how an ancient transposase may have evolved into various type V effectors and expand the nucleic acid detection and genome-editing toolbox. —SYM

Science, this issue p. 88

DRUG DEVELOPMENT

A long-lasting poison scavenger

Nerve agents are neurotoxic compounds found in pesticides and chemical weapons. They act by blocking the transmission of nerve impulses to the muscles, and exposure can be fatal within minutes. Zhang *et al*. developed a nanoparticle-based bioscavenger that breaks down organophosphate nerve agents into innocuous compounds. Prophylactic treatment of rats and guinea pigs confirmed low immunogenicity and good biodistribution. Treated animals were protected from repeated exposure to the nerve agent sarin over 7 days. This nanoscavenger might thus help prevent nerve-agent poisoning in at-risk subjects. —MM

Sci. Transl. Med. **11**, eaau7091 (2019).

IN OTHER JOURNALS

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

A Raggiana bird-of-paradise from the Southern Highlands, Papua New Guinea

HUMAN GENETICS

Alzheimer's disease in admixed people

Several genes have been identified that increase the risk of late-onset genetic disorders, such as Alzheimer's disease (AD). Specifically, the *ApoE* ε4 allele is associated with a higher risk of developing AD. However, individuals of African ancestry that carry this variant appear to be less prone to developing AD. Rajabli *et al*. examined AD cases and controls in admixed individuals of Puerto Rican and African-American descent and found that individuals who carried an African *ApoE* ε4 background had less risk of developing the disease. It seems the African variant of *ApoE* ε4 contains protective genetic variants. —LMZ

PLOS Genet. **14**, e1007791 (2018).

PLANT SCIENCE

Essential metal for plants

Although zinc (Zn) is an essential micronutrient for plants and humans, much of the world's agricultural land is deficient in Zn. Sinclair *et al*. studied plants that are unable to deliver Zn into their own xylem. The plant shoots were thus internally starved regardless of whether Zn was available from the root. The Zn-starved shoots signaled to roots to increase Zn supplies. In response, the roots up-regulated expression of the genes encoding metal transport/tolerance protein 2 (MTP2) and heavy metal ATPase 2 (HMA2). Local Zn deficiency in roots left these same genes unaffected. It seems that Zn taken up in lateral roots is transported into the endoplasmic reticulum by MTP2, thus gaining access to the intercellular symplastic network. The Zn then progresses from outer epidermal cells toward the core of the root, where it is exported by HMA2 into the xylem for transport to the shoot. The shoot asks for what it needs, and the root delivers. —PJH

Plant Cell **30**, 2463 (2018).

SKIN

Roots of acne

Most people experience a bout of acne at some stage in their life. For an unlucky few, the skin

PHOTOS: (LEFT TO RIGHT) SHAWSHOTS/ALAMY STOCK PHOTO; NATIONAL GEOGRAPHIC IMAGE COLLECTION/ALAMY STOCK PHOTO

EVOLUTION

Sing on high, dance on the floor

The frugivorous, polygamous, and wildly glamorous birds of paradise are a puzzle to evolutionary biologists. What is sexual selection acting on to result in such extremely visual, behavioral, and aural diversity among these related species? Ligon *et al*. analyzed 961 video clips, 176 audio clips, and 393 museum specimens. They concluded that females are selecting on the combined sensory assault from song, display, and plumage color, resulting in a "courtship phenotype." Although all elements are required for successful courtship, there is room for variation depending on environmental constraints. Song predominates in the canopy, where it is unimpeded by twigs and branches, whereas flashy behavioral display is most effective on the gloomy forest floor. —CA

PLOS Biol. **16**, e2006962 (2018).

condition feels relentless and can evade treatment. Petridis *et al*. performed a DNA study of individuals with acne vulgaris and found that those affected share similar, but surprising, genetic mutations. Homing in on 15 regions of the genome, they identified a series of culprit genes that controlled hair growth and follicle formation. This discovery lends weight to the idea that hair follicle shape creates a milieu susceptible to bacterial colonization and inflammation. —PNK

Nat. Commun. **9**, 5075 (2018).

PLASTIC POLLUTION

Scallops seasoned with nanoplastics

Microplastics are present in marine environments worldwide. As these particles break down further, they form nanoplastics, which are harder to detect. Nanoplastics also can enter the environment directly from commercial products such as paints and cosmetics. Al-Sid-Cheikh *et al*. investigate the uptake of such nanoplastics by scallops at predicted environmental concentrations. The authors use radiocarbon labeling to track the nanoplastics within the scallop tissues. Uptake differs depending on particle size: Larger nanoparticles accumulate in the intestine, whereas smaller nanoparticles are dispersed through the entire scallop body. After exposure to nanoplastics ceased, smaller nanoparticles were no longer detected after 14 days, but some larger nanoparticles persisted for more than 48 days. The presence of the smaller nanoparticles in muscle tissue suggests that the particles can cross epithelial membranes. —JFU

Environ. Sci. Technol. **52**, 14480 (2018).

The great scallop (*Pecten maximus*) reveals details about the uptake of nanoplastics by marine organisms.

PHOTO: NATURE PICTURE LIBRARY/ALAMY STOCK PHOTO

RELATIONSHIP SCIENCE

Ending a relationship

When deciding to end a relationship, people may consider the feelings of their partners as well as their own. Joel *et al*. investigated whether decisions to break up are driven in part by perceptions of a partner's dependence on the relationship. They found that participants were less likely to initiate a breakup with their partners when they felt that their partners were more dependent on the relationship for psychological well-being, even when participants were unsatisfied in the relationship. Even participants who were actively considering breaking up with their partners were less likely to do so if they felt their partners depended on the relationship. These results suggest that people exhibit costly, prosocial preferences in relationships even when they may wish to leave them. —TSR

J. Pers. Soc. Psychol. **115**, 805 (2018).

OPPORTUNITY DENIED

The inequality of innovation

A lack of social capital can undermine a child's likelihood of becoming an inventor, regardless of her inventive ability. Using U.S. patent records for 1.2 million inventors, combined with tax records and other data, Bell *et al*. show how children from high-income families are several times more likely to become inventors than those from lower-income families, even when they have comparable math abilities. Children who grow up in areas where innovation and patenting are more common are more likely to patent as well, and particularly in the same class of technologies that had a high innovation rate in their childhood communities. —BW

Quart. J. Econ. 10.1093/qje/qjy028 (2018).

ALSO IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

OPTICS

Exceptional points in optics

Many complex systems operate with loss. Mathematically, these systems can be described as non-Hermitian. A property of such a system is that there can exist certain conditions—exceptional points—where gain and loss can be perfectly balanced and exotic behavior is predicted to occur. Optical systems generally possess gain and loss and so are ideal systems for exploring exceptional point physics. Miri and Alù review the topic of exceptional points in photonics and explore some of the possible exotic behavior that might be expected from engineering such systems. —ISO

Science, this issue p. 42

NEUROSCIENCE

Forgetting and receptor removal

The trafficking of AMPA receptors to and from the surface of postsynaptic membranes regulates synaptic strength and underlies learning and memory. Awasthi *et al*. found that the integral membrane protein synaptotagmin-3 (Syt3) is predominantly found on postsynaptic endocytic zones of neurons, where it promotes AMPA receptor internalization (see the Perspective by Mandelberg and Tsien). In Syt3 overexpressing or knockdown neurons, synaptic transmission and short-term plasticity were unchanged. However, in neurons from Syt3 knockout mice, synaptic long-term depression was abolished and decaying long-term potentiation endured. In Syt3 knockout mice, spatial learning was unaltered; however, these animals showed signs of impaired forgetting and relearning during the water maze spatial memory task. —PRS

Science, this issue p. 44; see also p. 31

IMMUNOLOGY

Commensal-specific T cells are flexible

Barrier tissues, like the skin, are sites where noninvasive commensal microbes constantly interact with resident T cells. These encounters can result in commensal-specific T cell responses that promote, for example, host defense and tissue repair. Harrison *et al*. show that subsets of skin-resident commensal-specific interleukin-17A–producing $CD4^+$ and $CD8^+$ T cells have a dual nature: They coexpress transcription factors that direct antagonistic antimicrobial (type 17) and antiparasite and pro–tissue repair (type 2) programs. When skin is damaged, epithelial cell alarmins license type 17 T cells to turn on type 2 cytokines. Thus, commensal-specific type 17 T cells can direct antimicrobial activity under homeostatic conditions but rapidly turn on tissue repair in the context of injury. —STS

Science, this issue p. 43

PLANT SCIENCE

Fixing photosynthetic inefficiencies

In some of our most useful crops (such as rice and wheat), photosynthesis produces toxic by-products that reduce its efficiency. Photorespiration deals with these by-products, converting them into metabolically useful components, but at the cost of energy lost. South *et al*. constructed a metabolic pathway in transgenic tobacco plants that more efficiently recaptures the unproductive by-products of photosynthesis with less energy lost (see the Perspective by Eisenhut and Weber). In field trials, these transgenic tobacco plants were ~40% more productive than wild-type tobacco plants. —PJH

Science, this issue p. 45; see also p. 32

NANOMATERIALS

Nanographenes on oxides

The growth of nanographene islands and ribbons on metal surfaces can be accomplished on single-crystal metal surfaces through carbon-carbon coupling reactions, but the surfaces of oxides do not assist these reactions. Kolmer *et al*. show that fluorinated aryl groups can be coupled to form nanographenes on the rutile surface of titanium oxide. The fluorine substitution of the aryl groups was selected so that as the carbon-fluorine bonds were thermally activated, a stepwise process sequentially added aromatic rings around a central aryl group until it was completely substituted. —PDS

Science, this issue p. 57

MESOSCOPIC PHYSICS

A backward current

Two-dimensional materials in a magnetic field can exhibit the so-called quantum Hall effect. This regime is characterized by currents running along the edge of the sample in the "downstream" direction determined by the sign of the magnetic field. Lafont *et al*. studied electrical transport in GaAs-AlGaAs heterostructures, focusing on a previously less-studied spin-unpolarized state in the fractional quantum Hall regime. By considering various experimental configurations, they observed a component of the charge current flowing in the opposite, "upstream" direction. —JS

Science, this issue p. 54

NEUROSCIENCE

Another primary visual cortex

Most functional studies in the visual system have focused on the cortical representation of the geniculo-striate pathway that links the retina to the cortex. The parallel collicular pathway is believed to sparsely project throughout the visual cortex and have a modulatory role on cortical responses to visual stimuli. Beltramo and Scanziani found a visual cortical area that is entirely dedicated to the superior colliculus. This area can discriminate moving visual stimuli that the "classical" primary visual cortex cannot. Thus, the superior colliculus, a phylogenetically ancient structure, has its own projection in neocortex that provides this area with exquisite feature-detection abilities not found in the classical primary visual cortex. —PRS

Science, this issue p. 64

CLIMATE CHANGE

Deep Pacific cooling

Earth's climate cooled considerably across the transition from the Medieval Warm Period to the Little Ice Age about 700 years ago. Theoretically, owing to how the ocean circulates, this cooling should be recorded in Pacific deep-ocean temperatures, where water that was on the surface then is found today. Gebbie and Huybers used an ocean circulation model and observations from both the end of the 19th century and the end of the 20th century to detect and quantify this trend. The ongoing deep Pacific is cooling, which revises Earth's overall heat budget since 1750 downward by 35%. —HJS

Science, this issue p. 70

EVOLUTION

DNA breakage and adaptation

Adaptation to new environments often occurs in similar ways across different colonization events. Stickleback fish represent a classic example of this, in which repeated colonizations of freshwater have resulted in the loss of pelvic hind fins. Previous work has shown that a pelvic enhancer gene is involved. Xie *et al*. now show that this gene lies within a region of the genome

that is prone to double-stranded DNA breakage owing to a high thymine-guanine content. This enhanced region of breakage could lead to enhanced mutation rates that facilitate repeated adaptations to new environments. —SNV

Science, this issue p. 81

CONSERVATION

The distinctive sound of a biodiverse forest

Assessing the state of biodiversity in a forest is a time-consuming task that typically requires detailed on-the-ground surveys. In a Perspective, Burivalova *et al*. explain that recordings of soundscapes can provide an easier route to this information. By recording soundscapes from a forest over time and comparing them to a regional baseline, scientists can determine whether a forest's ecosystem is healthy or not. If the soundscape of a forest spared from conversion becomes impoverished and altered, an on-the-ground survey would be warranted. This approach may be particularly useful for companies interested in sustainability certification or zero-deforestation commitments. —JFU

Science, this issue p. 28

TUBERCULOSIS

Faulty kinase purged by tuberculosis?

Rare mutations in genes involved in interferon-γ–dependent immunity underpin human genetic susceptibility to severe mycobacterial diseases, including primary tuberculosis. Boisson-Dupuis *et al*. investigated whether two common missense variants of the TYK2 Janus kinase that have impaired catalytic activity conferred an increased risk of tuberculosis. Individuals homozygous for the P1104A (proline to alanine substitution at residue 1104) variant of TYK2 are markedly predisposed to developing primary tuberculosis, defining a common monogenic etiology for the "white plague." The current frequency of the P1104A allele in European populations is significantly decreased compared with its frequency in ancient European DNA samples. These findings suggest that negative selection against the TYK2 P1104A allele by endemic tuberculosis in Europe may have contributed to a slow genetic purge of this susceptibility allele during recent millennia. —IW

Sci. Immunol. **3**, eaau8714 (2018).

CANCER

Altering membrane potential for cancer

Polymorphisms in the G protein–coupled receptor GPR35 are associated with increased risk for certain inflammatory diseases that can progress to cancer. Schneditz *et al*. found that GPR35 promoted the activity of Na^+- and K^+-dependent adenosine triphosphatase (Na^+,K^+-dependent ATPase), a transmembrane pump that sets the membrane potential in cells. This effect was enhanced by a disease-associated GPR35 variant. Stimulation of Na^+,K^+-ATPase activity by GPR35 increased glycolysis and proliferation in intestinal epithelial cells. Na^+,K^+-ATPase deficiency or treatment with a pepducin targeting GPR35 decreased tumor burden in mouse models of intestinal cancer. —WW

Sci. Signal. **12**, eaau9048 (2019).

REVIEW SUMMARY

OPTICS

Exceptional points in optics and photonics

Mohammad-Ali Miri and Andrea Alù*

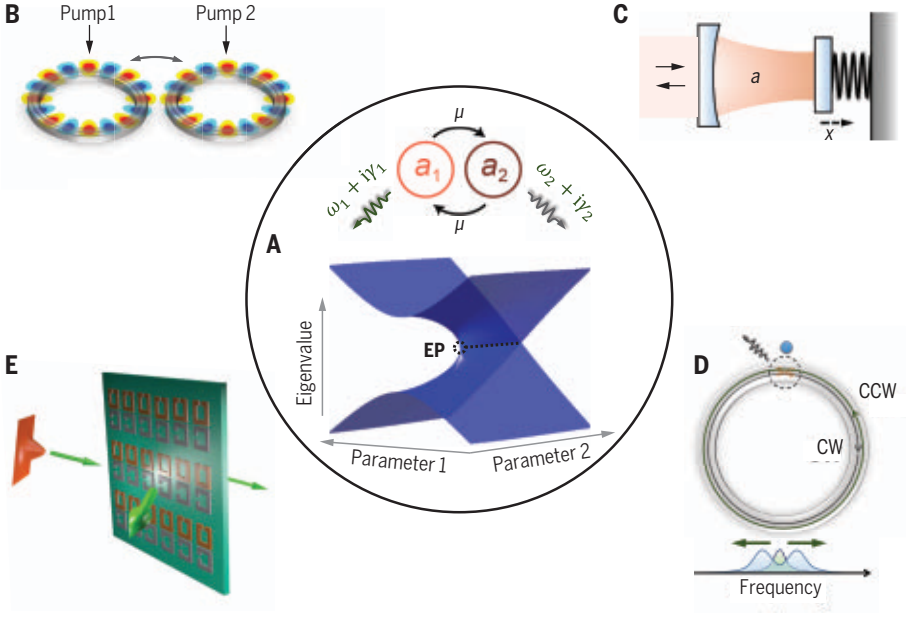
BACKGROUND: Singularities are critical points for which the behavior of a mathematical model governing a physical system is of a fundamentally different nature compared to the neighboring points. Exceptional points are spectral singularities in the parameter space of a system in which two or more eigenvalues, and their corresponding eigenvectors, simultaneously coalesce. Such degeneracies are peculiar features of nonconservative systems that exchange energy with their surrounding environment. In the past two decades, there has been a growing interest in investigating such nonconservative systems, particularly in connection with the quantum mechanics notions of parity-time symmetry, after the realization that some non-Hermitian Hamiltonians exhibit entirely real spectra. Lately, non-Hermitian systems have raised considerable attention in photonics, given that optical gain and loss can be integrated as nonconservative ingredients to create artificial materials and structures with altogether new optical properties.

ADVANCES: As we introduce gain and loss in a nanophotonic system, the emergence of exceptional point singularities dramatically alters the overall response, leading to a range of exotic functionalities associated with abrupt phase transitions in the eigenvalue spectrum. Even though such a peculiar effect has been known theoretically for several years, its controllable realization has not been made possible until recently and with advances in exploiting gain and loss in guided-wave photonic systems. As shown in a range of recent theoretical and experimental works, this property creates opportunities for ultrasensitive measurements and for manipulating the modal content of multimode lasers. In addition, adiabatic parametric evolution around exceptional points provides interesting schemes for topological energy transfer and designing mode and polarization converters in photonics. Lately, non-Hermitian degeneracies have also been exploited for the design of laser systems, new nonlinear optics phenomena, and exotic scattering features in open systems.

OUTLOOK: Thus far, non-Hermitian systems have been largely disregarded owing to the dominance of the Hermitian theories in most areas of physics. Recent advances in the theory of non-Hermitian systems in connection with exceptional point singularities has revolutionized our understanding of such complex systems. In the context of optics and photonics, in particular, this topic is highly important because of the ubiquity of nonconservative elements of gain and loss. In this regard, the theoretical developments in the field of non-Hermitian physics have allowed us to revisit some of the well-established platforms with a new angle of utilizing gain and loss as new degrees of freedom, in stark contrast with the traditional approach of avoiding these elements. On the experimental front, progress in fabrication technologies has allowed for harnessing gain and loss in chip-scale photonic systems. These theoretical and experimental developments have put forward new schemes for controlling the functionality of micro- and nanophotonic devices. This is mainly based on the anomalous parameter dependence in the response of non-Hermitian systems when operating around exceptional point singularities. Such effects can have important ramifications in controlling light in new nanophotonic device designs, which are fundamentally based on engineering the interplay of coupling and dissipation and amplification mechanisms in multimode systems. Potential applications of such designs reside in coupled-cavity laser sources with better coherence properties, coupled nonlinear resonators with engineered dispersion, compact polarization and spatial mode converters, and highly efficient reconfigurable diffraction surfaces. In addition, the notion of the exceptional point provides opportunities to take advantage of the inevitable dissipation in environments such as plasmonic and semiconductor materials, which play a key role in optoelectronics. Finally, emerging platforms such as optomechanical cavities provide opportunities to investigate exceptional points and their associated phenomena in multiphysics systems. ■

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aar7709



Ubiquity of non-Hermitian systems, supporting exceptional points, in photonics. (**A**) A generic non-Hermitian optical system involving two coupled modes with different detuning, $\pm\omega_{1,2}$, and gain-loss values, $\pm\gamma_{1,2}$, coupled at rate of μ. The real part of the associated eigenvalues in a two-dimensional parameter space of the system, revealing the emergence of an exceptional point (EP) singularity. a_1 and a_2 are the modal amplitudes. (**B** to **E**) A range of different photonic systems, which are all governed by the coupled-mode equations. (B) Two coupled lasers pumped at different rates. (C) Dynamical interaction between optical and mechanical degrees of freedom in an optomechanical cavity. (D) A resonator with counter-rotating whispering gallery modes. CW, clockwise; CCW, counterclockwise. (E) A thin metasurface composed of coupled nanoantennas as building blocks.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: aalu@gc.cuny.edu
Cite this article as M.-A. Miri and A. Alù, *Science* **363**, eaar7709 (2019). DOI: 10.1126/science.aar7709

REVIEW

OPTICS

Exceptional points in optics and photonics

Mohammad-Ali Miri[1,2,3] and Andrea Alù[4,3,5,1]*

Exceptional points are branch point singularities in the parameter space of a system at which two or more eigenvalues, and their corresponding eigenvectors, coalesce and become degenerate. Such peculiar degeneracies are distinct features of non-Hermitian systems, which do not obey conservation laws because they exchange energy with the surrounding environment. Non-Hermiticity has been of great interest in recent years, particularly in connection with the quantum mechanical notion of parity-time symmetry, after the realization that Hamiltonians satisfying this special symmetry can exhibit entirely real spectra. These concepts have become of particular interest in photonics because optical gain and loss can be integrated and controlled with high resolution in nanoscale structures, realizing an ideal playground for non-Hermitian physics, parity-time symmetry, and exceptional points. As we control dissipation and amplification in a nanophotonic system, the emergence of exceptional point singularities dramatically alters their overall response, leading to a range of exotic optical functionalities associated with abrupt phase transitions in the eigenvalue spectrum. These concepts enable ultrasensitive measurements, superior manipulation of the modal content of multimode lasers, and adiabatic control of topological energy transfer for mode and polarization conversion. Non-Hermitian degeneracies have also been exploited in exotic laser systems, new nonlinear optics schemes, and exotic scattering features in open systems. Here we review the opportunities offered by exceptional point physics in photonics, discuss recent developments in theoretical and experimental research based on photonic exceptional points, and examine future opportunities in this area from basic science to applied technology.

Hermiticity is a property of a wide variety of physical systems, under the assumptions of being conservative and obeying time-reversal symmetry. Hermitian operators play a key role in the theory of linear algebraic and differential operators (*1*–*4*), and they are known to exhibit real-valued eigenvalues, a property that stems from energy conservation. For a set of dynamical equations described through a Hermitian operator, the relation between initial and final states is governed by a unitary operation. Hermiticity has long been considered one of the pillars of mathematical and physical models, such as in quantum mechanics and electromagnetics. The elegance of such theories lies in powerful properties, including the completeness and orthogonality of the eigenbasis of the governing operators (*1*). However, these models are based on idealizations, like the assumption of complete isolation of a system from its surrounding environment. In principle, nonconservative elements arise ubiquitously in various forms; thus, a proper description of a realistic physical system requires a non-Hermitian Hamiltonian. Generally, nonconservative phenomena are introduced as small perturbations to otherwise Hermitian systems. Thus, the overall behavior of non-Hermitian systems has been largely extracted from their Hermitian counterparts. However, recent investigations have revealed that non-Hermitian phenomena can drastically alter the behavior of a system compared to its Hermitian counterpart. The best example of such deviation is the emergence of singularities, so-called exceptional points, at which two or more eigenvalues, and their associated eigenvectors, simultaneously coalesce and become degenerate (*5*).

The term "exceptional point" was first introduced in studying the perturbation of linear non-Hermitian operators (*6*), described by a general class of matrices $H(z)$ parameterized by the complex variable $z = x + iy$, where x is the real part, i is the imaginary unit, and y is the imaginary part. The eigenvalues $\sigma_n(z)$ and eigenvectors $|\psi_n(z)\rangle$ of H can be represented as analytic functions except at certain singularities $z = z_{EP}$ (EP, exceptional point). At such exceptional points, two eigenvalues coalesce, and the matrix H can no longer be diagonalized. The physical importance of exceptional points was pointed out in early works (*7*, *8*), in which the terminology of non-Hermitian degeneracy was used to distinguish such critical points from regular degeneracies occurring in Hermitian systems (*9*, *10*). In addition, exceptional points were referred to as branch-point singularities in investigating the quantum theory of resonances in the context of atomic, molecular, and nuclear reactions (*11*). Early experiments on microwave cavities revealed the peculiar topology of eigenvalue surfaces near exceptional points (*12*, *13*). The emergence of spectral singularities was also pointed out in the analysis of multimode laser cavities (*14*, *15*) and in time-modulated complex light potentials for matter waves (*16*).

Recently, interest in these peculiar spectral degeneracies has been sparked in a particular family of non-Hermitian Hamiltonians, the so-called parity-time (PT) symmetric systems. A Hamiltonian is PT symmetric as long as it commutes with the $\mathcal{PT}$ operator, that is, $[\mathcal{H}, \mathcal{PT}] = 0$, where the parity operator $\mathcal{P}$ represents a reflection with respect to a center of symmetry and the time operator $\mathcal{T}$ represents complex conjugation. It has been realized that PT-symmetric Hamiltonians, despite being non-Hermitian, can support entirely real eigenvalue spectra (*17*). More interestingly, it has been realized that commuting with the $\mathcal{PT}$ operator is not sufficient to ensure a real spectrum, as formally PT-symmetric Hamiltonians can undergo a phase transition to the spontaneously broken symmetry regime, in which complex eigenvalues appear. The phase transition happens as a result of a parametric variation in the Hamiltonian. Quite interestingly, the symmetry-breaking threshold point exhibits all properties of an exceptional point singularity (*17*–*23*).

Although these theoretical explorations originated in the realm of quantum mechanics, optics and photonics have proven to be the ideal platform to experimentally observe and utilize the rich physics of exceptional points (*24*–*27*). Owing to the abundance of nonconservative processes, photonics provides the necessary ingredients to realize controllable non-Hermitian Hamiltonians. Indeed, dissipation is ubiquitous in optics, because it arises from material absorption as well as radiation leakage to the outside environment. In addition, gain can be implemented in a locally controlled fashion through stimulated emission, which involves optical or electrical pumping of energy through an external source, or through parametric processes. Therefore, photonics provides a fertile ground to systematically investigate non-Hermitian Hamiltonians and exceptional points. Recent theoretical developments in the area of non-Hermitian physics have opened exciting opportunities to revisit fundamental concepts in nonconservative photonic systems with gain and loss, such as lasers, sensors, absorbers, and isolators. In these systems, exceptional points open pathways for totally new functionalities and performance. The interested reader may find detailed overviews of non-Hermitian and, in particular, PT-symmetric systems in the context of optics and photonics in recent review papers (*28*–*32*). In the present work, we discuss instead more broadly the concept of exceptional points in non-Hermitian systems. In the following, we provide an introduction to exceptional point physics and explain some of the fundamental concepts associated with such critical points. We then draw the connection with optics and photonics and show the universal occurrence of

[1]Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712, USA. [2]Department of Physics, Queens College of the City University of New York, Queens, NY 11367, USA. [3]Physics Program, Graduate Center of the City University of New York, New York, NY 10016, USA. [4]Photonics Initiative, Advanced Science Research Center, City University of New York, New York, NY 10031, USA. [5]Department of Electrical Engineering, City College of The City University of New York, New York, NY 10031, USA.
*Corresponding author. Email: aalu@gc.cuny.edu

exceptional points in optical settings. Finally, we review recent theoretical and experimental efforts in observing exceptional points in optics and their peculiar functionalities in practical devices, presenting an outlook for the future of this exciting area of research.

Theoretical background

We begin by investigating exceptional points in a generic two-level system. Assuming that $a_{1,2}$ are the modal amplitudes of two states that evolve with the variable ξ, representing the evolution time or propagation distance, the coupled mode equations can be generally written as

$$\frac{\mathrm{d}}{\mathrm{d}\xi}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = -\mathrm{i}\begin{pmatrix} \omega_1 - \mathrm{i}\gamma_1 & \mu \\ \mu & \omega_2 - \mathrm{i}\gamma_2 \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad (1)$$

where ω is the resonance frequency of the two coupled modes, μ is the coupling coefficient, and γ is their decay rate. This particular choice of Hamiltonian system, shown in Fig. 1A, represents a large class of structures and devices of large relevance in photonics, examples of which are given in Fig. 1, such as coupled cavities (Fig. 1B) (*33*), coupled waveguides (Fig. 1C) (*34*), polarization states in the presence of small perturbations in an optical waveguide (Fig. 1D) (*35*), counter-propagating waves in Bragg gratings (Fig. 1E) (*36*), wave mixing in nonlinear crystals (Fig. 1F) (*37*), coupled optical and mechanical modes in an optomechanical cavity (Fig. 1G) (*38*), and a two-level atom in a cavity (Fig. 1H) (*39*). In the case of coupled optical resonators, for instance, $\omega_{1,2}$ in Eq. 1 represent the individual frequencies of each element, $\gamma_{1,2}$ describe their loss or gain rate, and μ represents the mutual coupling. Assuming, harmonic solutions of the form $(a_1, a_2) = (a_1, a_2)e^{-i\sigma\xi}$, the eigenvalues of the system are

$$\sigma_{\pm} = \omega_{\mathrm{ave}} - \mathrm{i}\gamma_{\mathrm{ave}} \pm \sqrt{\mu^2 + (\omega_{\mathrm{diff}} + \mathrm{i}\gamma_{\mathrm{diff}})^2} \quad (2)$$

where $\omega_{\mathrm{ave}} = (\omega_1 + \omega_2)/2$ and $\gamma_{\mathrm{ave}} = (\gamma_1 + \gamma_2)/2$, respectively, represent the mean values of resonance frequencies and loss factors, whereas $\omega_{\mathrm{diff}} = (\omega_1 - \omega_2)/2$ and $\gamma_{\mathrm{diff}} = (\gamma_1 + \gamma_2)/2$ are the differences between their resonance frequencies and loss factors.

The Hamiltonian in Eq. 1 is a function of multiple parameters. In Fig. 2, A and B, we evaluate the evolution of real and imaginary parts of the eigenvalues in the parameter space (ω_{diff}, γ_{diff}), assuming a constant coupling coefficient μ. An exceptional point occurs when the square-root term in Eq. 2 is zero, as the two eigenvalues coalesce. Assuming a real coupling constant, this happens for ($\omega_{\mathrm{diff}} = 0$; $\gamma_{\mathrm{diff}} = \pm\mu$). Figure 2, A and B, highlights the interesting topology of the branch point singularity at the exceptional point, which has important implications in the optical response of the system around this parameter point, as we discuss in the following sections.

The two-body problem investigated here is the simplest case of a non-Hermitian system. In general, exceptional points appear ubiquitously in systems with spatially discrete or continuous degrees of freedom of multiple dimensionalities. In principle, when more than two eigenvalue surfaces are involved, it is also possible that more than two surfaces simultaneously collapse at one point, creating a higher-order exceptional point (*40*, *41*). A third-order exceptional point, for example, is formed when three eigenvalues simultaneously coalesce. In this scenario, the square-root dependence of the eigenvalues around the exceptional point in Eq. 2 is replaced by a cubic root. It is worth stressing that at an exceptional point, the coalescing eigenvalues do not support independent eigenvectors, implying that, in discrete systems described by a matrix Hamiltonian, the Jordan form is no longer diagonal (*42*). This is notably different from accidental degeneracies, which occur when two eigenvalues with different eigenvectors cross. In a two-dimensional parameter space, such accidental degeneracies appear when two eigenvalue surfaces form a double cone or "diablo," forming diabolic points (*43*). In contrast with exceptional points, at the diabolic points, the eigenvectors remain linearly independent. Diabolic points emerge in various Hermitian systems, most notably in molecular reactions (*44*) and in the electronic band diagram of graphene (*45*).

Exceptional point singularities are closely related to the phenomenon of level repulsion, which has been originally explored in the context of quantum chaos, because it explains the scarcity of closely spaced levels in Wigner distributions (*46*). In photonics, level repulsion is of great interest because it marks strong coupling and hybridization between states, which is manifested as a repulsion between closely spaced eigenvalues when a parameter is adiabatically

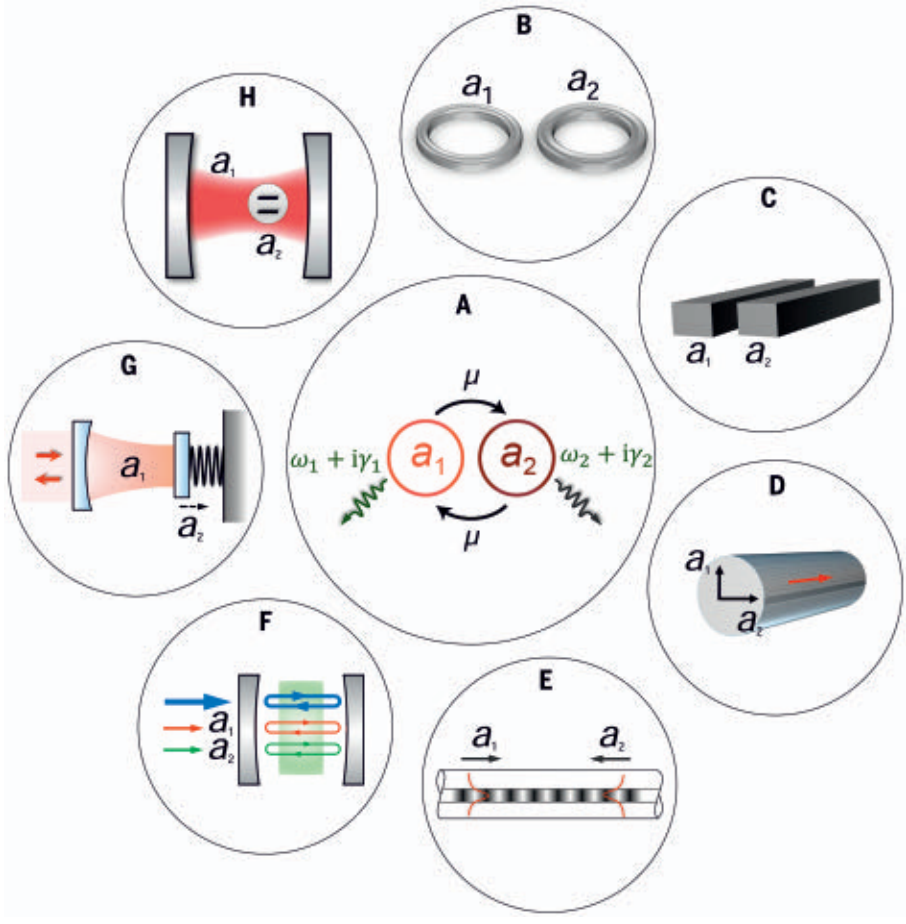


Fig. 1. A generic two-level system and its different realizations in optics and photonics. (**A**) A schematic representation of a generic two-level system composed of two coupled entities. (**B**) Two coupled optical cavities with spatially separated resonator modes. (**C**) Two evanescently coupled optical waveguides with spatially separated waveguide modes. (**D**) Coupled orthogonal polarization states in an optical waveguide. (**E**) Counter-propagating waves in a volume Bragg grating. (**F**) Signal and idler frequency components in a parametric amplifier. (**G**) Photonic and phononic degrees of freedom in an optomechanical cavity. (**H**) Coupling between a two-level atom and an optical cavity mode. The different platforms represented in (B) to (H) can be treated under a unified model depicted schematically in (A). The universality of nonconservative processes in these settings calls for a systematic understanding of non-Hermiticity in a basic two-level system as a first step toward a rigorous bottom-up approach for designing complex photonic systems in the presence of gain and loss. The arrows indicate electromagnetic waves, and different colors indicate different frequencies.

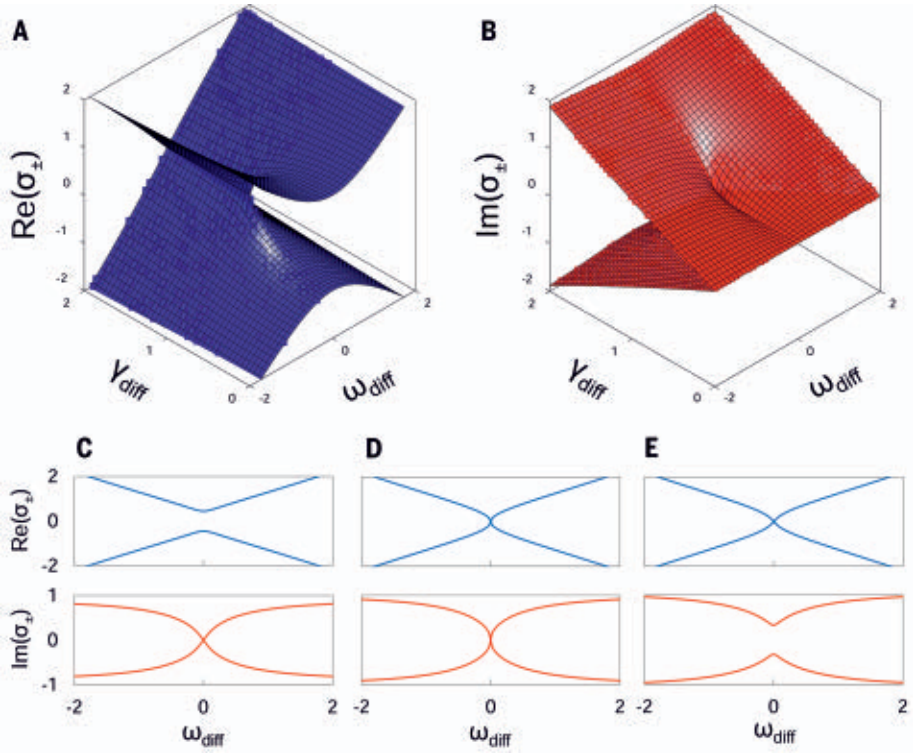


Fig. 2. Exceptional points in a non-Hermitian two-level system. (**A** and **B**) Evolution of the real (A) and imaginary (B) parts of the eigenvalues of the system described by Eq. 1 in the two-dimensional parameter space (ω_{diff}, γ_{diff}). These panels illustrate the exotic topology of the eigenvalue surfaces near an exceptional point singularity. (**C** to **E**) Eigenvalues versus ω_{diff} for different values of γ_{diff}, that is, cross sections of the surfaces depicted in (A) and (B). Owing to the presence of the exceptional point ($\gamma_{diff} = \gamma_{EP}$; $\omega_{diff} = \omega_{EP}$), depending on the value of the secondary parameter, different parameter dependence is observed for the eigenvalues. (C) For $\gamma_{diff} > \gamma_{EP}$, level repulsion occurs in the real part of the eigenvalues, whereas the imaginary parts cross. (D) For $\gamma_{diff} = \gamma_{EP}$, the real and imaginary parts coalesce at $\omega_{diff} = \omega_{EP}$. (E) For $\gamma_{diff} < \gamma_{EP}$, level crossing governs the real parts of the eigenvalues, whereas the imaginary parts repel each other.

tuned (*47*). They typically occur near an exceptional point in the real or complex parameter space. For instance, Fig. 2, C to E, shows cross sections of the eigenvalue surfaces in Fig. 2, A and B, for different values of γ_{diff}, highlighting level repulsion in either their real (Fig. 2C) or imaginary part (Fig. 2E) for values of γ_{diff} respectively larger or smaller than the critical value $\gamma_{diff} = \gamma_{EP}$, corresponding to the exceptional point condition (Fig. 2D). Level repulsion in the real (imaginary) part is accompanied by level crossing of the imaginary (real) part, as shown in Fig. 2, C to E (*48*, *49*). At the critical condition $\gamma_{diff} = \gamma_{EP}$, both real and imaginary parts of the eigenvalues coalesce, and an exceptional point is achieved. The different behavior in the three cases is determined by the topology of the involved Riemann surfaces at the given cross section. As a special case, level repulsion can arise also in Hermitian systems, such as in the case of two lossless optical resonators, in which level repulsion occurs as we detune their resonance frequency (*33*). Consistent with Fig. 2C, this phenomenon is associated with an exceptional point in the complex parameter space, as we operate at $\gamma_{diff} = 0 < \gamma_{EP}$.

In the context of exceptional points, an especially relevant class of non-Hermitian two-level systems are those satisfying PT symmetry. In the context of quantum mechanics, a Hamiltonian $\mathcal{H}$ is PT symmetric when $[\mathcal{H}, \mathcal{PT}] = 0$, where $\mathcal{P}$ and $\mathcal{T}$ respectively represent parity and time operators. In photonics, this corresponds to the case in which loss in one region is balanced by gain in another symmetric region (*50*). For the two-level system of Eq. 1, considering that the parity and time operators respectively act as $\mathcal{P}(a,b) = (b,a)$ and $\mathcal{T}(a,b) = (a^*, b^*)$, where a and b are two variables, the conditions of PT symmetry are satisfied for $\omega_1 = \omega_2 \equiv \omega$ and $\gamma_1 = -\gamma_2 \equiv \gamma$. The response of this system is governed by the interplay of two major processes: the gain and loss contrast γ and the mutual coupling μ. An exceptional point arises at the critical condition $\mu = \gamma$. Here, the exceptional point marks the onset of a transition from purely real eigenvalues, associated with oscillatory solutions $\exp(\pm i|\sigma_\pm|\xi)$, where ξ is the evolution variable, to purely imaginary eigenvalues associated with growing or decaying solutions $\exp(\pm|\sigma_\pm|\xi)$. This transition is often referred to as spontaneous symmetry breaking, because the eigenvalues change their behavior despite the fact that the governing evolution operator preserves its symmetry. The behavior of the eigenvalues of a PT-symmetric system is shown in Fig. 3A, highlighting the bifurcation associated with the spontaneous symmetry breakdown at the exceptional point.

In Eq. 1, we assumed that the coupling μ is a real parameter, whereas in principle, it can become complex, involving dissipation. For instance, in several scenarios, coupling between two states is mediated through a continuum of radiation modes, for which the energy partially leaks to the outside environment (*51*). Examples include radiative coupling between subwavelength nanoparticles (*52*) as well as channel-mediated coupling of microring lasers (*53*). Independent of the coupling mechanism, exceptional points also arise in this case. According to Eq. 2, assuming a purely imaginary coupling $\mu = i\mu_i$, exceptional points emerge for ($\omega_{diff} = \pm\mu_i$; $\gamma_{diff} = 0$). In this case, the exceptional point arises for a frequency detuning equal to the mutual coupling between cavities.

The discussion on exceptional points presented so far has been built on Hamiltonian systems, or, in broader terms, on dynamical systems, that evolve in time and space through a linear operator. A large body of photonic systems, however, are open, coupled to a continuum of radiation modes, as in the case of optical waveguides coupled to cavities or finite-sized scatterers illuminated by impinging optical fields. Such systems are better described through a scattering matrix, which directly relates outgoing waves and incoming waves. The scattering matrix can be compared with the time-evolution operator, that is, $\mathcal{U} = \exp(-i\mathcal{H}\xi)$ in Hamiltonian systems. Indeed, in a scattering medium without material gain or loss, the scattering matrix is unitary, with all its eigenvalues located on the unit circle (*54*). In the presence of loss and/or gain, however, the norms are not preserved, and the eigenvalues can, in general, be located inside or outside the unit circle. Quite interestingly, similar to Hamiltonian systems, exceptional points can also emerge in the scattering matrix formalism when two or more eigenvalues and their associated eigenvectors coalesce (*55*). A basic example is a PT-symmetric Fabry-Perot resonator involving two materials with balanced gain and loss (Fig. 3B). At a given frequency, for an increasing gain and loss contrast, the scattering-matrix eigenvalues bifurcate from the unit circle at an exceptional point singularity, as shown in Fig. 3B. Here, the exceptional point marks the onset of the broken symmetry regime, in which amplification of the wave excitation becomes the dominant response of the PT-symmetric scatterer.

Exceptional points in photonics

Exceptional points arise in several optical and photonic systems. In the previous section, we introduced a general class of two-level systems described through coupled-mode equations, pointing out the conditions to achieve a second-order exceptional point. Integrated photonic waveguides and cavities, in particular, provide a controllable platform to observe exceptional points. In

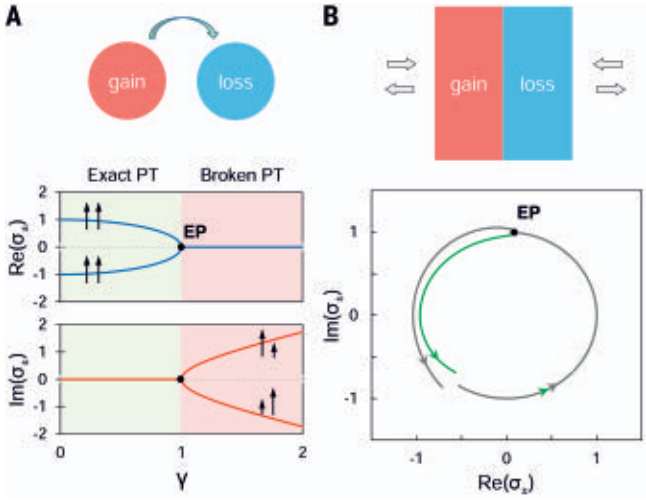


Fig. 3. PT symmetry in closed and open systems. PT-symmetric systems form an interesting class of non-Hermitian settings, which share certain similarities with Hermitian systems. In the case of a two-level system (Fig. 1), PT symmetry is realized for $\omega_1 = \omega_2 \equiv \omega$ and $\gamma_1 = -\gamma_2 \equiv \gamma$, that is, when the individual levels share the same real part but exhibit opposite values of the imaginary parts (gain and loss). (**A**) A PT-symmetric system of two coupled waveguides (top) with gain (red) and loss (blue), and the corresponding eigenvalues (bottom) versus the gain-loss contrast γ. This figure reveals a transition in the eigenvalues from purely real (exact PT symmetry) to purely imaginary (broken PT symmetry). Interestingly, the PT symmetry–breaking threshold point reveals all the properties of an exceptional point singularity. In this figure, the arrows represent the intensity of the eigenmodes in both the exact and broken PT regimes. (**B**) A PT-symmetric Fabry-Perot resonator (top) and the eigenvalues of its scattering matrix (bottom) evolving as a function of the frequency of excitation. In this case, an exceptional point marks a transition in the eigenvalue evolution, breaking away from the unit circle. The geometries of (A) and (B) represent examples of Hamiltonian and scattering settings.

integrated photonic platforms, exceptional points and phase transitions have been observed in coupled passive optical waveguides, where controllable loss in one of the channels was utilized (*56*) (Fig. 4, A and B). In the context of PT symmetry, spontaneous symmetry breaking at the exceptional point was demonstrated in a coupled arrangement of optical waveguides with balanced gain and loss (*50*). In other works, coupled optical cavities with gain and loss were utilized to observe a PT-symmetric phase transition (*57*, *58*) (Fig. 4, D and E). The first demonstration of exceptional points in periodic structures was achieved in time-domain lattices (*59*) (Fig. 4C), induced through the propagation of short laser pulses in two coupled fiber loops of a slightly different lengths with alternating gain and loss. This propagation creates a quantum walk of pulses governed by PT-symmetric evolution equations, described through a peculiar band structure as in spatially periodic structures. In addition, exceptional points have been demonstrated in photonic crystal slabs (*60*), in which out-of-plane radiation losses due to the finite thickness of the dielectric slab result in the merging of two eigenfrequency bands, inducing a ring of exceptional points in the wave number space. Among other realizations, exceptional points have also been experimentally demonstrated in chaotic optical cavities (*61*). In all these photonic systems, operation around the exceptional points enables a singular optical response.

The peculiar properties of exceptional points have also been investigated in open scattering systems involving gain and loss. In particular, it has been shown that a PT-symmetric Fabry-Perot cavity, similar to the one discussed in Fig. 3B, can simultaneously act as a laser and a coherent perfect absorber at the exceptional point (*55*, *62*). This interesting behavior, occurring as a result of the coalescence of a pair of poles and zeroes of the scattering matrix eigenvalue, has been recently demonstrated in an integrated semiconductor resonator with active and passive regions (*63*). Non-Hermitian optical gratings with alternating layers of materials with different levels of loss or gain reveal another interesting aspect of exceptional points (*64*, *65*). In such systems, whereas reciprocity enforces equal transmission in both directions, the reflection coefficients can be completely different. In a Hermitian system, equal transmission coefficients also require equal magnitude of the reflection coefficients, but in non-Hermitian systems, this is not the case. The contrast in reflection amplitudes is maximized at the exceptional point, where the reflection from one direction becomes zero and the reflection from the other direction can be very large, thus inducing unidirectional invisibility (*65*). In a similar fashion, it has been shown that a two-layer structure with gain and loss can exhibit one-way reflectionless behavior at a particular frequency, thus inducing an anisotropic transmission resonance (*66*). At the exceptional point, the photonic bandgap closes, whereas the coupling between counter-propagating waves becomes unidirectional (*67*). Unidirectional invisibility has been observed in different settings, including in integrated semiconductor waveguide gratings (*68*), organic composite films (*69*), time-domain lattices (*59*), and coupled acoustic resonators (*70*). Similar ideas have been utilized in microring resonators to create integrated laser devices supporting modes with definite angular momentum when the system is biased at an exceptional point (*71*). In addition, it has been shown that properly engineered defects in microring resonators can create an exceptional point that instead induces chirality between counter-rotating modes (*72*–*74*). It has also been shown that non-Hermitian scattering systems operating around the exceptional points can induce other interesting phenomena, such as negative refraction (*75*) and unidirectional cloaking (*76*, *77*).

Coherently prepared, multilevel warm atomic vapors provide another controllable platform to realize complex optical potentials. In such systems, strong pump laser beams can create waveguiding effects for weak probe beams where, under proper detuning, both gain and loss can be achieved in Raman-active systems (*78*). In this regard, the realization of complex potentials supporting exceptional points have been theoretically proposed in three- and four-level atoms (*79*, *80*) and experimentally demonstrated in coupled atomic vapor cells (*81*), as well as in PT-symmetric optical lattices (*82*).

Even though the discussion here is primarily focused on linear operators, it is important to also stress the relevance of exceptional points in nonlinear systems. The connection of non-Hermiticity to nonlinear systems is multifold: First, most nonlinear configurations in optics and photonics are accompanied by losses, and second, active devices are, by nature, nonlinear. Therefore, lasers, amplifiers, and saturable absorbers are all examples of devices in which nonlinearity and non-Hermiticity coexist. In addition, nonlinear optical effects can create interactions between different wave components. A high-intensity pump, for example, initiates energy exchange between lower-intensity wave components that are governed by a linearized operator. Such an operator is, by essence, non-Hermitian, given the energy exchange between pump and probe through the nonlinearity.

The interplay of nonconservative and nonlinear effects is of special interest, given that optical materials with strong nonlinearities necessarily suffer from large absorption (*83*). Therefore, concepts from non-Hermitian physics are sought to provide strategies to take advantage of losses in such nonlinear materials. In this regard, the conjunctive use of nonlinear processes with gain and loss have been suggested as a viable route to achieve optical nonreciprocity (*84*, *85*). In addition, it has been shown that laser systems exhibit exotic behavior such as anomalous pump dependence near the exceptional point singularity (*86*, *87*), as well as reduced lasing threshold with increased losses

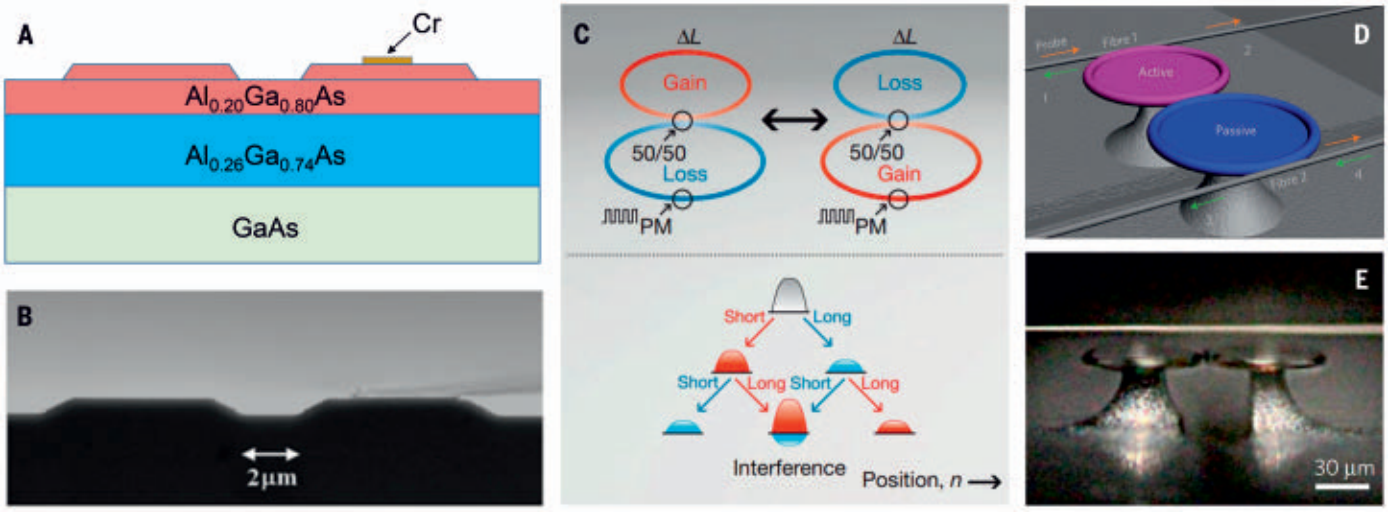


Fig. 4. Experimental demonstration of exceptional points in various optical settings. (**A** and **B**) Coupled integrated photonic waveguides (A) fabricated through a multilayer $Al_xGa_{1-x}As$ heterostructure (B), for which thin layers of chromium of different widths were utilized to impart different amount of losses in one of the waveguides (*56*). In this setting, couplers with different losses on one arm were used to observe mode symmetry breaking beyond the critical loss contrast associated with the exceptional point. (**C**) The propagation of laser pulses in coupled fiber loops of slightly different lengths (ΔL) with alternating gain and loss creates a quantum walk of pulses which is governed by a PT-symmetric operator (*59*). In this temporal lattice, the onset of complex eigenvalues associated with the band merging effect at the exceptional point was experimentally demonstrated. PM represents a phase modulator that creates an effective potential for the light pulses. (**D** and **E**) Coupled microring resonators with gain and loss have been used to probe the exceptional point through the mode splitting of the resonance eigenmodes (*57*, *58*). In (D), the numbers indicate the four ports that are used to probe the system, and orange and green arrows represent waves propagating in forward and backward directions, respectively. [Credits: (A) and (B) reprinted with permission from (*56*), copyright 2009 by the American Physical Society; (C), (D), and (E) reprinted from (*59*), (*57*), and (*59*), respectively, with permission from Springer Nature]

(*88*). The impact of non-Hermiticity on nonlinear waves in bulk and periodic systems has been also explored, after the realization that PT-symmetric potentials support optical solitons (*89*). Indeed, although dissipative nonlinear systems have been largely investigated (*90*), recent developments in the area of PT symmetry have sparked interest in the exploration of new integrated nonlinear systems combining gain and loss (*29*, *91*, *92*). In addition, solitary waves in PT-symmetric potentials have been experimentally demonstrated in time-domain lattices (*93*). Nonlinear wave-mixing processes, such as sum and difference frequency generation and optical parametric amplification, are other examples of non-Hermitian systems in which external coupling through a pump beam mediates the interactions (*94*).

At this point, it is worth stressing that exceptional points are not necessarily difficult to find in optical setups because they occur ubiquitously in the wave number space, even in conservative systems in which no gain or loss is involved. In these scenarios, a part of a Hermitian system can be considered non-Hermitian, because it exchanges energy with the rest of the system. Possibly the best-known example of these trivial exceptional points is the total internal reflection at the interface of two materials. In this case, light transmitted at the interface of two media critically depends on the incidence angle of the impinging light. In particular, at a critical angle, a phase transition occurs in the propagation wave number of the second medium, which goes from being real to complex valued. Other well-known examples of exceptional points in the wave number space are the cut-off frequency of a closed waveguide or the edge of a photonic bandgap in periodic structures. In addition, a volume Bragg grating, in which alternating layers of two different materials with refractive indices n_1 and n_2 create a photonic bandgap for a range of incoming frequencies, supports an exceptional point. In this structure, the wave number of the counter-propagating waves follows a square-root dispersion in terms of the incoming wave frequency. Whereas in the propagation band the wavenumber is real, inside the bandgap it becomes complex, and an exceptional point marks this transition. Similar to the exceptional points emerging in complex potentials, the photonic bandgap in gratings exhibits interesting properties, such as a vanishing group velocity (*95*).

Applications in nanophotonics

The exotic properties of exceptional points open interesting possibilities for advanced light manipulation. In this section, we present an overview of some of the recent theoretical and experimental developments in the exploration of exceptional points for applications in photonics. As in other areas of physics, in photonics, perturbation theory is an important mathematical tool to tackle a range of problems without having to deal with complex full-wave equations. Owing to the singularity at exceptional points, as well as the dimensionality collapse in the eigenvector space, standard perturbation theory, however, does not apply at such points. The perturbation problem can be introduced as $\mathcal{H} = \mathcal{H}_0 + \varepsilon\mathcal{H}_1$ where we want to find the behavior of the eigenvalues $\sigma_n(\varepsilon)$ and eigenvectors $|\psi_n(\varepsilon)\rangle$ of $\mathcal{H}$ for $\varepsilon \ll 1$, where ε is the perturbation parameter. In general, such a perturbation problem can be divided into regular and singular problems (*96*). In the regular case, a power-series solution with integral powers of ε exists, that is, $\sigma(\varepsilon) = \sigma_0 + \sum_{n=1}^{\infty} c_n\varepsilon^n$, where c_n are the series coefficients, with a finite radius of convergence. However, in the case of an exceptional point singularity, such a solution does not converge. At a singularity, the exact solution at $\varepsilon = 0$ is of a fundamentally different nature compared with its neighboring points $\varepsilon \to 0$ (*96*). At a second-order exceptional point, the series solution

$$\sigma_{\pm}(\varepsilon) = \sigma_0 + \sum_{n=1}^{\infty} (\pm 1)^n c_n \varepsilon^{n/2} \qquad (3)$$

exists, where σ_0 is the eigenvalue at the exceptional point. The radius of convergence of this series in the complex ε plane is determined by the nearest exceptional point. In a similar manner, for a kth-order exceptional point the nth term in the perturbation series is $\varepsilon^{n/k}$, with a dominant first-order term of $\varepsilon^{1/k}$. For small perturbations, this term is considerably larger than the linear term ε, which occurs at regular points, enabling extra sensitivity to the parameter ε of a system when biased at the

exceptional point singularity. This property has been proposed to achieve enhanced mode splitting between counter-propagating whispering gallery modes of a microring resonator in the presence of nanoparticles (*97*). The prospect of utilizing exceptional points for enhanced mode splitting has been experimentally demonstrated in microtoroid cavities (*98*, *99*) (Fig. 5A). In addition, integrated microring resonators with externally controllable perturbations have been utilized to induce second- and third-order exceptional points, where ½ and ⅓ power-law exponents in mode splitting have been demonstrated (*100*) (Fig. 5B). Although it has been pointed out that enhanced sensitivity at the exceptional point does not necessarily correspond to enhanced precision in sensing instruments (*101*) and that quantum noise should be considered to assess the ultimate performance of these exceptional point sensors (*102*), sensors appear to be an interesting application area for these concepts. In this area, it has also been shown that a scaled form of PT symmetry can be used for enhanced sensor telemetry (*103*).

Another interesting application of exceptional points is mode discrimination in multimode laser cavities (*104*). A common issue in laser systems is that often several transverse or longitudinal modes may simultaneously lase. In this regard, it has been suggested to complement the active multimode laser cavity with a passive cavity that ideally exhibits an equal amount of loss. In this scenario, the overall level of loss is increased in the entire system, given that each mode overlaps with the loss region, and thus the gain threshold is expected to increase. However, a large discrimination between lasing thresholds of different modes is obtained at the exceptional point supported by this PT-symmetric system. In this case, the modes are split into two classes that are equally distributed between the active and passive regions, as well as modes that are localized either in the gain or loss cavity. The first class of modes remains neutral, whereas the modes located in the gain enter the gain regime. As a result, the passive cavity prevents some of the modes from lasing. More interestingly, this structure creates a large discrimination between the lasing thresholds of the fundamental mode with its closest competing counterpart. Assuming g_0 and g_1 to be the gain coefficients for fundamental and competing modes, respectively, in the coupled-cavity system, the discrimination is governed by $\sqrt{g_0^2 - g_1^2}$, which can be considerably larger than $g_0 - g_1$ in a single laser cavity. This approach has been utilized to enforce single longitudinal-mode operation in coupled microring lasers (*105*) (Fig. 6A) and in single rings with embedded active-passive gratings (*106*) (Fig. 6B). Similar strategies have been utilized to filter out transverse modes in ring resonators with large cross sections (*107*), in optically and electrically pumped stripe lasers (*108*, *109*) (Fig. 6C), and in microdisc

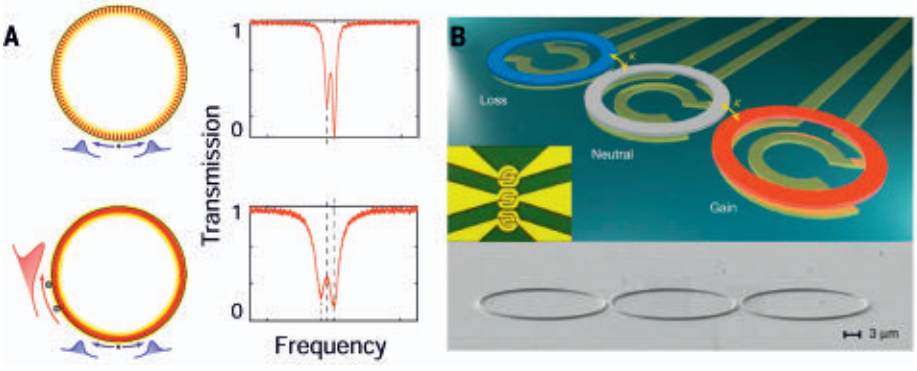


Fig. 5. Demonstration of enhanced perturbation near an exceptional point singularity. (**A**) Sensing a nanoparticle with a microtoroid resonator biased at an exceptional point (*99*). Blue arrows and curve indicate light pulses propagating in counter-rotating whispering gallery modes, and the red arrow and curve indicate a backscattering pulse due to the presence of additional scatterers (shown with two gray circles), which help to bias the system at an exceptional point. (**B**) Three coupled microring resonators creating a third-order exceptional point (*100*).κ represents the strength of coupling between adjacent microrings. [Credits: (A) and (B) reprinted from (*99*) and (*100*), respectively, with permission from Springer Nature]

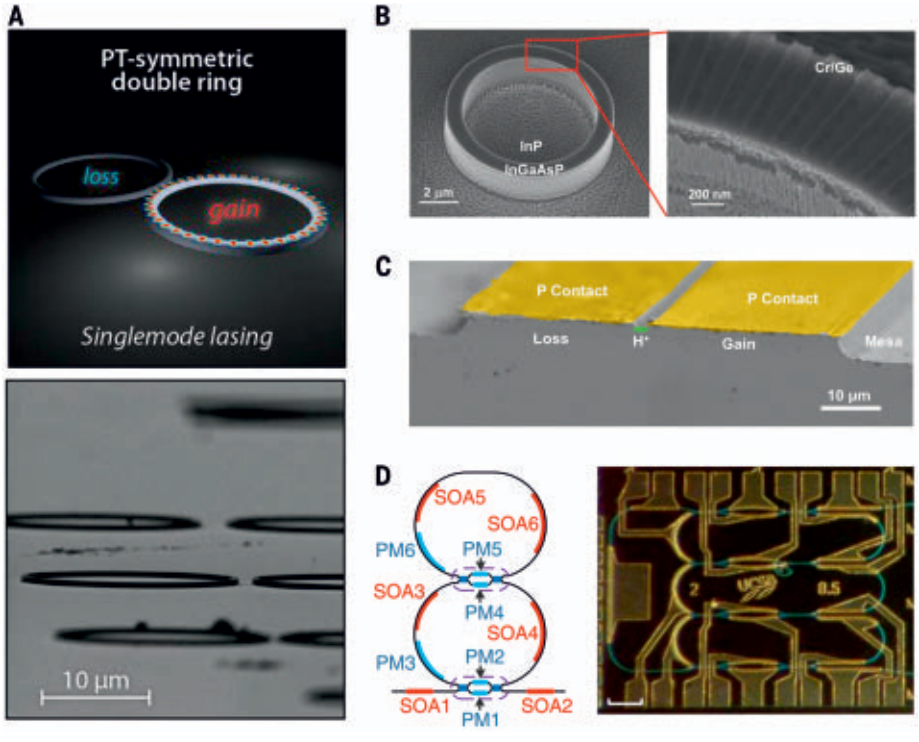


Fig. 6. PT-symmetric laser arrangement and its different realizations. (**A**) Coupled active-passive microring resonators (*105*), with a scanning electron microscope (SEM) image shown at the bottom. (**B**) SEM image of a microring resonator with an embedded gain-loss grating (*106*). (**C**) SEM image of coupled stripe lasers (*109*). (**D**) A schematic of integrated coupled microring lasers (left) and a photograph of the fabricated system (right), where the scale bar represents 200 μm (*111*). PM, phase modulator; SOA, semiconductor optical amplifier. [Credits: (A) and (B) reprinted from (*105*) and (*106*), respectively, with permission; (C) reprinted from (*109*) with permission from John Wiley and Sons; (D) reprinted from (*111*) with permission from Springer Nature]

lasers (*110*). In addition, integrated coupled microring lasers have been demonstrated with single-mode operation at telecommunication wavelengths (*111*) (Fig. 6D).

As illustrated in Fig. 7A, an interesting aspect of exceptional points consists of their exotic topological features in the parameter space. This discussion falls into the broad context of topological photonics, an area of optics research that has produced considerable excitement in recent years. Inspired by the unusual physics of topological insulators in condensed-matter physics, topological phenomena in photonics have been shown to arise in sophisticated periodic structures, ranging from gyromagnetic photonic crystals (*112*), arrays of helical waveguides (*113*), arrays of microring resonators (*114*), bianisotropic or magnetized metacrystals (*115*), dielectric chiral metasurfaces (*116*), and time-modulated lattices (*117*). In these systems, highly unusual photon transport, characterized by one-way propagation along the edges of the sample, arises within bandgaps delimited by bands with distinct topological properties. That their optical properties are related to a topological feature makes the response inherently robust to disorder and imperfections. Analogously, exceptional points represent an interesting example of topological features arising in simple coupled dynamical systems as a result of the interplay between interaction and dissipation. According to Fig. 7A, a loop of eigenvalues that encircle a base point identifies a topological object, given that it cannot be continuously deformed to a single point without crossing the base point.

The rigorous analysis of these features can be carried out using results from condensed-matter physics, in which the topological band theory of non-Hermitian Hamiltonians has been rigorously investigated in (*118*). Specifically, it was shown that non-Hermitian band structures exhibit a topological invariant associated with the gradient of the band in momentum space (*119*). Inspired by the periodic table of topological insulators, a systematic classification of topological phases of non-Hermitian systems has also been presented (*118*). An interesting problem in this context is to adiabatically change the parameters of a non-Hermitian system such that the exceptional point is dynamically encircled, as depicted in Fig. 7B. In a Hermitian system, when adiabatically changing the parameters along a closed path, the two eigenvectors are bound to return to their original form, apart from acquiring a possible geometric phase (*120*). In the case of non-Hermitian systems, instead, parametric cycling an exceptional point interchanges the instantaneous eigenvectors, whereas only one picks up the geometric phase (*13*, *121*–*123*). In principle, this behavior does not occur, even for arbitrarily slow dynamic cycling of the exceptional point, given that the adiabatic theorem breaks down in case of non-Hermitian systems. Indeed, under such conditions, depending on the direction of rotation, one of the two eigenstates dominates at the end of the parametric cycle. This interesting topological response provides a scheme for topologically robust energy conversion between different states.

On the basis of this principle, topological energy transfer has been recently demonstrated in a multimode optomechanical cavity in which two mechanical modes of a membrane are coupled and coherently controlled through a laser beam (*124*) (Fig. 7C). In addition, dynamical cycling of exceptional points is explored in a microwave waveguide in which a robust asymmetric transmission between even and odd modes is demonstrated (*125*) (Fig. 7D). In addition, it has been shown that this concept can provide opportunities for polarization manipulation (*126*, *127*). In particular, one can create an omnipolarizer in which the output light is polarized along a specific direction irrespective of the polarization of the input state (Fig. 7E). For propagation along the opposite direction, on the other hand, the output is populated in the orthogonal polarization.

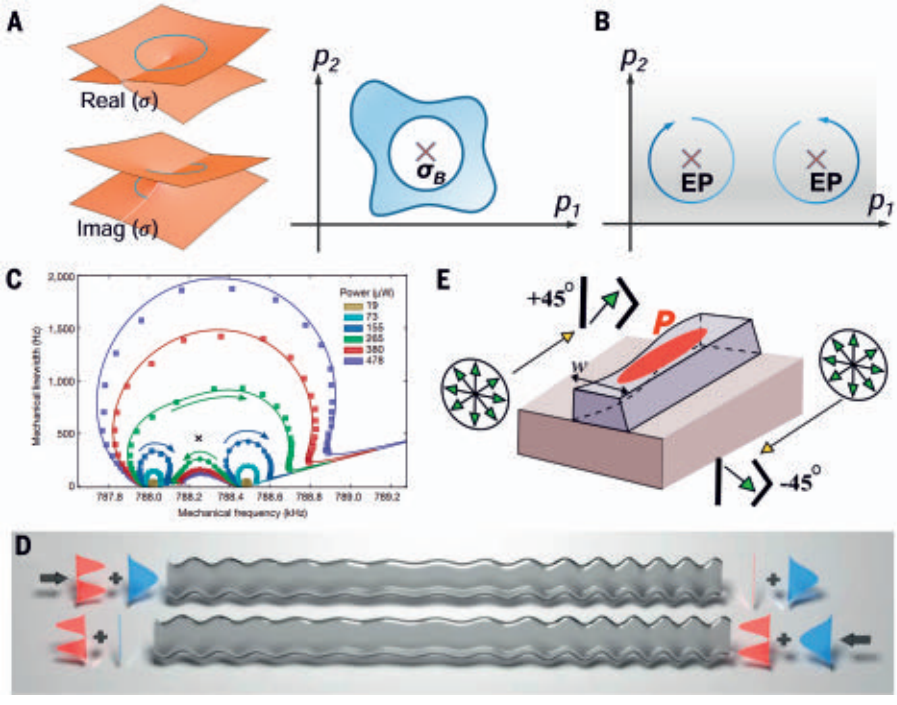


Fig. 7. Chiral mode conversion through dynamically cycling an exceptional point. (**A**) The eigenvalue surfaces near an exceptional point (left). Although a loop of eigenvalues containing a base point can be continuously deformed into a circle, it cannot be shrined into a point without crossing the base point (right) (*118*, *119*). p_1 and p_2 represent two parameters. (**B**) Two different possibilities of encircling an exceptional point (EP) cycling along opposite directions. (**C**) The experimental probing of the complex eigenvalues of two mechanical oscillators driven adiabatically through optical fields (*124*). The cross indicates the location of the exceptional point. (**D**) Asymmetric conversion between the even and odd modes of a waveguide, when the loss and detuning are adiabatically controlled in order to encircle an exceptional point (*125*). Blue and red curves indicate two modes of the waveguide, and the arrow indicates the direction of propagation. (**E**) An adiabatic conversion between orthogonal polarization states (*126*). Green arrows show the propagation direction, yellow arrows indicate the polarization state, *P* is the pumping, and *w* is the channel width. [Credits: (A) reprinted with permission from (*118*), copyright 2018 by the American Physical Society, and (*119*); (C) and (D) reprinted from (*124*) and (*125*) with permission from Springer Nature; (E) reprinted with permission from (*126*), copyright 2017 by the American Physical Society]

Conclusions and outlook

The peculiar features of exceptional points, associated with their unusual parameter dependence in the eigenvalue spectrum of non-Hermitian systems, enable exciting opportunities for a wide range of applications. These applications arise in scenarios in which interaction among different modes in the presence of dissipation and/or amplification is involved. In such circumstances, coupling and gain-loss mechanisms can be engineered and utilized to induce and control exceptional points, to take advantage of the strong and anomalous parameter dependence of the system around them.

We envision future opportunities to exploit these singular responses in photonics for advanced dispersion engineering. As a relevant recent example, level repulsion in the group velocity dispersion between coupled cavities has been used to control the modal dispersion of an individual cavity. This has been utilized to create anomalous dispersion, which is of great importance in four-wave mixing and parametric frequency comb generation (*128*–*130*). However, the full potential of coupled waveguide or cavity

arrangements for dispersion manipulation is still largely unexplored, and multiple coupled cavities or metamaterials may be envisioned to take full advantage of exceptional points in the context of dispersion engineering.

In a similar fashion, coupled-cavity arrangements offer exciting prospects to design new semiconductor lasers with highly desired functionalities. Although modern semiconductor laser sources exist in the entire optical spectrum, their coherence properties are not sufficient for many applications. In particular, key requirements for laser sources, such as stable and narrowband frequency operation, as well as frequency tunability, can be achieved through coupled-cavity geometries (*131*–*134*) (Fig. 8B). Even though this scheme has been previously applied to semiconductor lasers at specific frequencies, it remains to be explored in other, arguably more practical, sources and at different frequencies. In this regard, coupled-cavity techniques in conjunction with non-Hermitian designs provide an exciting strategy to systematically address the current challenges in integrated laser sources by taking advantage of the strong parameter dependence of such structures near exceptional points.

Mode conversion in a compact integrated photonic device is another important functionality that can largely benefit from exceptional points, in terms of reduced footprint and inherent robustness to disorder. Even though rigorous optimization techniques allow for inverse design of such structures, often resulting in complex structures that require advanced fabrication technologies, alternative designs with reduced complexity are highly desirable. In this vein, adiabatic perturbation of a structural parameter inducing an exceptional point–induced controllable level repulsion can provide a simple approach for hybridization and adiabatic exchange of modes. Recently, it has been shown that in optical ridge waveguides with different cladding and buffer materials, varying the waveguide width induces a strong coupling between transverse electric and magnetic polarizations of different spatial orders (*135*). As a result, adiabatic tapering of the waveguide width along the propagation direction can efficiently convert polarization states as well as spatial-mode orders (*136*, *137*). As shown schematically in Fig. 8C, the inclusion of selective gain and loss in such geometries provides an alternative degree of freedom to control the mode-conversion efficiency. In addition, hybridization between multiple modes through higher-order exceptional points can initiate the simultaneous conversion among a large number of modes. The full ramifications of these concepts become very powerful new tools in photonic engineering.

The quest for integration of optical setups on a chip requires integrated implementation of fundamental elements such as laser sources with critical power and coherence demands, isolators and circulators, mode convertors, and so on. In this regard, multimode structures have proven to provide a great opportunity to achieve desired functionalities and realize compact devices. This trend naturally calls for a bottom-up approach in designing photonic devices in an abstract modal picture in which three ingredients are relevant: (i) modal detuning, (ii) mode coupling, and (iii) modal gain and/or loss. The role of the first two processes has been largely explored in the past in the context of coupled-mode theory. The third mechanism, on the other hand, has been largely unexplored. As we discussed in this survey, the interplay of these phenomena can result into totally new opportunities for photonics, associated with the emergence of exceptional points that notably alter the eigenvalue surfaces. Therefore, notions from exceptional point physics can provide new designs for realizing multimode integrated photonic devices. This creates opportunities for theoretical and experimental research focused on exploring the fundamental bounds of accessible performance, such as bandwidth and sensitivity, of photonic devices operating at exceptional points. It is worth stressing that inducing exceptional points through gain and loss imposes difficulties in experimental photonics. This is because optical gain is limited to certain materials and is not generally compatible with all platforms, and loss is generally undesired for various purposes. At the same time, suitable settings for investigating and fruitfully exploiting exceptional points arise in systems that inherently involve optical gain or loss, such as semiconductor lasers, saturable absorbers, and plasmonic structures, among others.

Along different lines, remaining to be investigated are the interesting physics arising from the propagation of classical light at exceptional point singularities. Recent theoretical investigations, for example, suggest dynamical slowing and stopping of light in coupled waveguides at exceptional points (*138*), as well as photonic catastrophe in optical lattices (*139*). In addition, a point of interest would be to explore these phenomena in new platforms. An emerging playground to explore the rich physics of exceptional points is provided by hybrid photonic platforms

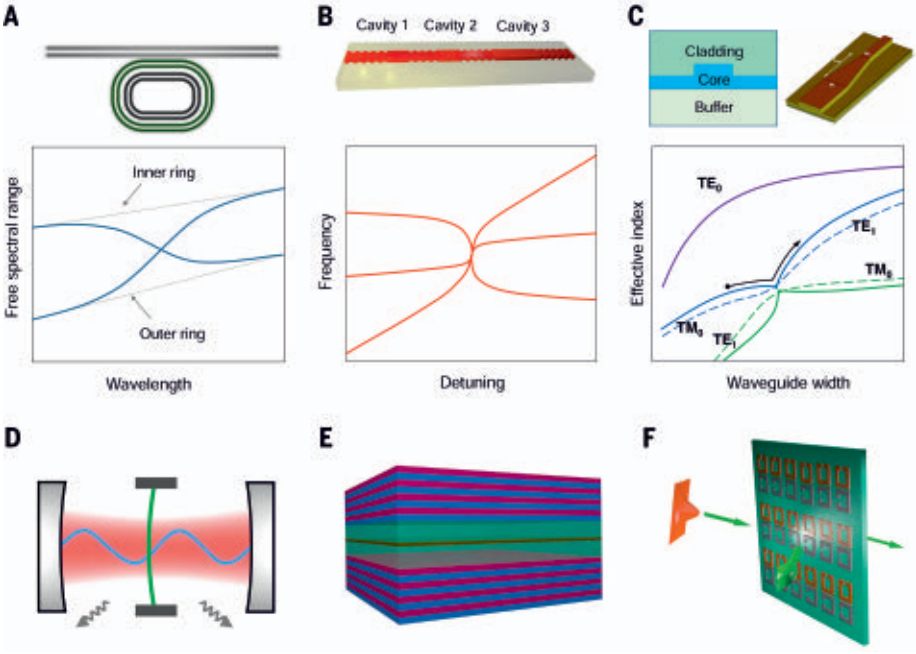


Fig. 8. Application of exceptional points in multimode photonic integrated circuits and new platforms to investigate exceptional points. (**A** to **C**) Applications. (A) Hybridization of eigenfrequencies in coupled microring resonators (top) creates two branches with strong dispersion (bottom) (*130*). The anomalous dispersion can be utilized for frequency comb generation. (B) Wavelength manipulation in three coupled-cavity lasers through a strong dispersion at a third-order exceptional point (*133*). (C) Level repulsion of modes with different polarization provides an opportunity for compact polarization mode conversion (*135*, *136*). A parametric evaluation of the eigenmodes of a rib waveguide (top left) versus the waveguide width reveals a level repulsion between transverse electric (TE) and transverse magnetic (TM) polarizations (bottom). Therefore, tapering of the waveguide width over a finite distance (top right) can result in an adiabatic polarization conversion. (**D** to **F**) New platforms. (D) Multimode optomechanical cavities provide a flexible platform for investigating exceptional points. (E) Exciton-polaritons in semiconductor cavities offer an alternative multiphysics structure for realizing exceptional points. (F) Coupled nanoantennas can be designed as non-Hermitian building blocks of optical metasurfaces. [Credits: (A) reprinted from (*130*) with permission from Springer Nature; (B) reprinted from (*133*) with permission from AIP Publishing; (C) reprinted with permission from (*135*) and (*136*), copyright 2011 and 2012, respectively, Optical Society of America]

that integrate other degrees of freedom beyond optics, exploiting the interaction between different phenomena. In particular, cavity optomechanics, relying on the strong coupling between optics and mechanical motion, offers a reconfigurable, inherently non-Hermitian platform that can be externally controlled through pump lasers with proper intensity and phase (*140*) (Fig. 8D). Operating in the red and blue sideband detuning of the pump beam can effectively control loss or gain for the optical modes involved, opening exciting opportunities for PT symmetry and exceptional points in a low-noise nanophotonic integrated environment. Similarly, cavity polaritons, because of their inherent non-Hermitian properties, can provide another platform for investigating and utilizing exceptional points (*141*) (Fig. 8E).

Finally, it is worth mentioning the potential of utilizing exceptional point singularities in optical scattering problems, where the coupling between discrete localized metastable states and a continuum of radiation states is concerned. Interest in photonic bound states embedded in the continuum is increasing, owing to their interesting properties (*142*–*144*). Such settings can, in general, be treated as non-Hermitian problems, for which a point of interest would be to explore the connection between radiation leakage and exceptional points emerging in the continuum, as observed in recent experiments (*145*). In addition, similar concepts can be utilized in designing coupled optical nanoantennas as non-Hermitian building blocks of metasurfaces in order to create scattering surfaces with desired phase, frequency, and polarization response (Fig. 8F). In addition to the radiative losses of dielectric inclusions, the inherent loss in metallic inclusions at optical frequencies can be turned into an opportunity to realize and exploit exceptional points in properly designed geometries (*146*). We envision exciting opportunities in translating the concepts of exceptional point physics to quantum nanophotonic and low-photon hybrid systems.

REFERENCES AND NOTES

1. N. J. Dunford, J. T. Schwartz, *Linear Operators, Parts I and II* (Wiley, 1988).
2. M. Morse, H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, 1953).
3. C. R. Wylie, *Advanced Engineering Mathematics* (McGraw-Hill, ed. 2, 1960).
4. G. Strang, *Introduction to Linear Algebra* (Wellesley-Cambridge Press, ed. 2, 1993).
5. W. D. Heiss, The physics of exceptional points. *J. Phys. A Math. Theor.* **45**, 444016 (2012). doi: 10.1088/1751-8113/45/44/444016
6. T. Kato, *Perturbation Theory of Linear Operators* (Springer, 1966).
7. M. V. Berry, D. H. J. O'Dell, Diffraction by volume gratings with imaginary potentials. *J. Phys. Math. Gen.* **31**, 2093–2101 (1998). doi: 10.1088/0305-4470/31/8/019
8. W. D. Heiss, Phases of wave functions and level repulsion. *Eur. Phys. J. D* **7**, 1–4 (1999). doi: 10.1007/s100530050339
9. M. V. Berry, Physics of nonhermitian degeneracies. *Czech. J. Phys.* **54**, 1039–1047 (2004). doi: 10.1023/B:CJOP.0000044002.05657.04
10. W. D. Heiss, Exceptional points of non-Hermitian operators. *J. Phys. Math. Gen.* **37**, 2455–2464 (2004). doi: 10.1088/0305-4470/37/6/034
11. N. Moiseyev, Quantum theory of resonances: Calculating energies, widths and cross-sections by complex scaling. *Phys. Rep.* **302**, 212–293 (1998). doi: 10.1016/S0370-1573(98)00002-7
12. E. Persson, I. Rotter, H. Stöckmann, M. Barth, Observation of resonance trapping in an open microwave cavity. *Phys. Rev. Lett.* **85**, 2478–2481 (2000). doi: 10.1103/PhysRevLett.85.2478; pmid: 10978086
13. C. Dembowski *et al*., Experimental observation of the topological structure of exceptional points. *Phys. Rev. Lett.* **86**, 787–790 (2001). doi: 10.1103/PhysRevLett.86.787; pmid: 11177940
14. H. Wenzel, U. Bandelow, H. J. Wunsche, J. Rehberg, Mechanisms of fast self pulsations in two-section DFB lasers. *IEEE J. Quantum Electron.* **32**, 69–78 (1996). doi: 10.1109/3.481922
15. M. V. Berry, Mode degeneracies and the Petermann excess-noise factor for unstable lasers. *J. Mod. Opt.* **50**, 63–81 (2003). doi: 10.1080/09500340308234532
16. S. Bernet *et al*., Matter waves in time-modulated complex light potentials. *Phys. Rev. A* **62**, 023606 (2000). doi: 10.1103/PhysRevA.62.023606
17. C. M. Bender, S. Boettcher, Real spectra in non-Hermitian Hamiltonians having PT symmetry. *Phys. Rev. Lett.* **80**, 5243–5246 (1998). doi: 10.1103/PhysRevLett.80.5243
18. G. Lévai, M. Znojil, Systematic search for PT-symmetric potentials with real energy spectra. *J. Phys. Math. Gen.* **33**, 7165–7180 (2000). doi: 10.1088/0305-4470/33/40/313
19. Z. Ahmed, Real and complex discrete eigenvalues in an exactly solvable onedimensional complex PT-invariant potential. *Phys. Lett. A* **282**, 343–348 (2001). doi: 10.1016/S0375-9601(01)00218-3
20. C. M. Bender, D. C. Brody, H. F. Jones, Complex extension of quantum mechanics. *Phys. Rev. Lett.* **89**, 270401 (2002). doi: 10.1103/PhysRevLett.89.270401; pmid: 12513185
21. A. Mostafazadeh, Pseudo-Hermiticity versus PT symmetry: The necessary condition for the reality of the spectrum of a non-Hermitian Hamiltonian. *J. Math. Phys.* **43**, 205–214 (2002). doi: 10.1063/1.1418246
22. M. Bender, Making sense of non-Hermitian Hamiltonians. *Rep. Prog. Phys.* **70**, 947–1018 (2007). doi: 10.1088/0034-4885/70/6/R03
23. N. Moiseyev, *Non-Hermitian Quantum Mechanics* (Cambridge Univ. Press, 2011).
24. R. El-Ganainy, K. G. Makris, D. N. Christodoulides, Z. H. Musslimani, Theory of coupled optical PT-symmetric structures. *Opt. Lett.* **32**, 2632–2634 (2007). doi: 10.1364/OL.32.002632; pmid: 17767329
25. K. G. Makris, R. El-Ganainy, D. N. Christodoulides, Z. H. Musslimani, Beam dynamics in PT symmetric optical lattices. *Phys. Rev. Lett.* **100**, 103904 (2008). doi: 10.1103/PhysRevLett.100.103904; pmid: 18352189
26. S. Klaiman, U. Günther, N. Moiseyev, Visualization of branch points in PT-symmetric waveguides. *Phys. Rev. Lett.* **101**, 080402 (2008). doi: 10.1103/PhysRevLett.101.080402; pmid: 18764593
27. S. Longhi, Bloch oscillations in complex crystals with PT symmetry. *Phys. Rev. Lett.* **103**, 123601 (2009). doi: 10.1103/PhysRevLett.103.123601; pmid: 19792436
28. A. A. Zyablovsky, A. P. Vinogradov, A. A. Pukhov, A. V. Dorofeenko, A. A. Lisyansky, PT-symmetry in optics. *Phys. Uspekhi* **57**, 1063–1082 (2014). doi: 10.3367/UFNe.0184.201411b.1177
29. V. V. Konotop, J. Yang, D. A. Zezyulin, Nonlinear waves in PT-symmetric systems. *Rev. Mod. Phys.* **88**, 035002 (2016). doi: 10.1103/RevModPhys.88.035002
30. L. Feng, R. El-Ganainy, L. Ge, Non-Hermitian photonics based on parity–time symmetry. *Nat. Photonics* **11**, 752–762 (2017). doi: 10.1038/s41566-017-0031-1
31. S. Longhi, Parity-time symmetry meets photonics: A new twist in non-Hermitian optics. *EPL* **120**, 64001 (2018). doi: 10.1209/0295-5075/120/64001
32. R. El-Ganainy *et al*., Non-Hermitian physics and PT symmetry. *Nat. Phys.* **14**, 11–19 (2018). doi: 10.1038/nphys4323
33. H. A. Haus, W. Huang, Coupled-mode theory. *Proc. IEEE* **79**, 1505–1518 (1991). doi: 10.1109/5.104225
34. A. Yariv, Coupled-mode theory for guided-wave optics. *IEEE J. Quantum Electron.* **9**, 919–933 (1973). doi: 10.1109/JQE.1973.1077767
35. J. P. Gordon, H. Kogelnik, PMD fundamentals: Polarization mode dispersion in optical fibers. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4541–4550 (2000). doi: 10.1073/pnas.97.9.4541; pmid: 10781059
36. H. Kogelnik, Coupled wave theory for thick hologram gratings. *Bell Syst. Tech. J.* **48**, 2909–2947 (1969). doi: 10.1002/j.1538-7305.1969.tb01198.x
37. R. Stolen, J. Bjorkholm, Parametric amplification and frequency conversion in optical fibers. *IEEE J. Quantum Electron.* **18**, 1062–1072 (1982). doi: 10.1109/JQE.1982.1071660
38. M. Aspelmeyer, T. J. Kippenberg, F. Marquardt, Cavity optomechanics. *Rev. Mod. Phys.* **86**, 1391–1452 (2014). doi: 10.1103/RevModPhys.86.1391
39. M. Fox, *Quantum Optics: An Introduction* (Oxford Univ. Press, 2006).
40. E. M. Graefe, U. Günther, H. J. Korsch, A. E. Niederle, A non-Hermitian PT-symmetric Bose–Hubbard model: Eigenvalue rings from unfolding higher-order exceptional points. *J. Phys. A Math. Theor.* **41**, 255206 (2008). doi: 10.1088/1751-8113/41/25/255206
41. G. Demange, E.-M. Graefe, Signatures of three coalescing eigenfunctions. *J. Phys. A Math. Theor.* **45**, 025303 (2012). doi: 10.1088/1751-8113/45/2/025303
42. J. Franklin, *Matrix Theory* (Dover Publications, 1993).
43. M. V. Berry, M. Wilkinson, Diabolical points in the spectra of triangles. *Proc. R. Soc. London Ser. A* **392**, 15–43 (1984). doi: 10.1098/rspa.1984.0022
44. D. R. Yarkony, Diabolical conical intersections. *Rev. Mod. Phys.* **68**, 985–1013 (1996). doi: 10.1103/RevModPhys.68.985
45. A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, A. K. Geim, The electronic properties of grapheme. *Rev. Mod. Phys.* **81**, 109–162 (2009). doi: 10.1103/RevModPhys.81.109
46. H.-J. Stockmann, *Quantum Chaos: An Introduction* (Cambridge Univ. Press, 1999).
47. L. Novotny, Strong coupling, energy splitting, and level crossings: A classical perspective. *Am. J. Phys.* **78**, 1199–1202 (2010). doi: 10.1119/1.3471177
48. W. D. Heiss, Repulsion of resonance states and exceptional points. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* **61**, 929–932 (2000). doi: 10.1103/PhysRevE.61.929; pmid: 11046343
49. P. von Brentano, M. Philipp, Crossing and anticrossing of energies and widths for unbound levels. *Phys. Lett. B* **454**, 171–175 (1999). doi: 10.1016/S0370-2693(99)00378-0
50. C. E. Rüter *et al*., Observation of parity-time symmetry in optics. *Nat. Phys.* **6**, 192–195 (2010). doi: 10.1038/nphys1515
51. A. I. Magunov, I. Rotter, S. I. Strakhova, Avoided level crossing and population trapping in atoms. *Physica E* **9**, 474–477 (2001). doi: 10.1016/S1386-9477(00)00247-2
52. S. Steshenko, F. Capolino, "Single dipole approximation for modeling collections of nanoscatterers," in *Theory and Phenomena of Metamaterials*, F. Capolino, Ed. (CRC Press, 2009).
53. C. M. Gentry, M. A. Popović, Dark state lasers. *Opt. Lett.* **39**, 4136–4139 (2014). doi: 10.1364/OL.39.004136; pmid: 25121670
54. R. E. Collin, *Field Theory of Guided Waves* (Wiley-IEEE Press, ed. 2, 1991).
55. Y. D. Chong, L. Ge, A. D. Stone, PT-symmetry breaking and laser-absorber modes in optical scattering systems. *Phys. Rev. Lett.* **106**, 093902 (2011). doi: 10.1103/PhysRevLett.106.093902; pmid: 21405622
56. A. Guo *et al*., Observation of PT-symmetry breaking in complex optical potentials. *Phys. Rev. Lett.* **103**, 093902 (2009). doi: 10.1103/PhysRevLett.103.093902; pmid: 19792798
57. B. Peng *et al*., Parity–time-symmetric whispering-gallery microcavities. *Nat. Phys.* **10**, 394–398 (2014). doi: 10.1038/nphys2927
58. L. Chang *et al*., Parity–time symmetry and variable optical isolation in active–passive-coupled microresonators. *Nat. Photonics* **8**, 524–529 (2014). doi: 10.1038/nphoton.2014.133
59. A. Regensburger *et al*., Parity-time synthetic photonic lattices. *Nature* **488**, 167–171 (2012). doi: 10.1038/nature11298; pmid: 22874962
60. B. Zhen *et al*., Spawning rings of exceptional points out of Dirac cones. *Nature* **525**, 354–358 (2015). doi: 10.1038/nature14889; pmid: 26352476
61. S.-B. Lee *et al*., Observation of an exceptional point in a chaotic optical microcavity. *Phys. Rev. Lett.* **103**, 134101 (2009). doi: 10.1103/PhysRevLett.103.134101; pmid: 19905515
62. S. Longhi, PT-symmetric laser absorber. *Phys. Rev. A* **82**, 031801 (2010). doi: 10.1103/PhysRevA.82.031801
63. Z. J. Wong *et al*., Lasing and anti-lasing in a single cavity. *Nat. Photonics* **10**, 796–801 (2016). doi: 10.1038/nphoton.2016.216

64. M. Kulishov, J. Laniel, N. Bélanger, J. Azaña, D. Plant, Nonreciprocal waveguide Bragg gratings. *Opt. Express* **13**, 3068–3078 (2005). doi: 10.1364/OPEX.13.003068; pmid: 19495203
65. Z. Lin *et al*., Unidirectional invisibility induced by PT-symmetric periodic structures. *Phys. Rev. Lett.* **106**, 213901 (2011). doi: 10.1103/PhysRevLett.106.213901; pmid: 21699297
66. G. Castaldi, S. Savoia, V. Galdi, A. Alù, N. Engheta, PT metamaterials via complex-coordinate transformation optics. *Phys. Rev. Lett.* **110**, 173901 (2013). doi: 10.1103/PhysRevLett.110.173901; pmid: 23679728
67. M.-A. Miri, A. B. Aceves, T. Kottos, V. Kovanis, D. N. Christodoulides, Bragg solitons in nonlinear PT-symmetric periodic potentials. *Phys. Rev. A* **86**, 033801 (2012). doi: 10.1103/PhysRevA.86.033801
68. L. Feng *et al*., Experimental demonstration of a unidirectional reflectionless parity-time metamaterial at optical frequencies. *Nat. Mater.* **12**, 108–113 (2013). doi: 10.1038/nmat3495; pmid: 23178268
69. Y. Yan, N. C. Giebink, Passive PT symmetry in organic composite films via complex refractive index modulation. *Adv. Opt. Mater.* **2**, 423–427 (2014). doi: 10.1002/adom.201400021
70. R. Fleury, D. Sounas, A. Alù, An invisible acoustic sensor based on parity-time symmetry. *Nat. Commun.* **6**, 5905 (2015). doi: 10.1038/ncomms6905; pmid: 25562746
71. P. Miao *et al*., Orbital angular momentum microlaser. *Science* **353**, 464–467 (2016). doi: 10.1126/science.aaf8533; pmid: 27471299
72. J. Wiersig *et al*., Nonorthogonal pairs of copropagating optical modes in deformed microdisk cavities. *Phys. Rev. A* **84**, 023845 (2011). doi: 10.1103/PhysRevA.84.023845
73. B. Peng *et al*., Chiral modes and directional lasing at exceptional points. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 6845–6850 (2016). doi: 10.1073/pnas.1603318113; pmid: 27274059
74. M. Kim, K. Kwon, J. Shim, Y. Jung, K. Yu, Partially directional microdisk laser with two Rayleigh scatterers. *Opt. Lett.* **39**, 2423–2426 (2014). doi: 10.1364/OL.39.002423; pmid: 24979009
75. R. Fleury, D. L. Sounas, A. Alù, Negative refraction and planar focusing based on parity-time symmetric metasurfaces. *Phys. Rev. Lett.* **113**, 023903 (2014). doi: 10.1103/PhysRevLett.113.023903; pmid: 25062184
76. X. Zhu, L. Feng, P. Zhang, X. Yin, X. Zhang, One-way invisible cloak using parity-time symmetric transformation optics. *Opt. Lett.* **38**, 2821–2824 (2013). doi: 10.1364/OL.38.002821; pmid: 23903152
77. D. L. Sounas, R. Fleury, A. Alù, Unidirectional cloaking based on metasurfaces with balanced loss and gain. *Phys. Rev. Appl.* **4**, 014005 (2015). doi: 10.1103/PhysRevApplied.4.014005
78. M. Fleischhauer, A. Imamoglu, J. P. Marangos, Electromagnetically induced transparency: Optics in coherent media. *Rev. Mod. Phys.* **77**, 633–673 (2005). doi: 10.1103/RevModPhys.77.633
79. C. Hang, G. Huang, V. V. Konotop, PT symmetry with a system of three-level atoms. *Phys. Rev. Lett.* **110**, 083604 (2013). doi: 10.1103/PhysRevLett.110.083604; pmid: 23473145
80. J. Sheng, M.-A. Miri, D. N. Christodoulides, M. Xiao, PT-symmetric optical potentials in a coherent atomic medium. *Phys. Rev. A* **88**, 041803 (2013). doi: 10.1103/PhysRevA.88.041803
81. P. Peng *et al*., Anti-parity–time symmetry with flying atoms. *Nat. Phys.* **12**, 1139–1145 (2016). doi: 10.1038/nphys3842
82. Z. Zhang *et al*., Observation of parity-time symmetry in optically induced atomic lattices. *Phys. Rev. Lett.* **117**, 123601 (2016). doi: 10.1103/PhysRevLett.117.123601; pmid: 27689270
83. M. J. Weber, *Handbook of Optical Materials* (CRC Press, 2002).
84. H. Ramezani, T. Kottos, R. El-Ganainy, D. N. Christodoulides, Unidirectional nonlinear PT-symmetric optical structures. *Phys. Rev. A* **82**, 043803 (2010). doi: 10.1103/PhysRevA.82.043803
85. P. Aleahmad, M. Khajavikhan, D. Christodoulides, P. LiKamWa, Integrated multi-port circulators for unidirectional optical information transport. *Sci. Rep.* **7**, 2129 (2017). doi: 10.1038/s41598-017-02340-9; pmid: 28522872
86. M. Liertzer *et al*., Pump-induced exceptional points in lasers. *Phys. Rev. Lett.* **108**, 173901 (2012). doi: 10.1103/PhysRevLett.108.173901; pmid: 22680867
87. M. Brandstetter *et al*., Reversing the pump dependence of a laser at an exceptional point. *Nat. Commun.* **5**, 4034 (2014). doi: 10.1038/ncomms5034; pmid: 24925314
88. B. Peng *et al*., Loss-induced suppression and revival of lasing. *Science* **346**, 328–332 (2014). doi: 10.1126/science.1258004; pmid: 25324384
89. Z. H. Musslimani, K. G. Makris, R. El-Ganainy, D. N. Christodoulides, Optical solitons in PT periodic potentials. *Phys. Rev. Lett.* **100**, 030402 (2008). doi: 10.1103/PhysRevLett.100.030402; pmid: 18232949
90. N. Akhmediev, A. Ankiewicz, *Dissipative Solitons* (Springer, 2005).
91. A. E. Miroshnichenko, B. A. Malomed, Y. S. Kivshar, Nonlinearly PT-symmetric systems: Spontaneous symmetry breaking and transmission resonances. *Phys. Rev. A* **84**, 012123 (2011). doi: 10.1103/PhysRevA.84.012123
92. S. Nixon, L. Ge, J. Yang, Stability analysis for solitons in PT-symmetric optical lattices. *Phys. Rev. A* **85**, 023822 (2012). doi: 10.1103/PhysRevA.85.023822
93. M. Wimmer *et al*., Observation of optical solitons in PT-symmetric lattices. *Nat. Commun.* **6**, 7782 (2015). doi: 10.1038/ncomms8782; pmid: 26215165
94. R. W. Boyd, *Nonlinear Optics* (Academic Press, 2003).
95. J. P. Dowling, M. Scalora, M. J. Bloemer, C. M. Bowden, The photonic band edge laser: A new approach to gain enhancement. *J. Appl. Phys.* **75**, 1896–1899 (1994). doi: 10.1063/1.356336
96. C. M. Bender, S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers I: Asymptotic Methods and Perturbation Theory* (McGraw-Hill, 1978).
97. J. Wiersig, Enhancing the sensitivity of frequency and energy splitting detection by using exceptional points: Application to microcavity sensors for single-particle detection. *Phys. Rev. Lett.* **112**, 203901 (2014). doi: 10.1103/PhysRevLett.112.203901
98. Z. P. Liu *et al*., Metrology with PT-symmetric cavities: Enhanced sensitivity near the PT-phase transition. *Phys. Rev. Lett.* **117**, 110802 (2016). doi: 10.1103/PhysRevLett.117.110802; pmid: 27661674
99. W. Chen, Ş. Kaya Özdemir, G. Zhao, J. Wiersig, L. Yang, Exceptional points enhance sensing in an optical microcavity. *Nature* **548**, 192–196 (2017). doi: 10.1038/nature23281; pmid: 28796206
100. H. Hodaei *et al*., Enhanced sensitivity at higher-order exceptional points. *Nature* **548**, 187–191 (2017). doi: 10.1038/nature23280; pmid: 28796201
101. W. Langbein, No exceptional precision of exceptional-point sensors. *Phys. Rev. A* **98**, 023805 (2018). doi: 10.1103/PhysRevA.98.023805
102. M. Zhang, W. Sweeney, C. W. Hsu, L. Yang, A. D. Stone, L. Jiang, Quantum noise theory of exceptional point sensors. arxiv:1805.12001 [quant-ph] (30 May 2018).
103. P.-Y. Chen *et al*., Generalized parity–time symmetry condition for enhanced sensor telemetry. *Nat. Electron.* **1**, 297–304 (2018). doi: 10.1038/s41928-018-0072-6
104. M.-A. Miri, P. LiKamWa, D. N. Christodoulides, Large area single-mode parity-time-symmetric laser amplifiers. *Opt. Lett.* **37**, 764–766 (2012). doi: 10.1364/OL.37.000764; pmid: 22378386
105. H. Hodaei, M.-A. Miri, M. Heinrich, D. N. Christodoulides, M. Khajavikhan, Parity-time-symmetric microring lasers. *Science* **346**, 975–978 (2014). doi: 10.1126/science.1258480; pmid: 25414308
106. L. Feng, Z. J. Wong, R. M. Ma, Y. Wang, X. Zhang, Single-mode laser by parity-time symmetry breaking. *Science* **346**, 972–975 (2014). doi: 10.1126/science.1258479; pmid: 25414307
107. H. Hodaei *et al*., Single mode lasing in transversely multi-moded PT-symmetric microring resonators. *Laser Photonics Rev.* **10**, 494–499 (2016). doi: 10.1002/lpor.201500292
108. Z. Gu *et al*., Experimental demonstration of PT-symmetric stripe lasers. *Laser Photonics Rev.* **10**, 588–594 (2016). doi: 10.1002/lpor.201500114
109. R. Yao, C.-S. Lee, V. Podolskiy, W. Guo, Electrically injected parity time–symmetric single transverse–mode lasers. *Laser Photonics Rev.* 10.1002/lpor.201800154 (2018). doi: 10.1002/lpor.201500114
110. N. Zhang *et al*., Quasiparity-time symmetric microdisk laser. *Laser Photonics Rev.* **11**, 1700052 (2017). doi: 10.1002/lpor.201700052
111. W. Liu *et al*., An integrated parity-time symmetric wavelength-tunable single-mode microring laser. *Nat. Commun.* **8**, 15389 (2017). doi: 10.1038/ncomms15389; pmid: 28497784
112. Z. Wang, Y. Chong, J. D. Joannopoulos, M. Soljačić, Observation of unidirectional backscattering-immune topological electromagnetic states. *Nature* **461**, 772–775 (2009). doi: 10.1038/nature08293; pmid: 19812669
113. M. C. Rechtsman *et al*., Photonic Floquet topological insulators. *Nature* **496**, 196–200 (2013). doi: 10.1038/nature12066; pmid: 23579677
114. M. Hafezi, S. Mittal, J. Fan, A. Migdall, J. M. Taylor, Imaging topological edge states in silicon photonics. *Nat. Photonics* **7**, 1001–1005 (2013). doi: 10.1038/nphoton.2013.274
115. X. Ni *et al*., Spin- and valley-polarized one-way Klein tunneling in photonic topological insulators. *Sci. Adv.* **4**, eaap8802 (2018). doi: 10.1126/sciadv.aap8802; pmid: 29756032
116. M. A. Gorlach *et al*., Far-field probing of leaky topological states in all-dielectric metasurfaces. *Nat. Commun.* **9**, 909 (2018). doi: 10.1038/s41467-018-03330-9; pmid: 29500466
117. R. Fleury, A. B. Khanikaev, A. Alù, Floquet topological insulators for sound. *Nat. Commun.* **7**, 11744 (2016). doi: 10.1038/ncomms11744; pmid: 27312175
118. Z. Gong *et al*., Topological phases of non-Hermitian systems. *Phys. Rev. X* **8**, 031079 (2018). doi: 10.1103/PhysRevLett.120.146402; pmid: 29694133
119. H. Shen, B. Zhen, L. Fu, Topological band theory for non-Hermitian Hamiltonians. *Phys. Rev. Lett.* **120**, 146402 (2018). doi: 10.1103/PhysRevLett.120.146402; pmid: 29694133
120. M. V. Berry, Quantal phase factors accompanying adiabatic changes. *Proc. R. Soc. London Ser. A* **392**, 45–57 (1984). doi: 10.1098/rspa.1984.0023
121. C. Dembowski *et al*., Encircling an exceptional point. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **69**, 056216 (2004). doi: 10.1103/PhysRevE.69.056216; pmid: 15244913
122. A. A. Mailybaev, O. N. Kirillov, A. P. Seyranian, Geometric phase around exceptional points. *Phys. Rev. A* **72**, 014104 (2005). doi: 10.1103/PhysRevA.72.014104
123. R. Uzdin, N. Moiseyev, Scattering from a waveguide by cycling a non-Hermitian degeneracy. *Phys. Rev. A* **85**, 031804 (2012). doi: 10.1103/PhysRevA.85.031804
124. H. Xu, D. Mason, L. Jiang, J. G. E. Harris, Topological energy transfer in an optomechanical system with exceptional points. *Nature* **537**, 80–83 (2016). doi: 10.1038/nature18604; pmid: 27454555
125. J. Doppler *et al*., Dynamically encircling an exceptional point for asymmetric mode switching. *Nature* **537**, 76–79 (2016). doi: 10.1038/nature18605; pmid: 27454554
126. A. U. Hassan, B. Zhen, M. Soljačić, M. Khajavikhan, D. N. Christodoulides, Dynamically encircling exceptional points: Exact evolution and polarization state conversion. *Phys. Rev. Lett.* **118**, 093002 (2017). doi: 10.1103/PhysRevLett.118.093002; pmid: 28306295
127. S. N. Ghosh, Y. D. Chong, Exceptional points and asymmetric mode conversion in quasi-guided dual-mode optical waveguides. *Sci. Rep.* **6**, 19837 (2016). doi: 10.1038/srep19837; pmid: 27101933
128. Y. Liu *et al*., Investigation of mode coupling in normal-dispersion silicon nitride microresonators for Kerr frequency comb generation. *Optica* **1**, 137–144 (2014). doi: 10.1364/OPTICA.1.000137
129. S. Ramelow *et al*., Strong polarization mode coupling in microresonators. *Opt. Lett.* **39**, 5134–5137 (2014). doi: 10.1364/OL.39.005134; pmid: 25166092
130. S. Kim *et al*., Dispersion engineering and frequency comb generation in thin silicon nitride concentric microresonators. *Nat. Commun.* **8**, 372 (2017). pmid: 28851874
131. W. T. Tsang, N. A. Olsson, R. A. Logan, Stable single-longitudinal-mode operation under high-speed direct modulation in cleaved-coupled-cavity GaInAsP semiconductor lasers. *Electron. Lett.* **19**, 488–490 (1983). doi: 10.1049/el:19830331
132. L. Coldren, T. Koch, Analysis and design of coupled-cavity lasers—Part I: Threshold gain analysis and design guidelines. *IEEE J. Quantum Electron.* **20**, 659–670 (1984). doi: 10.1109/JQE.1984.1072438
133. P. Pellandini *et al*., Dual-wavelength laser emission from a coupled semiconductor microcavity. *Appl. Phys. Lett.* **71**, 864–866 (1997). doi: 10.1063/1.119671
134. Z. Gao, S. T. M. Fryslie, B. J. Thompson, P. S. Carney, K. D. Choquette, Parity-time symmetry in coherently coupled vertical cavity laser arrays. *Optica* **4**, 323–329 (2017). doi: 10.1364/OPTICA.4.000323
135. D. Dai, J. E. Bowers, Novel concept for ultracompact polarization splitter-rotator based on silicon nanowires. *Opt. Express* **19**, 10940–10949 (2011). doi: 10.1364/OE.19.010940; pmid: 21643354

136. D. Dai, Y. Tang, J. E. Bowers, Mode conversion in tapered submicron silicon ridge optical waveguides. *Opt. Express* **20**, 13425–13439 (2012). doi: 10.1364/OE.20.013425; pmid: 22714370
137. Z. Zhang, X. Hu, J. Wang, On-chip optical mode exchange using tapered directional coupler. *Sci. Rep.* **5**, 16072 (2015). doi: 10.1038/srep16072; pmid: 26530728
138. T. Goldzak, A. A. Mailybaev, N. Moiseyev, Light stops at exceptional points. *Phys. Rev. Lett.* **120**, 013901 (2018). doi: 10.1103/PhysRevLett.120.013901; pmid: 29350937
139. S. Longhi, Exceptional points and photonic catastrophe. *Opt. Lett.* **43**, 2929–2932 (2018). doi: 10.1364/OL.43.002929; pmid: 29905726
140. E. Verhagen, A. Alù, Optomechanical nonreciprocity. *Nat. Phys.* **13**, 922–924 (2017). doi: 10.1038/nphys4283
141. T. Byrnes, N. Y. Kim, Y. Yamamoto, Exciton–polariton condensates. *Nat. Phys.* **10**, 803–813 (2014). doi: 10.1038/nphys3143
142. Y. Plotnik *et al*., Experimental observation of optical bound states in the continuum. *Phys. Rev. Lett.* **107**, 183901 (2011). doi: 10.1103/PhysRevLett.107.183901; pmid: 22107630
143. A. Kodigala *et al*., Lasing action from photonic bound states in continuum. *Nature* **541**, 196–199 (2017). doi: 10.1038/nature20799; pmid: 28079064
144. H. M. Doeleman, F. Monticone, W. den Hollander, A. Alù, A. F. Koenderink, Experimental observation of a polarization vortex at an optical bound state in the continuum. *Nat. Photonics* **12**, 397–401 (2018). doi: 10.1038/s41566-018-0177-5
145. A. Regensburger *et al*., Observation of defect states in PT-symmetric optical lattices. *Phys. Rev. Lett.* **110**, 223902 (2013). doi: 10.1103/PhysRevLett.110.223902; pmid: 23767725
146. J. B. Khurgin, How to deal with the loss in plasmonics and metamaterials. *Nat. Nanotechnol.* **10**, 2–6 (2015). doi: 10.1038/nnano.2014.310; pmid: 25559961

ACKNOWLEDGMENTS

Funding: This work was supported by the Office of Naval Research, the Simons Foundation, the Air Force Office of Scientific Research, and the National Science Foundation. **Competing interests:** None declared.

10.1126/science.aar7709

FORCEFORSCIENCE.ORG

STAND TOGETHER
Be a Force for Science

GET THE FACTS

Understand the science behind the issues that matter.

FOLLOW AAAS ADVOCACY

Champion public discussion and evidence-based policy.

TAKE ACTION

Learn ways you can become an advocate and stand up for science.

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

AAAS

RESEARCH ARTICLE SUMMARY

IMMUNOLOGY

Commensal-specific T cell plasticity promotes rapid tissue adaptation to injury

Oliver J. Harrison, Jonathan L. Linehan, Han-Yu Shih, Nicolas Bouladoux, Seong-Ji Han, Margery Smelkinson, Shurjo K. Sen, Allyson L. Byrd, Michel Enamorado, Chen Yao, Samira Tamoutounour, Francois Van Laethem, Charlotte Hurabielle, Nicholas Collins, Andrea Paun, Rosalba Salcedo, John J. O'Shea, Yasmine Belkaid*

INTRODUCTION: Barrier tissues are constitutive targets of environmental stressors and are home to a highly diverse microbiota. When the immune system encounters these noninvasive microbes, one possible result is the induction of cognate T cell responses that control various aspects of tissue function, including antimicrobial defense and tissue repair. Given the extraordinary number of antigens expressed by the microbiota, a substantial fraction of barrier tissue–resident T cells are expected to be commensal-specific, accumulating over time in response to successive exposure to new commensals. Because barrier tissues are defined by the constitutive coexistence of commensals and commensal-reactive lymphocytes, any understanding of tissue homeostasis, response to injury, and tissue-specific pathologies must occur in the context of this fundamental dialog.

RATIONALE: The skin serves as a primary interface with the environment and is consequently a constitutive target of environmental stressors mediated by physical damage or invasive pathogens. Tissue protection from these challenges relies on rapid and coordinated local responses tailored to both the microenvironment and the nature of the instigating injury. Our study explored whether commensal-specific T cells can act as tissue sentinels, allowing rapid adaptation to defined injuries, and how dysregulation of these responses may have pathogenic consequences.

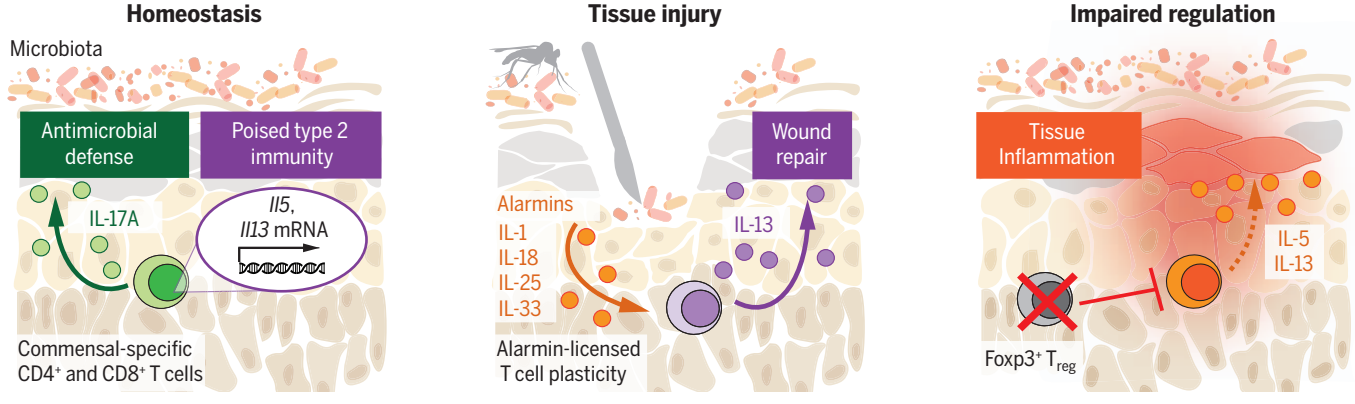
RESULTS: Homeostatic encounters with commensal microbes promoted the induction of commensal-specific interleukin-17A (IL-17A)–producing T cells [$CD4^+$ (T_H17) and $CD8^+$ (T_C17)] that persisted as tissue-resident memory cells. Surprisingly, commensal-specific T cells were characterized by coexpression of classically antagonistic transcription factors (RORγt and GATA-3) that control the respective expression of type 17 and type 2 programs. Consequently, commensal-specific T cells displayed a hybrid chromatin landscape that underlies the coexpression of a broad type 2 transcriptome, including the type 2 effector cytokines IL-5 and IL-13. Notably, during homeostasis, $ROR\gamma t^+$ T cells expressed type 2 cytokine mRNA without subsequent protein translation. By contrast, in the context of tissue challenges such as chitin injection or insect bites, commensal-specific $ROR\gamma t^+$ T cells were able to produce type 2 cytokines (IL-5 and IL-13). The spontaneous release of type 2 cytokines by these cells was also observed in the context of local defects in immune regulation associated with impaired regulatory T cell function. Alarmins associated with tissue damage and inflammation, such as IL-1, IL-18, IL-25, and IL-33, were able to superimpose a type 2 effector program on both T_C17 and T_H17 cells in the context of T cell receptor engagement. Using an IL-17A fate-mapping strategy, we found that IL-17A–committed $ROR\gamma t^+$ T cells and their IL-17A$^-$ $ROR\gamma t^+$ counterparts both produced type 2 cytokines in response to tissue alarmins. Such cellular plasticity allows commensal-specific type 17 cells to promote IL-17A–mediated antimicrobial defense under homeostatic conditions, as well as tissue repair in an IL-13–dependent manner in the context of tissue injury.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aat6280

CONCLUSION: Our work describes a tissue checkpoint that relies on the remarkable plasticity and adaptability of tissue-resident commensal-specific T cells. We propose that this feature may also have important implications in the etiology of tissue-specific inflammatory disorders. The extraordinary number of both commensal-derived antigens and T cells at barrier sites suggests that the ability of commensal-specific T cells to functionally adapt to injury may play a fundamental role in controlling tissue physiology. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: ybelkaid@niaid.nih.gov
Cite this article as O. J. Harrison *et al.*, *Science* **363**, eaat6280 (2019). DOI: 10.1126/science.aat6280



Poised type 2 immunity of commensal-specific T cells promotes rapid adaptation to tissue injury. Commensal-specific T cells produce IL-17A under homeostatic conditions for antimicrobial defense while harboring a poised type 2 transcriptome. Tissue injury licenses type 2 immune potential of commensal-specific type 17 T cells, thereby promoting tissue repair. Impaired immune regulation unleashes type 2 cytokine production from commensal-specific $CD8^+$ T_C17 cells.

RESEARCH ARTICLE

IMMUNOLOGY

Commensal-specific T cell plasticity promotes rapid tissue adaptation to injury

Oliver J. Harrison[1], Jonathan L. Linehan[1]*, Han-Yu Shih[2], Nicolas Bouladoux[1,3], Seong-Ji Han[1], Margery Smelkinson[4], Shurjo K. Sen[5], Allyson L. Byrd[1]*, Michel Enamorado[1], Chen Yao[2], Samira Tamoutounour[1], Francois Van Laethem[6]†, Charlotte Hurabielle[1,7], Nicholas Collins[1], Andrea Paun[8], Rosalba Salcedo[9], John J. O'Shea[2], Yasmine Belkaid[1,3]‡

Barrier tissues are primary targets of environmental stressors and are home to the largest number of antigen-experienced lymphocytes in the body, including commensal-specific T cells. We found that skin-resident commensal-specific T cells harbor a paradoxical program characterized by a type 17 program associated with a poised type 2 state. Thus, in the context of injury and exposure to inflammatory mediators such as interleukin-18, these cells rapidly release type 2 cytokines, thereby acquiring contextual functions. Such acquisition of a type 2 effector program promotes tissue repair. Aberrant type 2 responses can also be unleashed in the context of local defects in immunoregulation. Thus, commensal-specific T cells co-opt tissue residency and cell-intrinsic flexibility as a means to promote both local immunity and tissue adaptation to injury.

Barrier tissues are constitutive targets of environmental stressors as well as primary sites of exposure to symbiotic and pathogenic microbes. As such, under homeostasis, barrier tissues are home to vast numbers of antigen-experienced lymphocytes. The numerous and diverse microbes that colonize these tissues, referred to as the microbiota, play a fundamental role in the induction and quality of these local immune responses, including those that are directed at the microbiota itself (*1*–*4*). Indeed, far from being ignored, microbes at all barrier surfaces are actively recognized by the immune system. Encounters with noninvasive symbionts can lead to the induction of cognate T cell responses (*1*–*4*). This tonic recognition promotes a highly physiological form of adaptive immunity that can control distinct aspects of tissue function, including antimicrobial defense and tissue repair (*5*, *6*). Because of the extraordinary number of antigens expressed by the microbiota, a substantial fraction of barrier tissue-resident T cells are expected to be commensal-specific, accumulating over time in response to successive exposure to new commensals. This understanding of host-microbiota interactions has important implications for our understanding of host immunity and pathologies. Because barrier tissues are defined by the constitutive coexistence of commensals (and associated antigens) and commensal-reactive lymphocytes, our understanding of tissue homeostasis, response to injury, and tissue-specific pathologies must occur in the context of this fundamental dialog.

The skin serves as a primary interface with the environment and is consequently a constitutive target of environmental stressors mediated by physical damage, invasive pathogens, impaired immune regulation, or the nutritional state of the host. Tissue protection from these challenges relies on rapid and coordinated local responses tailored to both the microenvironment and the nature of the instigating injury. Recently, the discovery that cells such as innate lymphoid cells (ILCs) can rapidly respond to mediators released during tissue damage has provided a framework to begin to understand this phenomenon. Whether tissue-resident T cells, particularly those specific to commensals, can also act as tissue sentinels allowing rapid adaptation to defined injury remains unknown. Here, we explored the unique features of commensal-specific T cells and how their distinct wiring might promote physiological or pathological tissue adaptation.

Acute injury licenses type 2 cytokine production from commensal-specific type 17 T cells

The skin is home to a number of resident lymphocytes, some of which recognize the microbiota (*4*, *6*–*8*). We first assessed whether commensal-specific T cells could develop as nonrecirculating tissue-resident memory cells (T_{RM}), a subset of memory T cells previously shown to accumulate in tissues upon pathogen encounter and promote local immunity (*9*). *Staphylococcus epidermidis* colonization of the skin promotes the noninflammatory accumulation of both $CD4^+$ [T helper 1 (T_H1) and T_H17] and $CD8^+$ T cells [T cytotoxic 1 (T_C1) and T_C17] (*4*). A large fraction (>80%) of these *S. epidermidis*–specific polyclonal $CD8^+$ T cells are nonclassically restricted (*6*). *S. epidermidis*–specific $CD8^+$ T cells can be tracked via the use of a peptide–major histocompatibility complex (MHC) tetramer (f-MIIINA:H2-M3) (*6*) and newly generated T cell receptor (TCR)–transgenic mice ($Bowie^{Tg}$). Both tools recapitulate the *S. epidermidis*–specific polyclonal $CD8^+$ T cell response, including cytokine potential, skin-homing, and distribution of the tissue residency markers CD69 and CD103 (*9*) (Fig. 1, A to C). To assess tissue residency, we generated *S. epidermidis*–colonized parabiotic mice, which establish chimerism through joint circulation (*10*) (fig. S1A). In contrast to lymphoid organs, where cells equilibrated, f-MIIINA:H2-M3$^+$ $CD8^+$ T cells within the skin were host-derived (97.1 ± 2.4%) and coexpressed CD103 and CD69 (Fig. 1, D and E). Thus, commensal-specific T cells can develop as long-lived tissue-resident memory T cells.

Given the fundamental role of the skin as a protective barrier, we sought to determine the impact of environmental stressors on commensal-specific tissue-resident T cells. After colonization, *S. epidermidis*–specific polyclonal $CD8^+$ T cells were identified as T-bet$^+$CCR6$^-$ T_C1 cells or RORγt$^+$CCR6$^+$ T_C17 cells [of which ~30% have interleukin (IL)–17A production potential] (Fig. 1, F and G). Although the intradermal injection of chitin or sand fly (*Lutzomyia longipalpis*) bites had no impact on the potential for IL-17A and interferon (IFN)–γ production by T_C17 and T_C1 cells, respectively (Fig. 1H), both stressors revealed a surprising potential for the production of IL-5 and IL-13 from *S. epidermidis*–elicited T_C17 cells, including f-MIIINA:H2-M3$^+$ $CD8^+$ T cells (Fig. 1, H and I, and fig. S1, B and C). Increased type 2 cytokine production after chitin or sand fly challenge was also observed from RORγt-expressing $CD4^+$ T cells (T_H17) elicited by *S. epidermidis* (fig. S1D). Thus, RORγt$^+$ T cells (both $CD8^+$ and $CD4^+$ T cells) elicited by encounter with a commensal may have the unexpected potential to produce type 2 cytokines in response to defined tissue challenges.

[1]Mucosal Immunology Section, Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, Bethesda, MD 20892, USA. [2]Molecular Immunology and Inflammation Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, Bethesda, MD 20892, USA. [3]NIAID Microbiome Program, National Institute of Allergy and Infectious Diseases, Bethesda, MD 20892, USA. [4]Biological Imaging, Research Technology Branch, National Institute of Allergy and Infectious Diseases, Bethesda, MD 20892, USA. [5]Leidos Biomedical Research Inc., Basic Science Program, Cancer and Inflammation Program, Frederick National Laboratory for Cancer Research, Bethesda, MD 20892, USA. [6]Experimental Immunology Branch, National Cancer Institute, Bethesda, MD 20892, USA. [7]Inserm Unité 976, Hôpital Saint-Louis, Paris, France. [8]Intracellular Parasite Biology Section, Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, Bethesda, MD 20892, USA. [9]Cancer and Inflammation Program, National Cancer Institute, Bethesda, MD 20892, USA.
*Present address: Department of Cancer Immunology, Genentech, South San Francisco, CA 94080, USA. †Present address: Institut de Génétique Moléculaire de Montpellier, University of Montpellier, CNRS, Montpellier, France. ‡**Corresponding author. Email: ybelkaid@niaid.nih.gov**

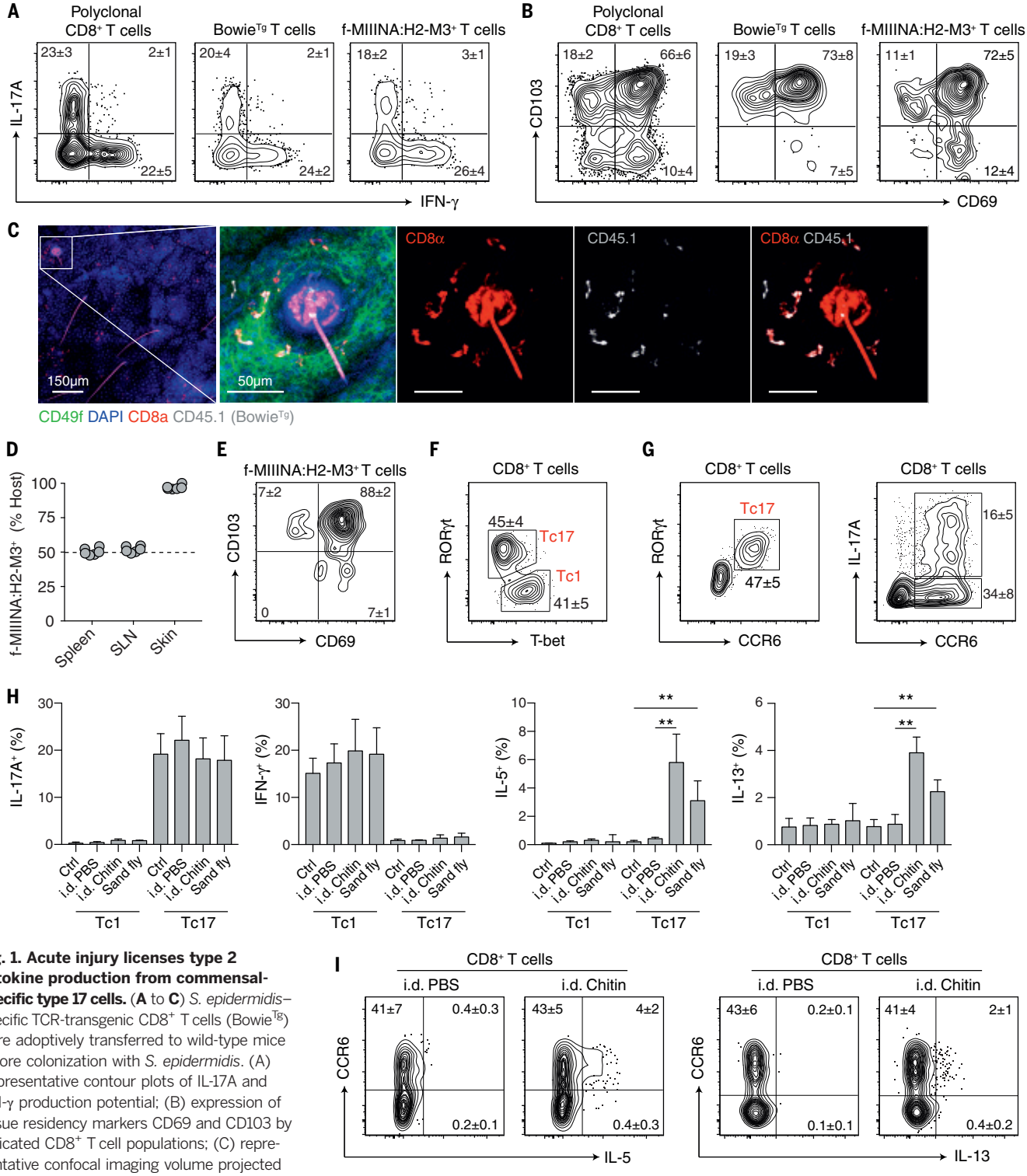


Fig. 1. Acute injury licenses type 2 cytokine production from commensal-specific type 17 cells. (**A** to **C**) *S. epidermidis*–specific TCR-transgenic CD8$^+$ T cells (Bowie^{Tg}) were adoptively transferred to wild-type mice before colonization with *S. epidermidis*. (A) Representative contour plots of IL-17A and IFN-γ production potential; (B) expression of tissue residency markers CD69 and CD103 by indicated CD8$^+$ T cell populations; (C) representative confocal imaging volume projected along the *z* axis of epidermal skin from *S. epidermidis*–colonized mice. (**D** and **E**) Conjoined pairs of *S. epidermidis*–colonized CD45.1 and CD45.2 mice were analyzed 90 days after parabiosis surgery for cellular origin and phenotype. (D) Frequency of host-derived f-MIIINA:H2-M3$^+$ CD8$^+$ T cells in indicated tissues; SLN, skin-draining lymph nodes. (E) Representative contour plot of CD69 and CD103 expression by skin f-MIIINA:H2-M3$^+$ CD8$^+$ T cells. (**F**) Representative contour plot of RORγt and T-bet expression by CD8$^+$ T cells from the skin of *S. epidermidis*–colonized wild-type mice. (**G**) Representative contour plots of RORγt, CCR6, and IL-17A expression by CD8$^+$ T cells from the skin of *S. epidermidis*–colonized wild-type mice. (**H** and **I**) *S. epidermidis*–colonized wild-type mice were exposed to bites from sand flies (*L. longipalpis*) or injected intradermally (i.d.) with PBS or chitin. (H) Frequencies of T_C1 and T_C17 cells with cytokine-producing potential from the skin of *S. epidermidis*–colonized wild-type mice after skin injury. (I) Representative contour plots of IL-5 and IL-13 production potential by CD8$^+$ T cells from the skin of *S. epidermidis*–colonized wild-type mice after skin injury. Numbers in representative plots indicate means ± SD. Bar graphs show means ± SD. Data represent at least two experiments with four to six mice per group. $^{**}P < 0.01$ (one-way ANOVA with Holm-Šidák multiple-comparison test).

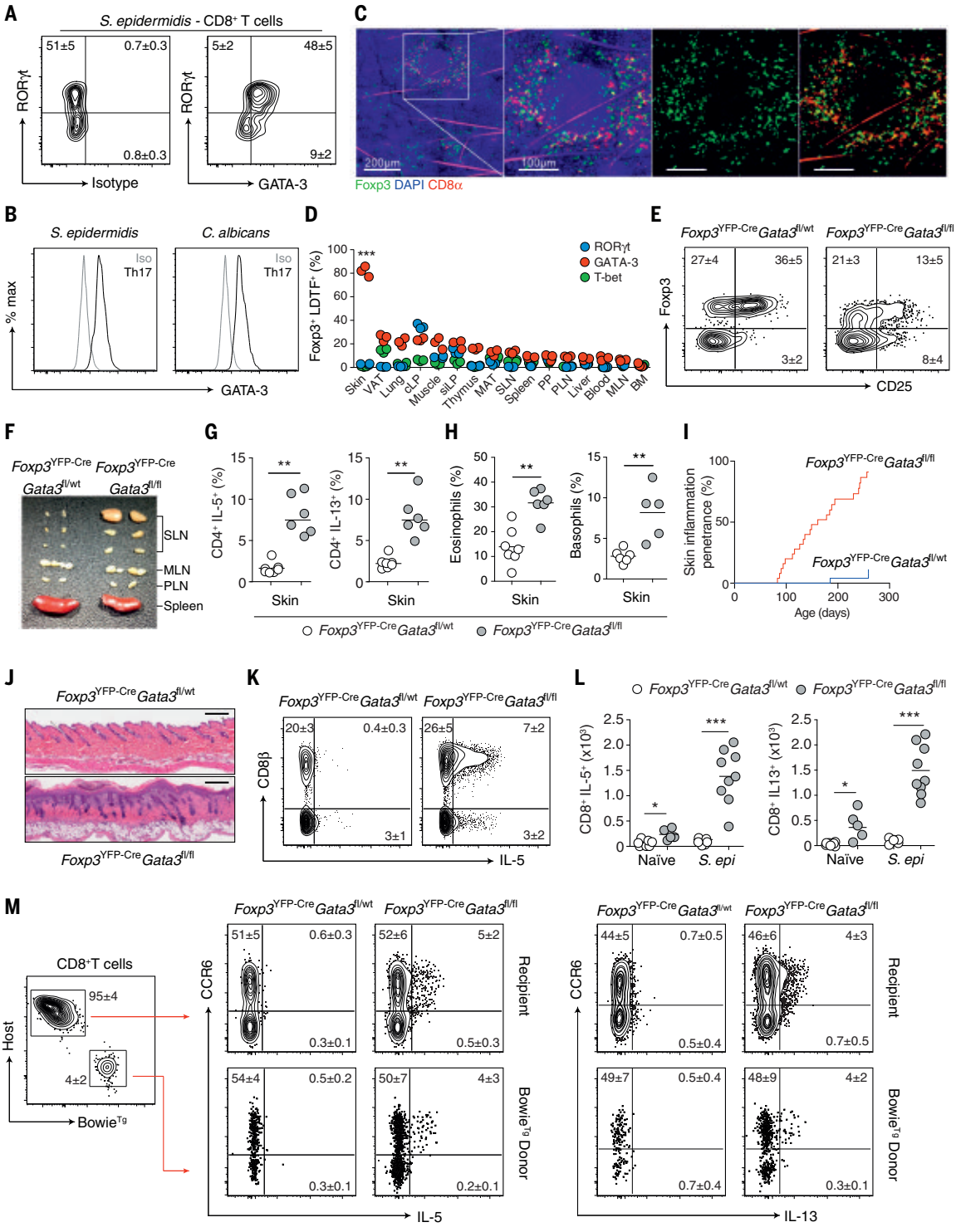


Fig. 2. Local defects in immunoregulation unleash type 2 immunity from commensal-specific T cells. (**A**) Representative contour plots of RORγt and GATA-3 expression by skin CD8$^+$ T cells from *S. epidermidis*–colonized wild-type mice. (**B**) Representative histogram plots of GATA-3 expression by RORγt$^+$ CD4$^+$ T_H17 cells from the skin of commensal-colonized wild-type mice. (**C**) Representative confocal imaging volume-projected along the *z* axis of epidermal skin from *S. epidermidis*–colonized *Foxp3*^{gfp} mice. (**D**) Frequencies of Foxp3$^+$ T_{reg} cells coexpressing lineage-defining transcription factors (LDTFs) within indicated tissues of naïve wild-type mice. VAT, visceral adipose tissue; cLP, colonic lamina propria; siLP, small intestinal lamina propria; MAT, mesenteric adipose tissue; SLN, skin-draining lymph node; PP, Peyer's patch; PLN, para-aortic lymph node; MLN, mesenteric lymph node; BM, bone marrow. (**E**) Representative contour plots of Foxp3 and CD25 expression by skin CD4$^+$ T cells from naïve *Foxp3*$^{YFP\text{-}Cre}$*Gata3*$^{fl/wt}$ and *Foxp3*$^{YFP\text{-}Cre}$*Gata3*$^{fl/fl}$ mice. (**F**) Representative cutaneous lymphadenopathy in *Foxp3*$^{YFP\text{-}Cre}$*Gata3*$^{fl/fl}$ compared to *Foxp3*$^{YFP\text{-}Cre}$*Gata3*$^{fl/wt}$ control mice. (**G**) Frequencies of IL-5– and IL-13–producing skin CD4$^+$ T cells from naïve *Foxp3*$^{YFP\text{-}Cre}$*Gata3*$^{fl/wt}$ and *Foxp3*$^{YFP\text{-}Cre}$*Gata3*$^{fl/fl}$ mice. (**H**) Frequencies of skin eosinophils and basophils from naïve *Foxp3*$^{YFP\text{-}Cre}$*Gata3*$^{fl/wt}$ and *Foxp3*$^{YFP\text{-}Cre}$*Gata3*$^{fl/fl}$ mice. (**I**) Cumulative incidence of skin inflammation among naïve *Foxp3*$^{YFP\text{-}Cre}$*Gata3*$^{fl/wt}$ and *Foxp3*$^{YFP\text{-}Cre}$*Gata3*$^{fl/fl}$ mice. (**J**) Representative histological micrograph of skin tissue from naïve *Foxp3*$^{YFP\text{-}Cre}$*Gata3*$^{fl/wt}$ and *Foxp3*$^{YFP\text{-}Cre}$*Gata3*$^{fl/fl}$ mice. Scale bars, 250 μm. (**K**) Representative contour plots of CD8β expression and IL-5 production potential by TCRβ$^+$ T cells from the skin of *S. epidermidis*–colonized *Foxp3*$^{YFP\text{-}Cre}$*Gata3*$^{fl/wt}$ and *Foxp3*$^{YFP\text{-}Cre}$*Gata3*$^{fl/fl}$ mice. (**L**) Total numbers of IL-5– and IL-13–producing CD8$^+$ T cells from the skin of *S. epidermidis*–colonized *Foxp3*$^{YFP\text{-}Cre}$*Gata3*$^{fl/wt}$ and *Foxp3*$^{YFP\text{-}Cre}$*Gata3*$^{fl/fl}$ mice. (**M**) Representative contour plots of IL-5 and IL-13 production by CD8$^+$ T cells in *Foxp3*$^{YFP\text{-}Cre}$*Gata3*$^{fl/wt}$ and *Foxp3*$^{YFP\text{-}Cre}$*Gata3*$^{fl/fl}$ mice adoptively transferred with Bowie^{Tg} T cells before colonization with *S. epidermidis*. Numbers in representative plots indicate means ± SD. Each dot represents an individual mouse. Data represent at least two experiments with three to seven mice per group. Cumulative skin inflammation data (I) represent 25 mice per genotype. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ as calculated using Student *t* test [(G), (H)] or one-way ANOVA with Holm-Šidák multiple comparison test [(D), (L)].

Local defects in immunoregulation unleash type 2 immunity from commensal-specific T cells

Flow cytometric analysis revealed that T_C17 cells coexpressed GATA-3, the lineage-defining transcription factor (LDTF) for both T_H2 cells and group 2 ILC (ILC2) (Fig. 2A). Such a phenotype was also detected among the very few $CD8^+$ T cells present in the skin of naïve mice (fig. S2A), and coexpression of RORγt and GATA-3 by *S. epidermidis*–specific $Bowie^{Tg}$ $CD8^+$ T cells was restricted to the skin and not detectable in secondary lymphoid organs; these findings suggested that GATA-3 expression is imprinted within the tissue microenvironment (fig. S2B). This phenotype was conserved across T cell lineages and distinct microbial exposures. Notably, T_H17 cells elicited by skin colonization with *S. epidermidis* or *Candida albicans* also expressed GATA-3 (Fig. 2B and fig. S2C). Thus, homeostatic encounter with bacterial or fungal commensal microbes can lead to the development of cells with a paradoxical phenotype characterized by the coexpression of classically antagonistic transcription factors.

The skin is highly enriched in $Foxp3^+$ regulatory T (T_{reg}) cells (*5*), and confocal imaging revealed colocalization of *S. epidermidis*–induced $CD8^+$ T cells and $Foxp3^+$ T_{reg} cells (Fig. 2C). As such, we assessed the possibility that skin $Foxp3^+$ T_{reg} cells could limit type 2 cytokine production by commensal-specific type 17 cells. Because complete ablation of $Foxp3^+$ T_{reg} cells results in severe local and systemic inflammatory responses and aberrant accumulation of T_C1 cells within the skin (*11*) (fig. S2, D and E), we used an approach allowing for a tissue-specific defect in immunoregulation. Within the skin, T_{reg} cells express high levels of GATA-3 (but not other LDTFs) (Fig. 2D and fig. S2F), a factor that contributes to T_{reg} cell stability and fitness (*12*–*15*). In mice in which T_{reg} cells were conditionally deleted of GATA-3 ($Foxp3^{YFP\text{-}Cre}Gata3^{fl/fl}$), $Foxp3^+$ cells were reduced in frequency and exhibited decreased Foxp3 and CD25 expression in the skin, but not in other tissues (Fig. 2E and fig. S2, G and H). Consistent with this observation, by 10 weeks of age, skin-draining lymph nodes (but not other lymphoid structures) were enlarged, and the skin compartments (but not other tissues) of these mice were characterized by a selective increase in the number of T cells producing IL-5 and IL-13 (Fig. 2, F and G, and fig. S2, I to K). Enhanced type 2 responses were associated with discrete elevated frequencies and absolute numbers of eosinophils and basophils in the skin of $Foxp3^{YFP\text{-}Cre}Gata3^{fl/fl}$ mice relative to control mice (Fig. 2H and fig. S2L). Of note, and in agreement with a skin-specific defect, naïve $Foxp3^{YFP\text{-}Cre}Gata3^{fl/fl}$ mice, with an endogenous skin microbiota but not *S. epidermidis*, spontaneously developed severe skin inflammation (but not systemic inflammation) with ~70% penetrance by 8 months of age (Fig. 2, I and J).

To assess the possibility that T cells producing type 2 cytokines within the skin of these mice are commensal-specific, we colonized young $Foxp3^{YFP\text{-}Cre}Gata3^{fl/fl}$ mice (before inflammation) and control mice with *S. epidermidis* and tracked the fate of *S. epidermidis*–specific T cells. Adoptively transferred $Bowie^{Tg}$ $CD8^+$ T cells (as well as host polyclonal *S. epidermidis*–induced $CD8^+$ T cells) expressed IL-5 and IL-13 proteins in the skin of $Foxp3^{YFP\text{-}Cre}Gata3^{fl/fl}$ mice but not control mice (Fig. 2, K to M). By contrast, the ability of *S. epidermidis*–elicited $CD8^+$ T cells to produce IL-17A or IFN-γ was unaffected (fig. S2M). Notably, type 2 cytokine production by *S. epidermidis*–specific polyclonal and adoptively transferred $Bowie^{Tg}$ $CD8^+$ T cells remained restricted to cells

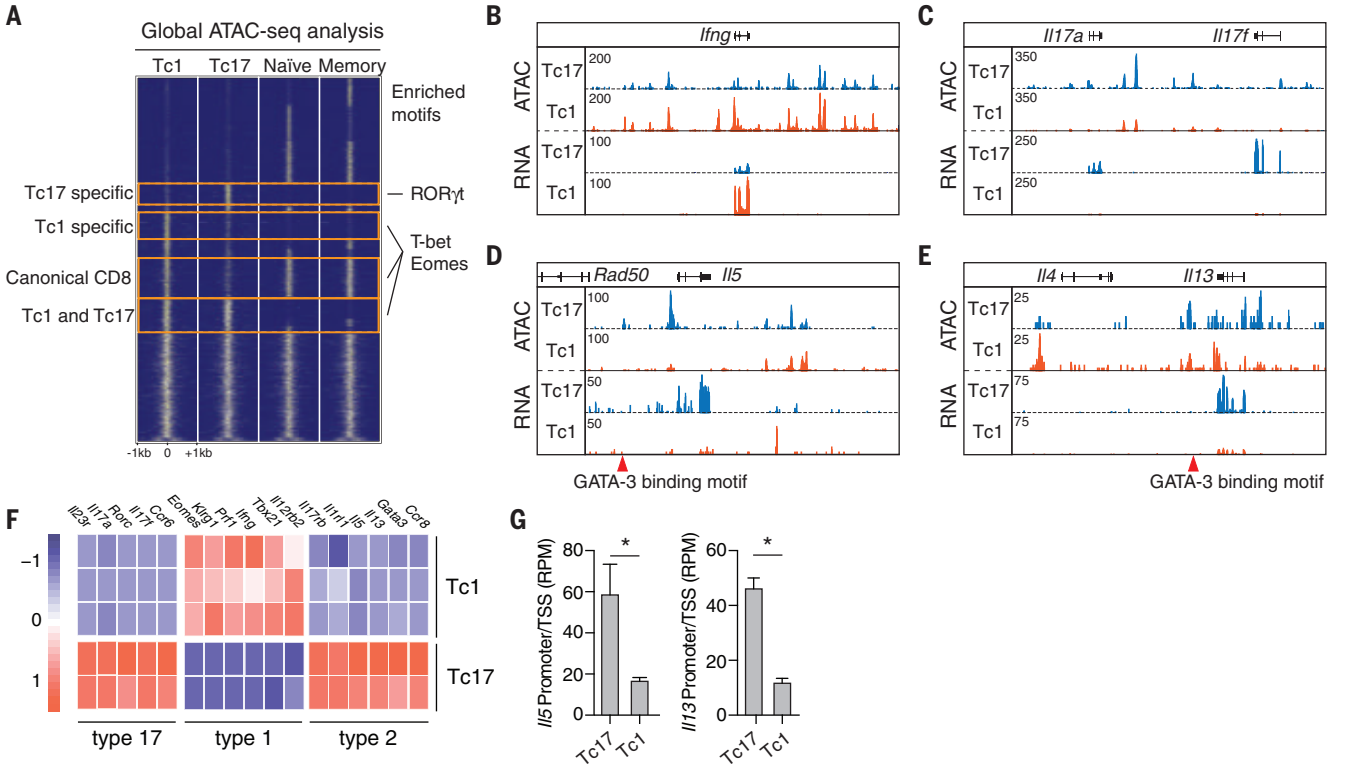


Fig. 3. *S. epidermidis*–specific T_C17 cells express a broad type 2 signature. T_C17 ($CD8^+CCR6^+$) and T_C1 ($CD8^+CCR6^-$) cells were isolated by FACS from the skin of *S. epidermidis*–colonized wild-type mice for transcriptomic and epigenetic analysis by RNA-seq and ATAC-seq. ATAC-seq signals from canonical naïve and memory $CD8^+$ T cells were from lymphoid tissue. (**A**) Global comparison of ATAC-seq signals in *S. epidermidis*–induced T_C17 and T_C1 and canonical naïve and memory $CD8^+$ T cells. Representative transcription factor binding motifs enriched in indicated groups are listed on the right. (**B** to **E**) Genomic tracks of ATAC-seq and RNA-seq signal profiles in T_C17 and T_C1 cells across signature cytokine genes. Genomic location of T_C17-specific regulatory elements with GATA-3 binding motifs are denoted by red triangles. (**F**) Heat map of lineage-specific signature genes expressed by T_C17 and T_C1 populations. (**G**) Chromatin accessibility at transcription start site (promoter ± 500 bp) of *Il5* and *Il13* in T_C17 and T_C1 cells. Bar graphs show means ± SD. Sequencing data represent two or three independent biological replicates. *$P < 0.05$ (Student *t* test).

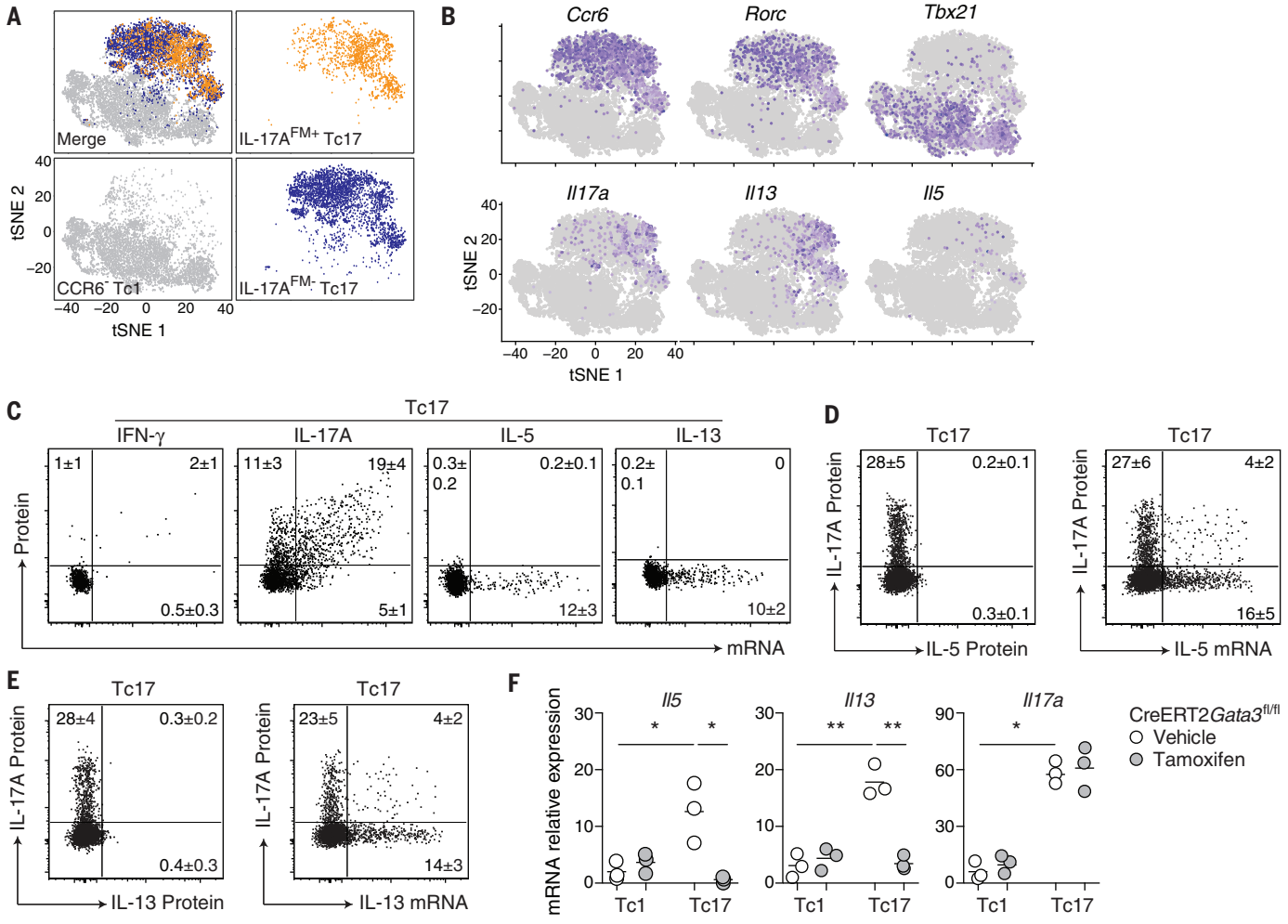


Fig. 4. *S. epidermidis*–specific T_C17 cells harbor a poised type 2 transcriptome. (**A** and **B**) T_C1 ($CD8^+CCR6^-$), IL-$17A^{FM+}$ T_C17 ($CD8^+CCR6^+eYFP^+$), and IL-$17A^{FM-}$ T_C17 ($CD8^+CCR6^+eYFP^-$) cells were isolated by FACS from the skin of *S. epidermidis*–colonized *Il17a*^{Cre}*R26R*^{eYFP} (IL-$17A^{FM}$) mice and analyzed by scRNA-seq. (A) tSNE plots of scRNA-seq expression highlighting T_C1 (gray), IL-$17A^{FM+}$ T_C17 (orange), and IL-$17A^{FM-}$ T_C17 (blue) populations. (B) Expression of LDTFs and cytokine genes projected onto a tSNE plot. (**C**) Representative dot plots of cytokine protein production potential and mRNA expression by T_C17 cells from the skin of *S. epidermidis*–colonized wild-type mice. (**D** and **E**) Representative dot plots of IL-17A and IL-5 or IL-13 production potential and *Il5* or *Il13* mRNA expression by T_C17 cells from the skin of *S. epidermidis*–colonized wild-type mice. (**F**) *S. epidermidis*–colonized CreERT2*Gata3*$^{fl/fl}$ mice received tamoxifen or vehicle control before cell sorting of skin T_C1 and T_C17 cells. Gene expression in the indicated populations was assessed by quantitative reverse transcription polymerase chain reaction (qRT-PCR). Numbers in representative plots indicate means ± SD. Flow cytometric data represent at least two experiments with four to six mice per group. qRT-PCR data represent three biological replicates of eight pooled mice per group. *$P < 0.05$, **$P < 0.01$ (one-way ANOVA with Holm-Šidák multiple-comparison test).

expressing CCR6; this result shows that in the context of local immune defects, type 2 cytokines can be unleashed from RORγt-committed T cells (Fig. 2M). Thus, impaired local immunoregulation promotes type 2 cytokine production by commensal-specific type 17 cells—a property that may predispose tissue to inflammation.

S. epidermidis–specific T_C17 cells harbor a poised type 2 transcriptome

To gain insight into the transcriptional and epigenetic landscape of commensal-specific T cells under homeostatic conditions, we identified global regulatory elements shared between, and unique to, *S. epidermidis*–specific polyclonal T_C17 ($CCR6^+$) and T_C1 ($CCR6^-$) cells from the skin and naïve and memory $CD8^+$ T cells from lymphoid tissue (*16*). Regulatory elements unique to skin T_C17 or T_C1 cells were enriched in binding sites for RORγt, and for T-bet and Eomes, respectively (Fig. 3A), consistent with subset-specific expression of these LDTFs (Fig. 1F). Elevated chromatin accessibility and transcript abundance of the signature cytokines *Ifng*, *Il17a*, and *Il17f* also confirmed the clear distinction between T_C1 and T_C17 cell subsets (Fig. 3, B and C, and fig. S3, A and B). Among regulatory elements unique to T_C17 cells, we identified previously described GATA-3–binding sites within *Il13* and the *Rad50*/T_H2 locus control region (*17*) (Fig. 3, D and E). Consequently, T_C17 cells demonstrated elevated chromatin accessibility at type 2 immune gene loci encoding *Il5* and *Il13* and expressed elevated levels of *Il5* and *Il13* mRNA transcripts relative to T_C1 cells (Fig. 3, D to G, and fig. S3C). Furthermore, T_C17 cells expressed a broad type 2 transcriptome, including a LDTF (*Gata3*) and cytokine and chemokine receptors (*Ccr8*, *Il1rl1*, and *Il17rb*), but neither *Il4* nor *Il10* mRNA, as previously described for tissue-derived T_H2 cells (*18*) (Fig. 3F and fig. S3D). The type 2–associated cytokine amphiregulin (*Areg*) was detectable in both cell subsets, albeit at higher abundance in T_C17 cells (fig. S3D). As such, commensal-specific T_C17 cells express a broad type 2 transcriptome under homeostatic conditions.

Of *S. epidermidis*–induced T_C17 cells, ~30% displayed the potential for IL-17A production (Fig. 1G), supporting the idea of possible phenotypic heterogeneity. However, using cells from IL-17A fate-mapping mice (IL-$17A^{FM}$–*Il17a*^{Cre}*R26R*^{eYFP})

and single-cell RNA sequencing (scRNA-seq), t-distributed stochastic neighbor embedding (tSNE) projection of T_C1, IL-17A^{FM+} T_C17, and IL-17A^{FM-} T_C17 cells demonstrated considerable transcriptional overlap between IL-17A^{FM+} and IL-17A^{FM-} T_C17 cell fractions, with type 2 cytokine mRNA–expressing cells present in both fractions (Fig. 4, A and B, and fig. S4A). Thus, commensal-specific T_C17 cells, including those already committed to IL-17A production, can be superimposed with the expression of a type 2 transcriptome. Furthermore, in situ hybridization for mRNA detection by flow cytometry revealed that *Il5* and *Il13* transcripts, but not protein, were expressed selectively by T_C17 cells from the skin of *S. epidermidis*–colonized mice (Fig. 4C and fig. S4B). In line with our scRNA-seq data (Fig. 4B and fig. S4A), *Il5*$^+$ and *Il13*$^+$ cells were found within both IL-17A–producing and IL-17A–nonproducing fractions of T_C17 cells (Fig. 4, D and E); this suggests that during homeostasis, commensal-specific T_C17 cells express type 2 cytokine mRNA without subsequent protein translation. The inducible deletion of *Gata3* at the peak of the CD8$^+$ T cell response to *S. epidermidis* revealed that sustained GATA-3 expression by T_C17 cells was required for the constitutive expression of *Il5* and *Il13*, but, as expected, not for *Il17a* (Fig. 4F). Thus, *S. epidermidis*–specific T_C17 cells express a poised type 2 transcriptome dependent on continued GATA-3 expression.

Accordingly, type 2 cytokine competency (mRNA expression) and licensing (stimuli-induced protein production) are temporally decoupled in *S. epidermidis*–elicited T_C17 cells—a process likely involving the posttranscriptional regulation of cytokine mRNA stability and protein translation. Under inflammatory conditions, previous work revealed that distinct stimuli can govern competency and licensing of type 2 immunity within injured tissues, ensuring tissue-restricted effector function during pathogen infection (*19*). Recent findings also suggest that IFN-γ production by CD8$^+$ T cells is actively regulated at the level of translation, thereby preventing chronic immune activation (*20*–*22*).

Alarmins license type 2 cytokine production by commensal-specific T_C17 cells

Our work proposes that such a phenomenon may also apply to commensal-specific T cells generated under homeostatic conditions. To identify the factors capable of licensing poised type 2 immunity from commensal-specific T cells, we used an ex vivo screening approach, stimulating T_C17 and T_C1 cells with cytokines and alarmins previously shown to be associated with tissue damage. Cytokine stimulation alone did not promote type 2 cytokine production by skin T cells, demonstrating that these cells cannot be licensed in a TCR-independent manner (fig. S5A). Because commensal microbes persist within the skin, this result is consistent with the expectation that exposure to alarmins will occur in the context of antigen exposure. However, in line with the role of IL-1 within the skin (*4*), IL-1α significantly increased the ex vivo production of IL-17A from T_C17 cells in the context of TCR stimulation (Fig. 5A). As previously reported, IL-18 and IL-33 promoted IFN-γ production by T_C1 cells (*23*, *24*) (Fig. 5B). Notably, several alarmins promoted the production of IL-5 (IL-18, IL-25, and IL-33) or IL-13 (IL-1α, IL-1β, IL-18, and IL-33) (Fig. 5, C and D). IL-25 potently promoted the production of IL-5 but not IL-13 (Fig. 5, C and D), supporting the idea that distinct classes of injury may have different impacts on commensal-specific T cell responses. Strikingly, IL-18, a cytokine widely linked to the initiation of type 1 responses, was particularly potent at eliciting the release of both IL-5 and IL-13 from T_C17 cells ex vivo (Fig. 5, C and D). IL-18 also promoted IL-17A production by T_C17 cells, further supporting the idea that this alarmin can superimpose type 2 responses upon a precommitted type 17 program (Fig. 5A). Under these conditions, IL-4 and IL-10 were undetectable (fig. S5B), but both T_C1 and T_C17 cells produced amphiregulin upon TCR stimulation, a response that was also enhanced by IL-18 (fig. S5C). Type 2 responses to IL-18 were not restricted to CD8$^+$ T cells nor to *S. epidermidis*–elicited cells. Indeed, skin CD4$^+$ T cells induced by *S. epidermidis* or *C. albicans* colonization (including T_H17 cells) also produced higher levels of IL-5 and IL-13 upon IL-18 and TCR stimulation in vitro (fig. S5, D to F). Thus, such poised type 2 potential may be the norm for type 17 commensal-specific T cells raised under homeostatic conditions. In this context, local inflammatory factors including IL-1, IL-18, IL-25, and IL-33 can superimpose a type 2 effector program.

To assess the impact of a single defined alarmin on commensal-specific T cells, we next focused on the impact of IL-18 in vivo. A single injection of IL-18 licensed both IL-5 and IL-13 protein production by *S. epidermidis*–elicited T_C17 (including f-MIIINA:H2-M3$^+$ cells) and CD4$^+$ T cells (including T_H17 cells) (Fig. 5, E to G, and fig. S5, G and H). Type 2 cytokine licensing by IL-18 occurred at the expense of IL-17A production, suggesting dynamic regulation of cytokine production by commensal-specific T_C17 and T_H17 cells in vivo (Fig. 5, E to G). The ability of T_C17 and T_H17 cells to produce type 2 cytokines in response to IL-18 was dependent on T cell–intrinsic IL-18R1 signaling (Fig. 5, H and I) and was sustained up to 60 days after colonization (fig. S5I). After chitin injection, type 2 licensing of T_C17 and T_H17 cells was also IL-18R1 signaling–dependent (Fig. 5, H and I); these findings support the idea that in defined inflammatory settings, IL-18 alone may be sufficient to impose this response.

Commensal-specific T cell plasticity and IL-13 production promote wound repair

The co-production of cytokines associated with distinct T cell subsets can occur during inflammation. For example, IL-17A$^+$IFN-γ$^+$ cells are present during intestinal and central nervous system inflammation, and IL-17A$^+$IL-4$^+$ cells are found during allergic asthma and helminth infection (*25*–*29*). Previous studies also demonstrated plasticity of effector T_H17 cells to convert to T_H1, follicular helper (T_{FH}), and T_{reg} cell phenotypes in a context-dependent manner (*26*, *30*, *31*). Our work supports the idea that such plasticity may be a fundamental feature of tissue-resident commensal-specific T cells. To specifically address this point, we used IL-17A^{FM} mice to assess in vivo the heritage of T_C17 cells licensed for type 2 cytokine production. In line with the finding that both IL-17A^{FM-} and IL-17A^{FM+} T_C17 cells display poised *Il5* and *Il13* mRNA expression (Fig. 4, B to E), IL-18 triggered type 2 cytokine production from both T_C17 and T_H17 cells regardless of whether they had previously expressed IL-17A (IL-17A^{FM+} and IL-17A^{FM-}) (Fig. 6, A and B). Thus, within commensal-induced T_C17 and T_H17 cell populations, plasticity among IL-17A^{FM+} cells and local licensing of IL-17A^{FM-} cells both contribute to alarmin-mediated induction of type 2 cytokine production.

Although a few reports have suggested that IL-18 can potentially promote type 2 and regulatory responses (*32*–*34*), this cytokine is more widely considered to promote type 1 immunity. In support of a major role for IL-18 in the promotion of skin type 2 responses, IL-18 injection promoted type 2 cytokine production not only by T cells but also by ILC2, as recently described (Fig. 6C) (*35*). In contrast to transient ILC2 responses, induction of type 2 cytokine expression by T cells was sustained up to 4 days after injection (Fig. 6C). Thus, type 2 cytokine licensing by IL-18 may have a profound effect on skin physiology via the broad impact of a defined alarmin on both tissue-resident commensal-specific T cells and ILC2 (*35*). Indeed, IL-18 injection promoted an IL-5–dependent eosinophil accumulation within the skin compartment of *S. epidermidis*–colonized mice (Fig. 6D and fig. S6A). Thus, tissue-resident commensal-specific type 17 T cells can adapt to defined injury by direct sensing of alarmins and inflammatory mediators.

Because of the known contribution of type 2 immunity and IL-13 in particular to tissue repair, we next used a model of skin wounding to assess the potential contribution of commensal-specific type 2 cytokine licensing to this fundamental process. Although IL-13 did not contribute to the healing process in unassociated mice, IL-13 neutralization or genetic *Il13* deficiency impaired *S. epidermidis*–accelerated wound repair (Fig. 6, E and F). Adoptive transfer of wild-type Bowie^{Tg} CD8$^+$ T cells rescued this defect in an IL-13–dependent manner (Fig. 6F). In agreement with the role of IL-13 in tissue repair (*36*), whole-tissue RNA-seq of skin after wounding revealed an IL-13–dependent transcriptional signature dominated by pathways associated with muscle contractility and extracellular matrix reorganization (Fig. 6G and fig. S6B). Notably, in line with the fact that punch biopsies can trigger the release of numerous factors able to license type 17 cells (Fig. 5, C and D), IL-18 was insufficient

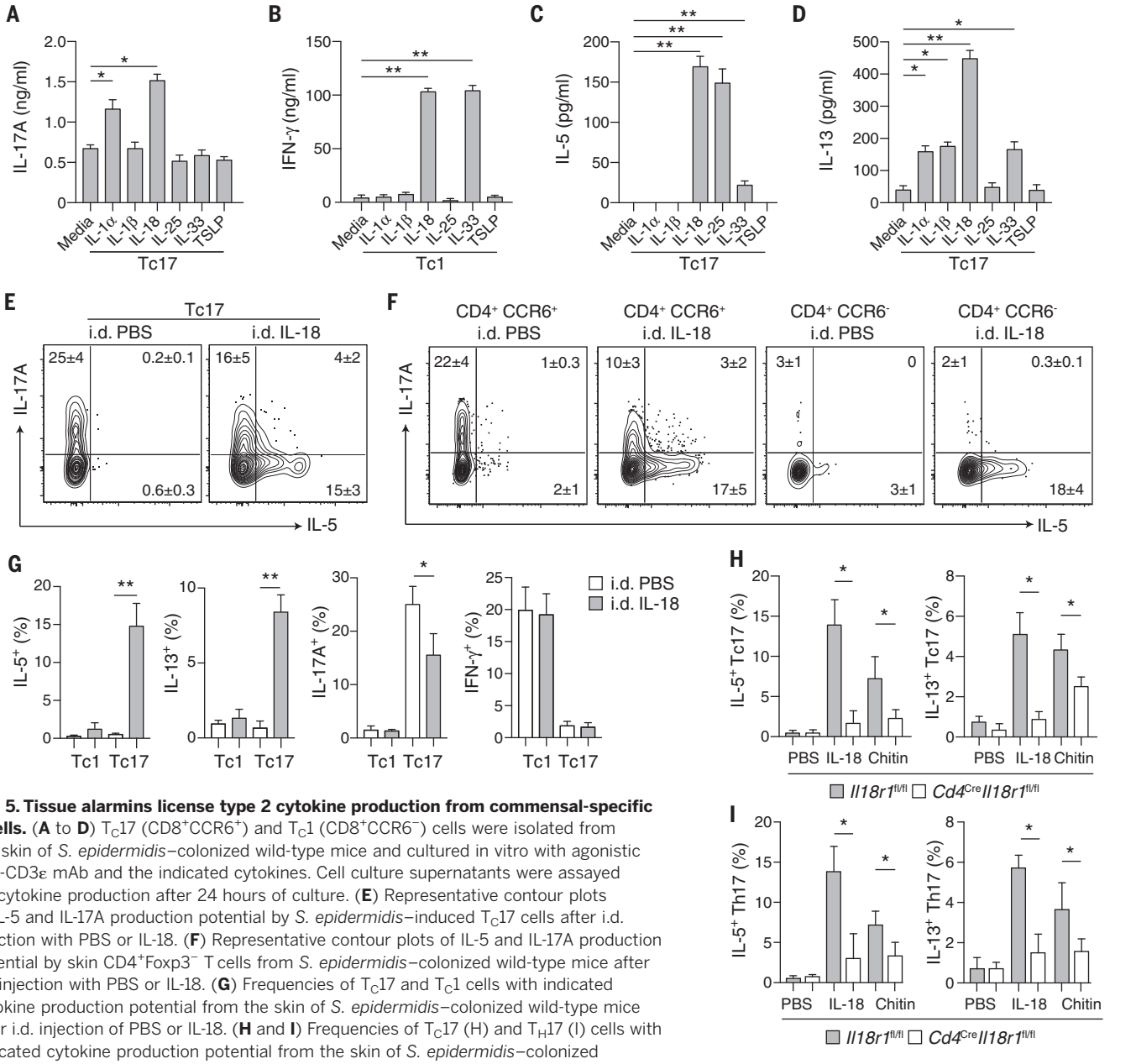


Fig. 5. Tissue alarmins license type 2 cytokine production from commensal-specific T cells. (**A** to **D**) T_C17 ($CD8^+CCR6^+$) and T_C1 ($CD8^+CCR6^-$) cells were isolated from the skin of *S. epidermidis*–colonized wild-type mice and cultured in vitro with agonistic anti-CD3ε mAb and the indicated cytokines. Cell culture supernatants were assayed for cytokine production after 24 hours of culture. (**E**) Representative contour plots of IL-5 and IL-17A production potential by *S. epidermidis*–induced T_C17 cells after i.d. injection with PBS or IL-18. (**F**) Representative contour plots of IL-5 and IL-17A production potential by skin $CD4^+Foxp3^-$ T cells from *S. epidermidis*–colonized wild-type mice after i.d. injection with PBS or IL-18. (**G**) Frequencies of T_C17 and T_C1 cells with indicated cytokine production potential from the skin of *S. epidermidis*–colonized wild-type mice after i.d. injection of PBS or IL-18. (**H** and **I**) Frequencies of T_C17 (H) and T_H17 (I) cells with indicated cytokine production potential from the skin of *S. epidermidis*–colonized *Cd4*^{Cre}*Il18r1*$^{fl/fl}$ and control mice after i.d. injection with PBS, IL-18, or chitin. Numbers in representative plots indicate means ± SD. Bar graphs show means ± SD. Data represent at least two experiments with three to six mice per group. *$P < 0.05$, **$P < 0.01$ as calculated using one-way [(A) to (D), (G)] or two-way [(H), (I)] ANOVA with Holm-Šidák multiple-comparison test.

to promote these responses (fig. S6C). Thus, the poised type 2 immune potential of commensal-specific T_C17 cells allows for local adaptation to injury, thereby promoting tissue repair.

Conclusion

Barrier tissues are constitutively exposed to environmental stressors and are primary targets of chronic inflammatory disorders. The maintenance of tissue integrity and function represent a complex challenge that requires both resilience and adaptation. Under steady-state conditions, tissue resilience is, in part, mediated by innate and adaptive immunity to the microbiota, which reinforces barrier function and immunity (*5*). Our results show that adaptation of tissue to injuries can also be mediated by immunity to the microbiota, a fundamental but poorly understood class of immunity. Notably, we found that homeostatic immunity to bacteria or fungal commensals is characterized by the coexpression of paradoxical programs, allowing commensal-specific T cells, when entering and persisting within tissues, to adopt a type 17 program compatible with tissue homeostasis and immunity while maintaining a type 2–poised state. As such, in the context of injury and consequent exposure to inflammatory mediators and cognate antigens, commensal-specific T cells rapidly release type 2 cytokines, allowing these cells to exert pleiotropic and contextual functions including tissue repair. Thus, we describe a tissue checkpoint that relies on the remarkable plasticity and adaptability of tissue-resident commensal-specific T cells. We propose that this feature may also have important implications in the etiology of tissue-specific inflammatory disorders. Given the extraordinary number of both commensal-derived antigens and T cells at barrier sites, such a feature may represent a fundamental component of host immunity.

Materials and methods

Mice

Wild-type (WT) C57BL/6 Specific Pathogen Free (SPF) mice were purchased from Taconic

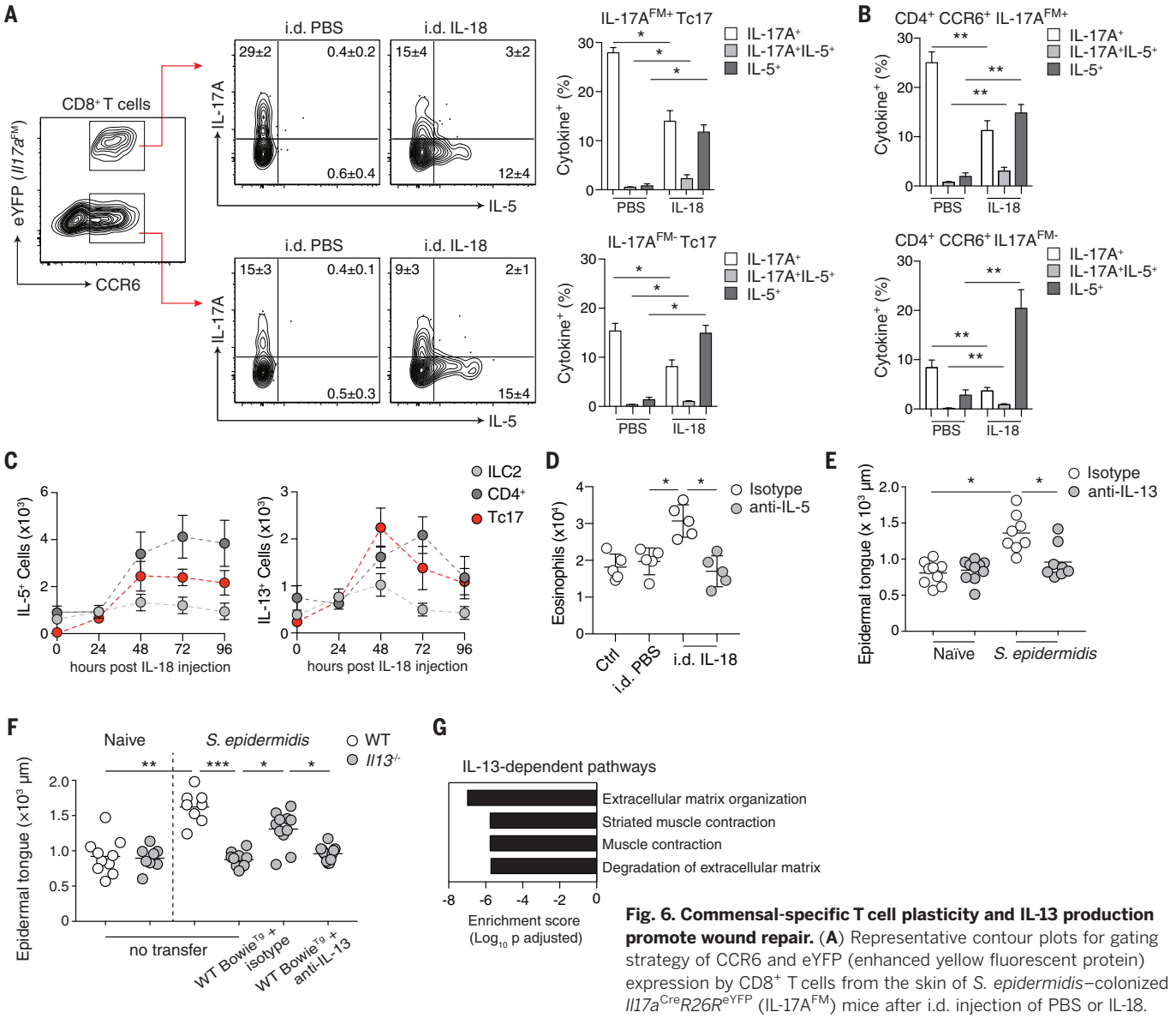


Fig. 6. Commensal-specific T cell plasticity and IL-13 production promote wound repair. (**A**) Representative contour plots for gating strategy of CCR6 and eYFP (enhanced yellow fluorescent protein) expression by $CD8^+$ T cells from the skin of *S. epidermidis*–colonized *Il17a*^{Cre}*R26R*^{eYFP} (IL-17A^{FM}) mice after i.d. injection of PBS or IL-18. Contour plots represent IL-5 and IL-17A production potential of IL-17A^{FM+} T_C17 ($CD8^+CCR6^+eYFP^+$) or IL-17A^{FM-} T_C17 ($CD8^+CCR6^+eYFP^-$) T cells after i.d. injection of PBS or IL-18. (**B**) Frequencies of T_H17 cells with IL-17A– or IL-5–producing potential from the skin of *S. epidermidis*–colonized IL-17A^{FM} mice after i.d. injection of PBS or IL-18. (**C**) Absolute cell number of IL-5– and IL-13–producing lymphocyte subsets in the skin of *S. epidermidis*–colonized wild-type mice after i.d. injection of IL-18. Data are means ± SD of five mice per group. (**D**) Absolute number of eosinophils from the skin of *S. epidermidis*–colonized wild-type mice after i.d. injection with PBS or IL-18, and i.p. injection with anti–IL-5 or isotype control. (**E** and **F**) Naïve and *S. epidermidis*–colonized wild-type and *Il13*$^{-/-}$ mice, with or without adoptive transfer of Bowie^{Tg} $CD8^+$ T cells before colonization and isotype or anti–IL-13 antibodies at the time of wounding, were subjected to back-skin punch biopsy. Epithelial tongue length of wound bed–infiltrating keratinocytes was quantified 5 days after wounding. (**G**) Pathway analysis using differentially expressed genes between d3 isotype and d3 anti–IL-13 wounding groups was performed using Enrichr and graphed according to enrichment score for significant Reactome biological processes. Numbers in representative plots indicate means ± SD. Bar graphs show means ± SD. Data represent at least two experiments with three to seven mice per group. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ as calculated using one-way [(A), (B), (D)] or two-way [(E), (F)] ANOVA with Holm-Šidák multiple-comparison test.

Biosciences. *Gata3*$^{fl/fl}$ (*37*), *Foxp3*$^{YFP\text{-}Cre}$ (*38*) and *Il17a*^{Cre} (*26*) have been previously described and were generously provided by J. Zhu (NIAID, NIH), A. Rudensky (Memorial Sloan Kettering Cancer Center), and B. Stockinger (Francis Crick Institute), respectively. *Foxp3*^{gfp} (*39*), CD45.1 (B6.SJL-Ptprc^a Pepc^b/BoyJ), *Tcra*$^{-/-}$ (B6.129S2-Tcra^{tm1Mom}/J) (*40*), CD45.1 *Rag1*$^{-/-}$, *Il13*$^{-/-}$ (*41*), and CreERT2-*Gata3*$^{fl/fl}$ mice (*42*) were purchased from the NIAID-Taconic Exchange. *Tcra*$^{+/-}$ mice were generated by breeding *Tcra*$^{-/-}$ mice with C57BL/6 WT mice. *Foxp3*^{DTR} (B6.129(Cg)-Foxp3$^{tm3(DTR/GFP)}$ Ayr/J) (*11*) and *R26R*^{eYFP} (B6.129X1-Gt(ROSA) 26Sor$^{tm1(EYFP)Cos}$/J) (*43*) mice were purchased from The Jackson Laboratory. *CD4*^{Cre}*Il18r1*$^{fl/fl}$ and *Il18r1*$^{fl/fl}$ control mice were kindly provided by G. Trinchieri (NCI, NIH). All mice were bred and maintained under SPF conditions at an American Association for the Accreditation of Laboratory Animal Care (AAALAC)–accredited animal facility at NIAID and housed in accordance with the procedures outlined in the Guide for the Care and Use of Laboratory Animals. All experiments were performed at NIAID under an Animal Study Proposal (LPD-11E or LPD-68E) approved by the NIAID Animal Care and Use Committee. Sex- and age-matched mice between

6 and 35 weeks of age were used for each experiment.

Commensal culture and colonization

Staphylococcus epidermidis NIHLM087 (*44*) was cultured for 18 hours in Tryptic Soy Broth at 37°C. *Candida albicans* (*8*) was cultured for 18 hours in Tryptic Soy Broth at 37°C (shaking 180 rpm). For colonization with commensal microbes, as before (*45*), each mouse was topically associated by placing 5 ml of culture suspension (approximately 10^9 CFU/ml) across the entire skin surface (approximately 36 cm^2) using a sterile swab. Application of commensal microbes was repeated every other day a total of four times. Skin tissue was analyzed 14 days after initial colonization, unless otherwise indicated.

Inducible deletion of Gata3

Deletion of *Gata3* in CreERT2*Gata3*$^{fl/fl}$ mice was induced by intraperitoneal injection of 5 mg tamoxifen in a corn oil–ethanol (90:10) mixture daily for 3 days before cellular isolation and subsequent analysis.

Global T_{reg} cell depletion

Naïve or *S. epidermidis*–colonized *Foxp3*^{DTR} mice received ~1 μg (50 μg/kg) of diphtheria toxin (Sigma-Aldrich) in phosphate-buffered saline (PBS), or PBS alone, by intraperitoneal (i.p.) injection on days 3, 5, 7, and 9 after initial *S. epidermidis* colonization. Flow cytometric analysis of skin leukocytes was performed 12 days after initial colonization.

Generation of Bowie^{Tg} mice

Tcra$^{+/-}$ mice were colonized with *S. epidermidis*, and CD8$^+$CCR6$^+$ T cells were isolated from skin tissue by fluorescence-activated cell sorting (FACS) and subjected to single-cell sequencing of TCR α and β chains (*46*). Clonal TCR pairs were identified and used in a hybridoma reconstitution screening assay to identify *S. epidermidis*–reactive TCR heterodimers. A single *S. epidermidis*–specific TCR pair was cloned into a hCD2-expression vector (*47*) and used to generate TCR-transgenic mice (Bowie^{Tg}), to track *S. epidermidis*–specific T cells in vivo.

Adoptive transfer of Bowie^{Tg} CD8$^+$ T cells

Bowie^{Tg} mice were backcrossed to a CD45.1 *Rag1*$^{-/-}$ background to exclude the possibility of dual TCR expression and facilitate identification of transferred cells. Recipient mice (CD45.2) received 4×10^5 Bowie^{Tg} CD45.1 *Rag1*$^{-/-}$ CD8$^+$ T cells by intravenous injection 24 hours before the first application of *S. epidermidis*.

Parabiosis experiments and surgery

Congenically distinct, age- and weight-matched mice were co-housed for 2 weeks before colonization with *S. epidermidis*. Both mice were colonized to control for bacterial spread. Forty days after initial colonization, parabiosis surgery was performed as described (*10*). Briefly, mice were sedated and longitudinal incisions were made from the elbow to the knee joint of each mouse. Excess skin was removed and mice were joined at the joints. The skin of the two animals was then connected and sutured together. Animals were kept on oral antibiotics for 2 weeks and remained conjoined for 90 to 95 days before analysis. Analysis was performed on ear pinnae skin tissue.

Acute intradermal challenge

S. epidermidis–colonized mice were anesthetized with ketamine-xylazine and injected intradermally (10 μl per ear pinnae) with either sterile PBS (vehicle control), 250 ng of carrier-free recombinant IL-18 (R&D Systems), or 500 ng of chitin (Sigma-Aldrich). Unless otherwise indicated, skin tissue was analyzed for cytokine production potential 48 hours after injury.

Sand fly bite exposure

S. epidermidis–colonized mice were exposed to sand fly bites as described (*48*). Briefly, mice were anesthetized with ketamine-xylazine. Twenty female *Lutzomiya longipalpus* were transferred to plastic vials (volume 12.2 cm^2, height 4.8 cm, diameter 1.8 cm) covered at one end with 0.25 mm of nylon mesh. Specially designed clamps were used to bring the mesh end of each vial flat against the ear, allowing flies to feed on exposed skin for a period of 1 hour in the dark at 26°C and 50% humidity. The number of flies with blood meals was employed as a means of checking for equivalent exposure to bites among animals. At indicated time points after exposure, tissues were analyzed for cytokine production.

Ex vivo cytokine screening

CD4$^+$ and CD8$^+$ T cell subsets from the skin of *S. epidermidis*– *or C. albicans*–colonized mice were isolated by FACS (> 97% purity) and cultured for 24 hours in the presence of cytokines (IL-1α, IL-1β, IL-18, IL-25, IL-33, or TSLP; R&D Systems) (10 ng/ml) and presence or absence of TCR stimulation (1 μg/ml plate bound anti-CD3 mAb, clone 145-2C11). Culture supernatants were assayed for cytokine production by FlowCytomix bead array (eBioscience).

Tissue processing

Cells from the skin-draining lymph nodes, spleen, and ear pinnae were isolated as described (*6*). Cells from lymph nodes and spleen were mashed through a 70-μm cell strainer to generate single-cell suspensions. Ear pinnae were excised and separated into ventral and dorsal sheets. Ear pinnae were digested in RPMI 1640 media supplemented with 2 mM L-glutamine, 1 mM sodium pyruvate, 1 mM non-essential amino acids, 50 μM β-mercaptoethanol, 20 mM HEPES, 100 U/ml of penicillin, 100 mg/ml of streptomycin, and 0.25 mg/ml of Liberase TL purified enzyme blend (Roche), and incubated for 90 min at 37°C and 5% CO_2. Digested skin sheets were homogenized using the Medicon/Medimachine tissue homogenizer system (Becton Dickinson).

In vitro restimulation

For detection of basal cytokine potential, single-cell suspensions from various tissues were cultured directly ex vivo in a 96-well U-bottom plate in complete medium (RPMI 1640 supplemented with 10% fetal bovine serum, 2 mM L-glutamine, 1 mM sodium pyruvate, 1 mM nonessential amino acids, 20 mM HEPES, 100 U/ml penicillin, 100 mg/ml streptomycin, and 50 μM β-mercaptoethanol) and stimulated with 50 ng/ml of phorbol myristate acetate (PMA) (Sigma-Aldrich) and 5 mg/ml of ionomycin (Sigma-Aldrich) in the presence of brefeldin A (1:1000, GolgiPlug, BD Biosciences) for 3 hours at 37°C in 5% CO_2. After stimulation, cells were assessed for intracellular cytokine production as described below.

Flow cytometric analysis

Single-cell suspensions were incubated with combinations of fluorophore-conjugated antibodies against the following surface markers: CCR6 (29-2L17), CD3ε (145-2C11), CD4 (RM4-5), CD8β (53-6.7), CD11b (M1/70), CD19 (6D5), CD44 (IM7), CD45 (30-F11), CD45.1 (A20), CD45.2 (104), CD69 (H1.2F3), CD103 (2E7), MHCII (M5/114.15.2), TCRβ (H57-597), and/or Thy1.2 (30-H12) in Hank's buffered salt solution (HBSS) for 20 min at 4°C (RT for 30 min for CCR6) and then washed. LIVE/DEAD Fixable Blue Dead Cell Stain Kit (Invitrogen Life Technologies) was used to exclude dead cells. Cells were then fixed for 30 min at 4°C using BD Cytofix/Cytoperm (Becton Dickinson) and washed twice. For intracellular cytokine staining, cells were stained with fluorophore-conjugated antibodies against IFN-γ (XMG-1.2), IL-5 (TRK5), IL-13 (eBio13A), and IL-17A (eBio17B7) in BD Perm/Wash Buffer (Becton Dickinson) for 60 min at 4°C. For transcription factor staining, cells were fixed and permeabilized with the Foxp3/Transcription Factor staining buffer set (eBioscience) and stained with fluorophore-conjugated antibodies against Foxp3 (FJK-16s), GATA-3 (L50-823 or TWAJ), RORγt (B2D), or T-bet (eBio4B10) for 45 min at 4°C. Each staining was performed in the presence of purified anti-mouse CD16/32 (clone 93) and purified rat gamma globulin (Jackson Immunoresearch). All antibodies were purchased from eBioscience, Biolegend, BD Biosciences, or Miltenyi Biotec. Cell acquisition was performed on a BD Fortessa X-20 flow cytometer using FACSDiVa software (BD Biosciences) and data were analyzed using FlowJo software (TreeStar).

RNA staining

Skin tissue single-cell suspensions were analyzed for mRNA and protein expression using the PrimeFlow RNA assay (eBioscience) and standard mouse probe sets for *Ifng*, *Il5*, *Il13*, and *Il17a*, as per manufacturer's instructions for 96-well-plate staining.

Tetramer-based cell enrichment

f-MIIINA:H2-M3-specific CD8$^+$ T cells from secondary lymphoid organs were subjected to magnetic bead based enrichment, as previously described (*49*). Briefly, spleen and lymph node cells from parabiotic pairs were stained for 1 hour at room temperature with f-MIIINA:H2-M3-streptavidin-phycoerythrin (PE) tetramer. Samples were then incubated with anti-PE beads (Miltenyi Biotech)

and enriched for bead-bound cells on magnetized columns.

RNA-sequencing and transcriptome analysis

T cells were isolated by flow cytometric cell sorting from the ear skin tissue of C57BL/6 mice 2 weeks after colonization with *S. epidermidis* NIHLM087. Groups included: T_C1 (Viable Lineage$^-$CD45$^+$ CD90.2$^+$TCRβ$^+$CD8β$^+$CCR6$^-$) and T_C17 cells (Viable Lineage$^-$CD45$^+$CD90.2$^+$TCRβ$^+$CD8β$^+$CCR6$^+$). Sorted cells were lysed in Trizol reagent and total RNA isolated by phenol-chloroform extraction with GlycoBlue as a co-precipitant (7 μg per sample; Life Technologies). Single-end libraries were prepared with 0.25 to 1 μg of total RNA using the TruSeq RNA Sample Preparation Kit V2 and sequenced for 50 cycles with a HiSeq 2500 instrument (4 to 6 samples multiplexed per lane; Illumina). Sequencing quality of the raw read data was assessed using FASTQC v0.11.5. Using a custom Perl script, 10 bp were trimmed from the 3′ end of the 50-bp reads. Subsequently, FASTQ files were used as input for RSEM v1.3.0 (*50*) (internally configured to use the bowtie aligner, v1.1.1). Expected read counts from RSEM were imported into the DESeq2 Bioconductor package (*51*), normalized using the geometric-mean based approach built into this package and then tested for differential expression between groups using a Wald test with multiple testing correction using Benjamini-Hochberg false discovery.

ATAC sequencing and epigenome analysis

T cells were isolated as for RNA sequencing. ATAC-seq was performed according to a published protocol (*16*). ATAC-seq reads from two biological replicates for each sample were mapped to the mouse genome (mm10 assembly) using STAR (*52*). Duplicate reads were removed using FastUniq (*53*), and reads mapping to mitochondrial loci removed based upon ENCODE blacklists. Regions of open chromatin were identified by MACS (version 1.4.2) using a *P*-value threshold of 1×10^{-5}. Only regions called in both replicates were used in downstream analysis. Downstream analysis and heatmap generation were performed with the Hypergeometric Optimization of Motif EnRichment program (HOMER) version 4.8 (*54*).

Single-cell RNA sequencing

T cells were isolated as for bulk RNA sequencing, from *S. epidermidis*–colonized IL-17A-fate-mapping mice, with three groups: T_C1 (Viable Lineage$^-$ CD45$^+$CD90.2$^+$TCRβ$^+$CD8β$^+$CCR6$^-$), T_C17 IL-17A^{FM-} (Viable Lineage$^-$CD45$^+$CD90.2$^+$TCRβ$^+$CD8β$^+$ CCR6$^+$eYFP$^-$), and T_C17 IL-17A^{FM+} (Viable Lineage$^-$ CD45$^+$CD90.2$^+$TCRβ$^+$CD8β$^+$CCR6$^+$eYFP$^+$). Freshly isolated cells were encapsulated into droplets, and libraries prepared using Chromium Single Cell 3′ Reagent Kits v2 (10X Genomics). The generated scRNA-seq libraries were sequenced using 26 cycles of Read 1, 8 cycles of i7 Index, and 98 cycles of Read2 with a HiSeq 3000 (Illumina).

Single-cell RNA sequencing analysis

Sequence reads were processed and aggregated using Cell Ranger software. Aggregated data were further analyzed using Seurat (*55*).

Confocal microscopy

Ear pinnae were split with forceps, fixed in 1% paraformaldehyde in PBS (Electron Microscopy Sciences) overnight at 4°C, and blocked in 1% BSA + 0.25% Triton X in PBS for 2 hours at room temperature. Tissues were first stained with anti-CD4 (RM4-5, eBioscience), anti-CD8α (clone 53-6.7, eBioscience), anti-CD45.1 (A20, eBioscience), anti-CD49f (GoH3, eBioscience), and/or anti-GFP (A21311, Life Technologies) antibodies overnight at 4°C, washed three times with PBS and then stained with 4,6-diamidino-2-phenylindole (DAPI, Sigma-Aldrich) overnight before being mounted with ProLong Gold (Molecular Probes) antifade reagent. Ear pinnae images were captured on a Leica TCS SP8 confocal microscope equipped with HyD and PMT detectors and a 40× oil objective (HC PL APO 40×/1.3 oil). Images were analyzed using Imaris software (Bitplane).

Back-skin wounding and epifluorescence microscopy of back-skin wounds

Tissue wounding and quantitation of wound healing were performed as previously described (*56*). Briefly, male mice in the telogen phase of the hair cycle were anesthetized and punch biopsies performed on back skin. Dorsal hair was shaved with clippers and a 6-mm biopsy punch was used to partially perforate the skin. Iris scissors were then used to cut epidermal and dermal tissue to create a full thickness wound in a circular shape. Back-skin tissue was excised 5 days after wounding, fixed in 4% paraformaldehyde in PBS, incubated overnight in 30% sucrose in PBS, embedded in OCT compound (Tissue-Tek), frozen on dry ice, and cryo-sectioned (20-μm section thickness). Sections were fixed in 4% paraformaldehyde in PBS, rinsed with PBS, permeabilized with 0.1% Triton X-100 in PBS (Sigma-Aldrich), and blocked for 1 hour in blocking buffer (2.5% Normal Goat Serum, 1% BSA, 0.3% Triton X-100 in PBS). Primary antibody to Keratin 14 (chicken, Poly9060, 1:400, Biolegend) was diluted in blocking buffer with rat gamma globulin and anti-CD16/32 and incubated overnight. After washing with PBS, a secondary antibody conjugated with Alexa647 (goat anti-chicken, Jackson ImmunoResearch) was added for 1 hour at room temperature. Slides were washed with PBS, counterstained with DAPI and mounted in Prolong Gold. Wound images were captured with a Leica DMI 6000 widefield epifluorescence microscope equipped with a Leica DFC360X monochrome camera. Tiled and stitched images of wounds were collected using a 20×/0.4NA dry objective. Images were analyzed using Imaris software (Bitplane).

In vivo cytokine blockade

Naïve or *S. epidermidis*–colonized WT or *Il13*$^{-/-}$ mice received 0.5 mg of anti–IL-13 monoclonal antibody (clone 262A-5-1, Genentech) or mouse IgG1 isotype control (clone MOPC-21, BioXCell), or 1 mg of anti–IL-5 monoclonal antibody (clone TRFK5, BioXCell) or rat IgG1 isotype control (clone TNP6A7, BioXCell), or 1 mg of anti–IL-18 monoclonal antibody [clone SK113AE-4 (*57*)] or isotype control by i.p. injection 1 day before skin injury.

Total tissue RNA-seq

A ~1-mm skin region surrounding the wound site was microdissected at indicated time points after wounding, submerged in RNAlater (Sigma-Aldrich), and stored at −20°C. Total tissue RNA was isolated from skin tissue using the RNeasy Fibrous Tissue Mini kit (Qiagen), as per manufacturer's instructions. A 3′ mRNA sequencing library was prepared using 200 to 500 ng of total input RNA with the QuantSeq 3′ mRNA-Seq Library Prep Kit FWD for Illumina (Lexogen) as per manufacturer's instructions. Libraries were quantified using an Agilent Tapestation (High Sensitivity D1000 ScreenTape) and Qubit (Thermo Fisher Scientific). Libraries (n = 20) were pooled at equimolar concentrations and sequenced on an Illumina Nextseq 500 using the High Output v2 kit (75 cycles). Resultant data was demultiplexed on Illumina Basespace server using blc2fastq tool. The reads from the Illumina Next-seq sequencer in fastq format were verified for quality control using FastQC software package, aligned to mouse GRCM38 using RSEM package (*50*) calling STAR aligner (*52*). The RSEM expected counts were rounded to the nearest integer value and the transcripts with zero counts across all samples filtered out. Differential expression analysis and principal components analysis was performed using DESeq2 (*51*).

Statistics

Groups were compared with Prism V7.0 software (GraphPad) using the two-tailed unpaired Student *t* test, one-way analysis of variance (ANOVA) with Holm-Šidák multiple-comparison test, or two-way ANOVA with Holm-Šidák multiple-comparison test where appropriate. Differences were considered to be statistically significant when $P < 0.05$.

REFERENCES AND NOTES

1. Y. Cong, T. Feng, K. Fujihashi, T. R. Schoeb, C. O. Elson, A dominant, coordinated T regulatory cell-IgA response to the intestinal microbiota. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19256–19261 (2009). doi: 10.1073/pnas.0812681106; pmid: 19889972
2. T. W. Hand *et al.*, Acute gastrointestinal infection induces long-lived microbiota-specific T cell responses. *Science* **337**, 1553–1556 (2012). doi: 10.1126/science.1220961; pmid: 22923434
3. Y. Yang *et al.*, Focused specificity of intestinal TH17 cells towards commensal bacterial antigens. *Nature* **510**, 152–156 (2014). doi: 10.1038/nature13279; pmid: 24739972
4. S. Naik *et al.*, Commensal-dendritic-cell interaction specifies a unique protective skin immune signature. *Nature* **520**, 104–108 (2015). doi: 10.1038/nature14052; pmid: 25539086
5. Y. Belkaid, O. J. Harrison, Homeostatic Immunity and the Microbiota. *Immunity* **46**, 562–576 (2017). doi: 10.1016/j.immuni.2017.04.008; pmid: 28423337
6. J. L. Linehan *et al.*, Non-classical Immunity Controls Microbiota Impact on Skin Immunity and Tissue Repair. *Cell* **172**, 784–796. e18 (2018). doi: 10.1016/j.cell.2017.12.033; pmid: 29358051
7. T. C. Scharschmidt *et al.*, A Wave of Regulatory T Cells into Neonatal Skin Mediates Tolerance to Commensal Microbes. *Immunity* **43**, 1011–1021 (2015). doi: 10.1016/j.immuni.2015.10.016; pmid: 26588783

8. V. K. Ridaura *et al*., Contextual control of skin immunity and inflammation by *Corynebacterium*. *J. Exp. Med.* **215**, 785–799 (2018). doi: 10.1084/jem.20171079; pmid: 29382696
9. J. M. Schenkel, D. Masopust, Tissue-resident memory T cells. *Immunity* **41**, 886–897 (2014). doi: 10.1016/j.immuni.2014.12.007; pmid: 25526304
10. D. E. Wright, A. J. Wagers, A. P. Gulati, F. L. Johnson, I. L. Weissman, Physiological migration of hematopoietic stem and progenitor cells. *Science* **294**, 1933–1936 (2001). doi: 10.1126/science.1064081; pmid: 11729320
11. J. M. Kim, J. P. Rasmussen, A. Y. Rudensky, Regulatory T cells prevent catastrophic autoimmunity throughout the lifespan of mice. *Nat. Immunol.* **8**, 191–197 (2007). doi: 10.1038/ni1428; pmid: 17136045
12. E. A. Wohlfert *et al*., GATA3 controls Foxp3$^+$ regulatory T cell fate during inflammation in mice. *J. Clin. Invest.* **121**, 4503–4515 (2011). doi: 10.1172/JCI57456; pmid: 21965331
13. M. Delacher *et al*., Genome-wide DNA-methylation landscape defines specialization of regulatory T cells in tissues. *Nat. Immunol.* **18**, 1160–1172 (2017). doi: 10.1038/ni.3799; pmid: 28783152
14. Y. Wang, M. A. Su, Y. Y. Wan, An essential role of the transcription factor GATA-3 for the function of regulatory T cells. *Immunity* **35**, 337–348 (2011). doi: 10.1016/j.immuni.2011.08.012; pmid: 21924928
15. D. Rudra *et al*., Transcription factor Foxp3 and its protein partners form a complex regulatory network. *Nat. Immunol.* **13**, 1010–1019 (2012). doi: 10.1038/ni.2402; pmid: 22922362
16. H. Y. Shih *et al*., Developmental Acquisition of Regulomes Underlies Innate Lymphoid Cell Functionality. *Cell* **165**, 1120–1133 (2016). doi: 10.1016/j.cell.2016.04.029; pmid: 27156451
17. G. R. Lee, P. E. Fields, T. J. Griffin IV, R. A. Flavell, Regulation of the Th2 cytokine locus by a locus control region. *Immunity* **19**, 145–153 (2003). doi: 10.1016/S1074-7613(03)00179-1; pmid: 12871646
18. H. E. Liang *et al*., Divergent expression patterns of IL-4 and IL-13 define unique functions in allergic immunity. *Nat. Immunol.* **13**, 58–66 (2011). doi: 10.1038/ni.2182; pmid: 22138715
19. K. Mohrs, A. E. Wakil, N. Killeen, R. M. Locksley, M. Mohrs, A two-step process for cytokine production revealed by IL-4 dual-reporter mice. *Immunity* **23**, 419–429 (2005). doi: 10.1016/j.immuni.2005.09.006; pmid: 16226507
20. C. H. Chang *et al*., Posttranscriptional control of T cell effector function by aerobic glycolysis. *Cell* **153**, 1239–1251 (2013). doi: 10.1016/j.cell.2013.05.016; pmid: 23746840
21. K. Araki *et al*., Translation is actively regulated during the differentiation of CD8$^+$ effector T cells. *Nat. Immunol.* **18**, 1046–1057 (2017). pmid: 28714979
22. F. Salerno *et al*., Translational repression of pre-formed cytokine-encoding mRNA prevents chronic activation of memory T cells. *Nat. Immunol.* **19**, 828–837 (2018). doi: 10.1038/s41590-018-0155-6; pmid: 29988089
23. W. V. Bonilla *et al*., The alarmin interleukin-33 drives protective antiviral CD8$^+$ T cell responses. *Science* **335**, 984–989 (2012). doi: 10.1126/science.1215418; pmid: 22323740
24. I. Okamoto, K. Kohno, T. Tanimoto, H. Ikegami, M. Kurimoto, Development of CD8+ effector T cells is differentially regulated by IL-18 and IL-12. *J. Immunol.* **162**, 3202–3211 (1999). pmid: 10092771
25. P. P. Ahern *et al*., Interleukin-23 drives intestinal inflammation through direct activity on T cells. *Immunity* **33**, 279–288 (2010). doi: 10.1016/j.immuni.2010.08.010; pmid: 20732640
26. K. Hirota *et al*., Fate mapping of IL-17-producing T cells in inflammatory responses. *Nat. Immunol.* **12**, 255–263 (2011). doi: 10.1038/ni.1993; pmid: 21278737
27. Y. H. Wang *et al*., A novel subset of CD4$^+$ T_H2 memory/effector cells that produce inflammatory IL-17 cytokine and promote the exacerbation of chronic allergic asthma. *J. Exp. Med.* **207**, 2479–2491 (2010). doi: 10.1084/jem.20101376; pmid: 20921287
28. M. Panzer *et al*., Rapid in vivo conversion of effector T cells into Th2 cells during helminth infection. *J. Immunol.* **188**, 615–623 (2012). doi: 10.4049/jimmunol.1101164; pmid: 22156341
29. C. Irvin *et al*., Increased frequency of dual-positive TH2/TH17 cells in bronchoalveolar lavage fluid characterizes a population of patients with severe asthma. *J. Allergy Clin. Immunol.* **134**, 1175–1186.e7 (2014). doi: 10.1016/j.jaci.2014.05.038; pmid: 25042748
30. K. Hirota *et al*., Plasticity of Th17 cells in Peyer's patches is responsible for the induction of T cell-dependent IgA responses. *Nat. Immunol.* **14**, 372–379 (2013). doi: 10.1038/ni.2552; pmid: 23475182
31. N. Gagliani *et al*., Th17 cells transdifferentiate into regulatory T cells during resolution of inflammation. *Nature* **523**, 221–225 (2015). doi: 10.1038/nature14452; pmid: 25924064
32. K. Nakanishi, T. Yoshimoto, H. Tsutsui, H. Okamura, Interleukin-18 regulates both Th1 and Th2 responses. *Annu. Rev. Immunol.* **19**, 423–474 (2001). doi: 10.1146/annurev.immunol.19.1.423; pmid: 11244043
33. O. J. Harrison *et al*., Epithelial-derived IL-18 regulates Th17 cell differentiation and Foxp3$^+$ Treg cell function in the intestine. *Mucosal Immunol.* **8**, 1226–1236 (2015). doi: 10.1038/mi.2015.13; pmid: 25736457
34. N. Arpaia *et al*., A Distinct Function of Regulatory T Cells in Tissue Protection. *Cell* **162**, 1078–1089 (2015). doi: 10.1016/j.cell.2015.08.021; pmid: 26317471
35. R. R. Ricardo-Gonzalez *et al*., Tissue signals imprint ILC2 identity with anticipatory function. *Nat. Immunol.* **19**, 1093–1099 (2018). doi: 10.1038/s41590-018-0201-4; pmid: 30201992
36. S. A. Eming, T. A. Wynn, P. Martin, Inflammation and metabolism in tissue repair and regeneration. *Science* **356**, 1026–1030 (2017). doi: 10.1126/science.aam7928; pmid: 28596335
37. J. Zhu *et al*., Conditional deletion of Gata3 shows its essential function in T(H)1-T(H)2 responses. *Nat. Immunol.* **5**, 1157–1165 (2004). doi: 10.1038/ni1128; pmid: 15475959
38. Y. P. Rubtsov *et al*., Stability of the regulatory T cell lineage in vivo. *Science* **329**, 1667–1671 (2010). doi: 10.1126/science.1191996; pmid: 20929851
39. E. Bettelli *et al*., Reciprocal developmental pathways for the generation of pathogenic effector TH17 and regulatory T cells. *Nature* **441**, 235–238 (2006). doi: 10.1038/nature04753; pmid: 16648838
40. P. Mombaerts *et al*., Mutations in T-cell antigen receptor genes alpha and beta block thymocyte development at different stages. *Nature* **360**, 225–231 (1992). doi: 10.1038/360225a0; pmid: 1359428
41. G. J. McKenzie *et al*., Impaired development of Th2 cells in IL-13-deficient mice. *Immunity* **9**, 423–432 (1998). doi: 10.1016/S1074-7613(00)80625-1; pmid: 9768762
42. R. Yagi *et al*., The transcription factor GATA3 is critical for the development of all IL-7Rα-expressing innate lymphoid cells. *Immunity* **40**, 378–388 (2014). doi: 10.1016/j.immuni.2014.01.012; pmid: 24631153
43. S. Srinivas *et al*., Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev. Biol.* **1**, 4 (2001). doi: 10.1186/1471-213X-1-4; pmid: 11299042
44. S. Conlan *et al*., Staphylococcus epidermidis pan-genome sequence analysis reveals diversity of skin commensal and hospital infection-associated isolates. *Genome Biol.* **13**, R64 (2012). doi: 10.1186/gb-2012-13-7-r64; pmid: 22830599
45. S. Naik *et al*., Compartmentalized control of skin immunity by resident commensals. *Science* **337**, 1115–1119 (2012). doi: 10.1126/science.1225152; pmid: 22837383
46. P. Dash *et al*., Paired analysis of TCRα and TCRβ chains at the single-cell level in mice. *J. Clin. Invest.* **121**, 288–295 (2011). doi: 10.1172/JCI44752; pmid: 21135507
47. G. Sun *et al*., The zinc finger protein cKrox directs CD4 lineage differentiation during intrathymic T cell positive selection. *Nat. Immunol.* **6**, 373–381 (2005). doi: 10.1038/ni1183; pmid: 15750595
48. N. Kimblin *et al*., Quantification of the infectious dose of Leishmania major transmitted to the skin by single sand flies. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10125–10130 (2008). doi: 10.1073/pnas.0802331105; pmid: 18626016
49. J. J. Moon *et al*., Naive CD4$^+$ T cell frequency varies for different epitopes and predicts repertoire diversity and response magnitude. *Immunity* **27**, 203–213 (2007). doi: 10.1016/j.immuni.2007.07.007; pmid: 17707129
50. B. Li, C. N. Dewey, RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011). doi: 10.1186/1471-2105-12-323; pmid: 21816040
51. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014). doi: 10.1186/s13059-014-0550-8; pmid: 25516281
52. A. Dobin *et al*., STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013). doi: 10.1093/bioinformatics/bts635; pmid: 23104886
53. H. Xu *et al*., FastUniq: A fast de novo duplicates removal tool for paired short reads. *PLOS ONE* **7**, e52249 (2012). doi: 10.1371/journal.pone.0052249; pmid: 23284954
54. S. Heinz *et al*., Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010). doi: 10.1016/j.molcel.2010.05.004; pmid: 20513432
55. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018). doi: 10.1038/nbt.4096; pmid: 29608179
56. B. E. Keyes *et al*., Impaired Epidermal to Dendritic T Cell Signaling Slows Wound Repair in Aged Skin. *Cell* **167**, 1323–1338.e14 (2016). doi: 10.1016/j.cell.2016.10.052; pmid: 27863246
57. M. Lochner, H. Wagner, M. Classen, I. Förster, Generation of neutralizing mouse anti-mouse IL-18 antibodies for inhibition of inflammatory responses in vivo. *J. Immunol. Methods* **259**, 149–157 (2002). doi: 10.1016/S0022-1759(01)00505-1; pmid: 11730850

ACKNOWLEDGMENTS

We thank the National Institute of Allergy and Infectious Diseases (NIAID) animal facility staff; K. Holmes, E. Stregevsky, and T. Hawley (NIAID Flow Cytometry facility); G. Gutierrez-Cruz, S. Dell'Orso, and H.-W. Sun (NIAMS Genome Analysis Core facility); J. Kehr for editorial assistance; K. Beacht and S. Mistry for technical assistance; and I. Förster (University of Bonn) for generous provision of the anti–IL-18 hybridoma. f-MIIINA:H2-M3-tetramer reagents were obtained from the NIH Tetramer Core Facility. This study used the Office of Cyber Infrastructure and Computational Biology (OCICB) High Performance Computing (HPC) cluster at NIAID and the high-performance computational capabilities of the Biowulf Linux cluster at NIH. **Funding:** Supported by the NIAID Division of Intramural Research (ZIA-AI001115, ZIA-AI001132) (Y.B.); the Division of Intramural Research of the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS; ZIA-AR041159, ZIA-AR041167) (J.J.O.); a National Psoriasis Foundation Early Career Research Grant (O.J.H.); the National Institute of General Medical Sciences (NIGMS) Postdoctoral Research Associate (PRAT) fellowship program (J.L.L.); a European Molecular Biology Organization (EMBO) fellowship (S.T.); and Collège des Enseignants de Dermatologie Français, Société Française de Dermatologie, Philippe Foundation, and Fondation pour la Recherche Médicale (C.H.). **Author contributions:** O.J.H. and Y.B. designed the study, experiments, and wrote the manuscript; O.J.H. performed the experiments and analyzed the data; J.L.L., S.-J.H., N.B., H.-Y.S., M.S., S.K.S., A.L.B., M.E., S.T., F.V.L., C.H., N.C., A.P., R.S., and J.J.O. participated in performing experiments, provided intellectual expertise, and helped to interpret experimental results; J.L.L. generated Bowie^{Tg} mice, performed wounding experiments, and analyzed data; H.-Y.S., S.K.S., A.L.B., and C.Y. assisted with RNA-seq and ATAC-seq studies; N.B. and S.T. performed flow cytometric analysis of skin immune cells; S.J.H. performed confocal microscopy analysis; M.S. performed epifluorescence microscopy of wounds; M.E. assisted with wounding experiments; S.J.H. and N.C. performed parabiotic surgeries; F.V.L. shared expertise for the generation of TCR-expressing hybridomas and TCR-transgenic mice; A.P. conducted sand fly exposures; and C.H. performed *C. albicans* experiments. **Competing interests:** Authors declare no competing interests. **Data and materials availability:** Anti–IL-13 (clone 262A-5-1) is available under a material agreement with Genentech. Anti–IL-18 (clone SK113AE-4) is available from I. Förster under a material agreement with the University of Bonn. The accession number for the RNA-seq and ATAC-seq datasets is NCBI BioProject: PRJNA486019. All other data needed to evaluate the conclusions in this paper are present either in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/363/6422/eaat6280/suppl/DC1
Figs. S1 to S6

19 March 2018; accepted 9 November 2018
Published online 6 December 2018
10.1126/science.aat6280

RESEARCH ARTICLE SUMMARY

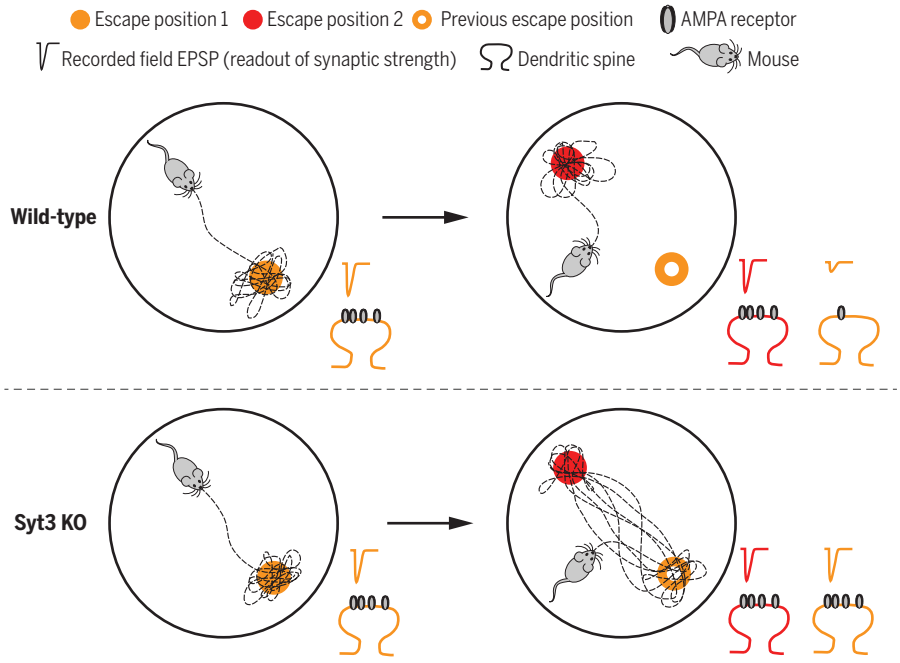
NEUROSCIENCE

Synaptotagmin-3 drives AMPA receptor endocytosis, depression of synapse strength, and forgetting

Ankit Awasthi*, Binu Ramachandran*, Saheeb Ahmed, Eva Benito, Yo Shinoda, Noam Nitzan, Alina Heukamp, Sabine Rannio, Henrik Martens, Jonas Barth, Katja Burk, Yu Tian Wang, Andre Fischer, Camin Dean†

INTRODUCTION: Memories are stored as molecular and cellular changes in the brain. Synapses, the nodes of connection between neurons, can store memories by virtue of their ability to tune the efficacy of communication between neurons. This property of synaptic plasticity makes it possible for the brain to store and retrieve memories—to replay patterns of electrical activity that occurred during an important event. Forgetting leads to the inability to retrieve memories by making them latent or decaying them below any useful quality. However, what determines whether a memory is forgotten? A mechanism is the regulation of neurotransmitter receptor numbers on the postsynaptic plasma membrane. These receptors mediate synaptic transmission by transducing presynaptically released neurotransmitters into an electrical signal. Neuronal activity strengthens synapses by inserting receptors or weakens synapses by removing receptors from the postsynaptic membrane. Receptor trafficking is controlled through calcium influx into the neuron; however, the calcium sensors mediating this control are not known.

RATIONALE: Synaptotagmin proteins sense calcium to trigger membrane fusion. Stimulating neuronal cultures elicits a calcium-mediated externalization of most synaptotagmin isoforms into plasma membranes, but synaptotagmin-3 (Syt3) internalizes from postsynaptic membranes. Stimulating AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic) or NMDA (*N*-methyl D-aspartate) receptors induces internalization of AMPA receptors (which mediate most of the fast synaptic transmission in the brain) and Syt3. This raised the intriguing possibility that Syt3 mediates activity-induced internalization of receptors to weaken synapses and cause forgetting. We imaged Syt3 using an isoform-specific antibody, tested its role in receptor trafficking using electrophysiological methods in brain slices and neuronal cultures, and tested its role in forgetting using spatial memory tasks in mice.



Syt3 knockout mice do not forget. Both wild-type mice and Syt3 knockout mice can learn an escape position in the water maze, in which corresponding synapses are strengthened through the increase of AMPA receptors. These synapses are weakened by the removal of receptors if the memory is no longer needed—for example, when a new escape position is learned. Syt3 knockout mice cannot remove receptors and therefore cannot forget previous escape positions.

RESULTS: Syt3 is on postsynaptic membranes at endocytic zones, which are clathrin-rich regions close to the postsynaptic density. Syt3 binds the GluA2 AMPA receptor subunit and also binds AP2 and BRAG2, two proteins implicated in activity-dependent internalization of AMPA receptors via clathrin-mediated endocytosis. Syt3 does not affect basal AMPA receptor trafficking. However, knocking out Syt3—or expressing calcium-binding–deficient Syt3—abolishes AMPA receptor internalization induced by AMPA, NMDA, or electrophysiological stimulation of long-term depression of synaptic strength. It also blocks the AMPA receptor internalization that normally decays long-term potentiation of synaptic strength. These effects are mimicked in a wild-type background through acute application of the Tat-GluA2-3Y peptide, which competitively inhibits binding of Syt3 to a tyrosine-rich (3Y) motif on the cytoplasmic tail of GluA2. In spatial memory tasks, mice in which Syt3 was knocked out (Syt3 knockout mice) learn escape positions normally but persevere to previously learned positions, which can be explained by a lack of forgetting previously acquired memories. Injecting the Tat-GluA2-3Y peptide in wild-type mice mimics the lack of forgetting of spatial memories, and this effect is occluded in Syt3 knockout mice.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aav1483

CONCLUSION: The persistence or degradation of memories is governed by a poorly understood molecular machinery. We have discovered a distinct synaptotagmin isoform that triggers calcium-mediated internalization of AMPA receptors, resulting in a weakening of synaptic transmission and forgetting of spatial memories in mice. ■

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: c.dean@eni-g.de
Cite this article as A. Awasthi *et al.*, *Science* **363**, eaav1483 (2019). DOI: 10.1126/science.aav1483

RESEARCH ARTICLE

NEUROSCIENCE

Synaptotagmin-3 drives AMPA receptor endocytosis, depression of synapse strength, and forgetting

Ankit Awasthi[1]*, Binu Ramachandran[1]*, Saheeb Ahmed[1]†, Eva Benito[2,3], Yo Shinoda[1]‡, Noam Nitzan[1], Alina Heukamp[1]§, Sabine Rannio[1], Henrik Martens[4], Jonas Barth[2,3], Katja Burk[1], Yu Tian Wang[5], Andre Fischer[2,3], Camin Dean[1]¶

Forgetting is important. Without it, the relative importance of acquired memories in a changing environment is lost. We discovered that synaptotagmin-3 (Syt3) localizes to postsynaptic endocytic zones and removes AMPA receptors from synaptic plasma membranes in response to stimulation. AMPA receptor internalization, long-term depression (LTD), and decay of long-term potentiation (LTP) of synaptic strength required calcium-sensing by Syt3 and were abolished through Syt3 knockout. In spatial memory tasks, mice in which Syt3 was knocked out learned normally but exhibited a lack of forgetting. Disrupting Syt3:GluA2 binding in a wild-type background mimicked the lack of LTP decay and lack of forgetting, and these effects were occluded in the Syt3 knockout background. Our findings provide evidence for a molecular mechanism in which Syt3 internalizes AMPA receptors to depress synaptic strength and promote forgetting.

A fundamental property of the brain is the ability to learn from experience by modulating the strength of synaptic connections between neurons. The trafficking of AMPA receptors to and from the surface of postsynaptic membranes is a key determinant in the regulation of synaptic strength (*1*–*3*). High-frequency stimulation increases surface receptors and promotes long-term potentiation (LTP) of synaptic strength. Low-frequency stimulation removes receptors from the postsynaptic membrane and causes long-term depression (LTD) of synaptic strength. This phenomenon has been most extensively studied in *N*-methyl-D-aspartate (NMDA)–receptor–dependent plasticity of CA3-CA1 synapses in the hippocampus (*4*). LTP is widely believed to underlie learning (*5*–*7*). Conversely, weakening of potentiated synapses, and LTD, can promote forgetting (*5*, *8*, *9*).

Both LTD (*10*, *11*) and the decay of LTP depend on activity-dependent removal of postsynaptic GluA2-containing AMPA receptors from the plasma membrane (*8*, *12*, *13*). Ca^{2+} influx into dendrites is critical for virtually all types of synaptic plasticity (*4*), including decay of LTP (*14*, *15*) and active internalization of GluA2-containing AMPA receptors (*16*, *17*). The same signaling entity can have divergent consequences for the synapse: Fast, high [Ca^{2+}] influx can promote LTP, whereas gradual, low [Ca^{2+}] influx can promote LTD (*4*). Plasma membrane–localized Ca^{2+}-binding proteins are likely needed to remove or add receptors to the postsynaptic membrane through regulated endo- or exocytosis. Synaptotagmins (Syts) are candidates for such a function (*18*). This family of integral membrane proteins contain a short luminal tail, transmembrane domain, and two conserved cytoplasmic Ca^{2+}-binding C2 domains (*19*) that regulate Ca^{2+}-dependent membrane recycling events.

Of the 17 mammalian Syt isoforms, Syt3 is the third most abundant in the brain. Unlike Syt1 and Syt2, which are predominantly thought to be localized to synaptic vesicle membranes, Syt3 was reported to be localized to the plasma membrane (*20*, *21*) and to have a 10-fold higher Ca^{2+} affinity than that of Syt1 and Syt2 (*22*). In a pHluorin-Syt screen of Syt isoforms in response to stimulation, Syt3 exhibited distinct recycling properties: It was the only isoform to endocytose in response to stimulation and to recycle exclusively in dendrites (*23*). Because the kinetics of stimulus-induced pHluorin-Syt3 endocytosis resembled those of pHluorin-GluA2 (*24*, *25*), we hypothesized that Syt3 may regulate Ca^{2+}- and activity-dependent postsynaptic receptor endocytosis to affect synaptic plasticity.

[1]Trans-synaptic Signaling Group, European Neuroscience Institute, 37077 Goettingen, Germany. [2]German Center for Neurodegenerative Disease, 37075 Goettingen, Germany. [3]Department of Psychiatry and Psychotherapy, University Medical Center Goettingen, 37075 Goettingen, Germany. [4]Synaptic Systems GmbH, 37079 Goettingen, Germany. [5]Brain Research Center and Department of Medicine, University of British Columbia, Vancouver, BC V6T2B5, Canada.
*These authors contributed equally to this work.
†Present address: Department of Diagnostic and Interventional Radiology, University Medical Center Goettingen, 37075 Goettingen, Germany.
‡Present address: Department of Environmental Health, School of Pharmacy, Tokyo University of Pharmacy and Life Sciences, Tokyo 192-0392 Japan. §Present address: Department of Neurobiology, Weizmann Institute of Science, 7610001 Rehovot, Israel.
¶**Corresponding author. Email: c.dean@eni-g.de**

Results

Syt3 is on postsynaptic plasma membranes

To examine the location of Syt3, we developed a highly specific antibody (fig. S1, A and B) that recognized a single band in brain homogenates, which was absent in Syt3 knockouts; an antibody developed by Neuromab showed similar results (Fig. 1A). Syt3 was most abundant in adipose tissue, heart, and brain (fig. S1C), where it was found in the hippocampus, cortex, thalamus, and striatum (fig. S1D). Expression in the brain began embryonically and remained high throughout adulthood (fig. S1E). Immunostains revealed Syt3 signal on neuronal cell bodies and dendrites in the CA1 (Fig. 1B), CA3, and dentate gyrus (fig. S1F) in wild-type, but not Syt3 knockout, hippocampal slices.

To localize Syt3 subcellularly, we expressed cytosolic green fluorescent protein (GFP) in cultured hippocampal neurons and immunostained for Syt3 and MAP2 to distinguish dendrites (MAP2-positive) from axons (GFP-positive/MAP2-negative). Syt3 signal was predominantly detected in dendrites (90.2 ± 3.0% of total signal) compared with axons (9.8 ± 5.1% of total signal) ($P < 0.001$) (Fig. 1C). Syt3 exhibited a punctate pattern in dendrites and colocalized with the pre- and postsynaptic markers synaptophysin, and PSD95 or GluA1, respectively, at synapses (Fig. 1D); 70.6 ± 5.3% of Syt3 signal was at excitatory synapses (fig. S1G), and 29.4 ± 1.9% was at inhibitory synapses ($P < 0.001$) (fig. S1H). In subcellular fractions, Syt3 was associated with synaptosomal plasma membranes and not with synaptic vesicles (fig. S1I), which is in agreement with a previous report (*20*). Immuno-organelle isolation of synaptic vesicles with antibodies to Syt1 or Syb2 further confirmed that Syt3 is not present on synaptic vesicles (fig. S1J). We further tested whether Syt3 is specifically localized to postsynaptic membranes using a trypsin cleavage assay of synaptosomes (*26*). Synaptosomes form in such a way that the presynaptic terminal seals, trapping synaptic vesicles and other presynaptic components inside (Fig. 1E), whereas the postsynaptic membrane does not reseal. Presynaptic proteins were therefore protected from trypsin cleavage, whereas postsynaptic proteins, including Syt3, were cleaved (Fig. 1F).

Stimulation induces endocytosis of Syt3

The presence of Syt3 on postsynaptic membranes suggests it may regulate a post-synaptic recycling event. Time-lapse imaging of pHluorin-Syt3 in transfected hippocampal neurons during depolarization with 45 mM KCl (*23*) or field stimulation revealed a calcium-dependent fluorescence decrease, requiring NMDA/AMPA receptors and L-type calcium channels, likely corresponding to endocytosis (fig. S2, A to D). PHluorin-Syt3 also exhibited calcium-dependent endocytosis in response to AMPA (Fig. 2A) and NMDA (Fig. 2B)—stimuli that promote AMPA receptor internalization—with kinetics similar to those of pHluorin-tagged AMPA receptors (*24*, *25*).

Because interpretation of pHluorin experiments may be confounded by intracellular acidification

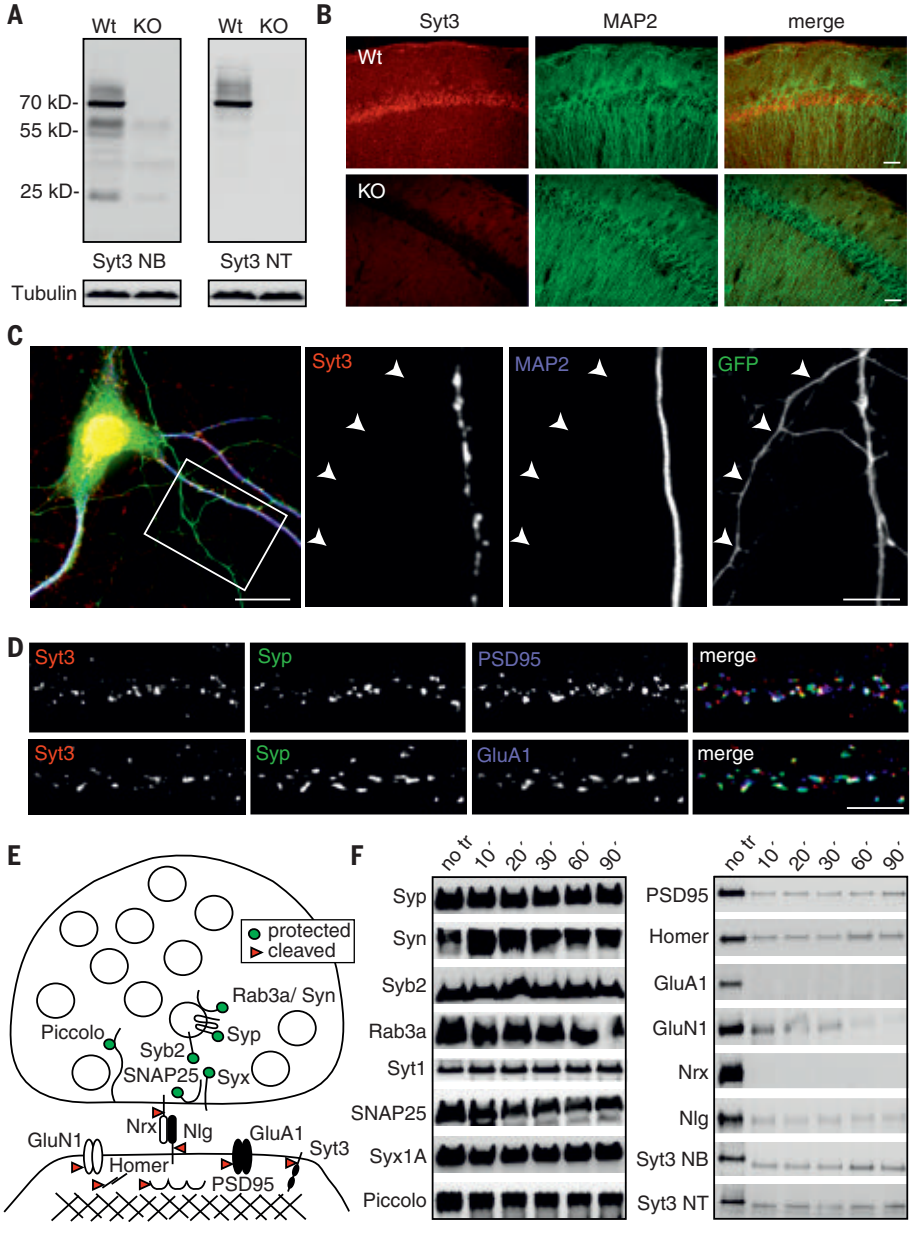


Fig. 1. Syt3 is postsynaptic. (**A**) Syt3 antibodies recognize a specific band of the expected molecular weight in wild-type (WT) but not in Syt3 knockout (KO) mouse brain homogenates in Western blots. Syt3 NB is a monoclonal antibody from Neuromab, and Syt3 NT is a polyclonal antibody developed by Synaptic Systems. Tubulin is a loading control. (**B**) Syt3 immunoreactivity is detected on pyramidal cell bodies and dendrites in the CA1 region of hippocampal slices but not in Syt3 knockouts. MAP2 marks dendrites. Scale bar, 50 μm. (**C**) Syt3 is predominantly in dendrites in hippocampal neurons transfected with GFP and immunostained with MAP2 to mark dendrites (MAP2-positive) or axons (GFP-positive/MAP2-negative) (n = 20 neurons/3 cultures). Scale bar, 20 μm (left), 5 μm (right). (**D**) Syt3 localizes to synapses marked by synaptophysin and PSD95 (top) or GluA1 (bottom) in hippocampal cultures. Scale bar, 5 μm. (**E**) Schematic of a synaptosome. The presynaptic side reseals, whereas the postsynaptic side does not, leaving postsynaptic proteins accessible to trypsin cleavage. (**F**) Presynaptic proteins are protected from trypsin cleavage, whereas postsynaptic proteins (including Syt3) are cleaved.

in response to AMPA and NMDA stimulation (*27*), we performed an additional assay to test internalization of Syt3. We expressed GFP-Syt3 in Syt3 knockout neurons and labeled surface GFP-Syt3 with an antibody to GFP, stimulated neurons with AMPA or NMDA, and then labeled remaining surface GFP-Syt3 with a secondary antibody of one color, permeabilized cells, and labeled internalized GFP-Syt3 with a secondary antibody of another color so as to quantify Syt3 internalized in response to stimulation. We found a significant increase in internalized GFP-Syt3 after stimulation with AMPA and NMDA compared with that in control conditions (Fig. 2C).

AMPA receptor endocytosis is thought to occur at postsynaptic endocytic zones—marked by fluorescently tagged Clathrin light chain—which is adjacent to, but distinct from, postsynaptic densities marked by fluorescently tagged Homer (*28*–*31*). To visualize these markers exclusively at postsynaptic sites, neuronal cultures were sparsely transfected so that dendrites of transfected neurons can be identified and fluorescent marker colocalization quantified at postsynaptic sites in these dendrites. We first verified that Clathrin and Homer can be distinguished; only 4.4 ± 1.1% of Clathrin light chain-DsRed and Homer-myc puncta overlapped (Fig. 2D). We further found that only 4.0 ± 1.3% of GFP-Syt3 puncta colocalized with Homer-myc at postsynaptic densities, whereas 84.0 ± 1.8% of GFP-Syt3 puncta in dendrites colocalized with Clathrin-DsRed at endocytic zones (Fig. 2D).

Syt3 internalizes AMPA receptors

Does Syt3 endocytose receptors? As a first test, we pulled down binding partners of Syt3 from brain homogenates. Recombinant Syt3 pulled down GluA2 but not other receptor subunits and also pulled down the endocytic proteins AP-2 (*32*) and BRAG2 (*10*)—two proteins implicated in receptor endocytosis—but not GRIP or PICK1 (Fig. 2E and fig. S3), which bind distinct regions of the GluA2 C-terminal tail (Fig. 2F) (*1*). Syt3 bound GluA2 in a calcium-dependent manner and bound AP-2 and BRAG2 in the absence of calcium, and at concentrations up to 10 μM in the case of AP2 (Fig. 2G). Because the nine–amino acid 3Y tail of GluA2 is important for receptor internalization, we tested whether Syt3 interacts with this region. Incubation of brain lysates with 1 μM Tat-GluA2-3Y peptide—which competitively inhibits protein binding to the GluA2 3Y region and blocks activity-dependent endocytosis of receptors (*10*, *13*)—indeed disrupted Syt3:GluA2 binding (Fig. 2H).

Because Syt3 pulled down GluA2 receptors, and GluA1/GluA2 heteromers comprise the majority of AMPA receptors in the hippocampus (*33*), we examined internalization of GluA1- and GluA2-containing receptors in response to stimulation (*17*, *32*, *34*) in dissociated hippocampal neurons in which Syt3 was overexpressed or knocked down postsynaptically, using sparse transfection. Surface epitopes of GluA1 or GluA2 in hippocampal cultures were labeled with primary antibodies, followed by stimulation with AMPA or NMDA to promote receptor internalization. Surface receptors were then labeled with a secondary antibody of one color and internalized receptors labeled with a second color after permeabilization, whereby the ratio of internal/surface signal reports the extent of receptor internalization. Overexpression of GFP-Syt3 increased internalization of GluA1 and GluA2, whereas Syt3 knockdown or expression of a calcium-binding–deficient mutant of Syt3 abolished stimulation-induced internalization of

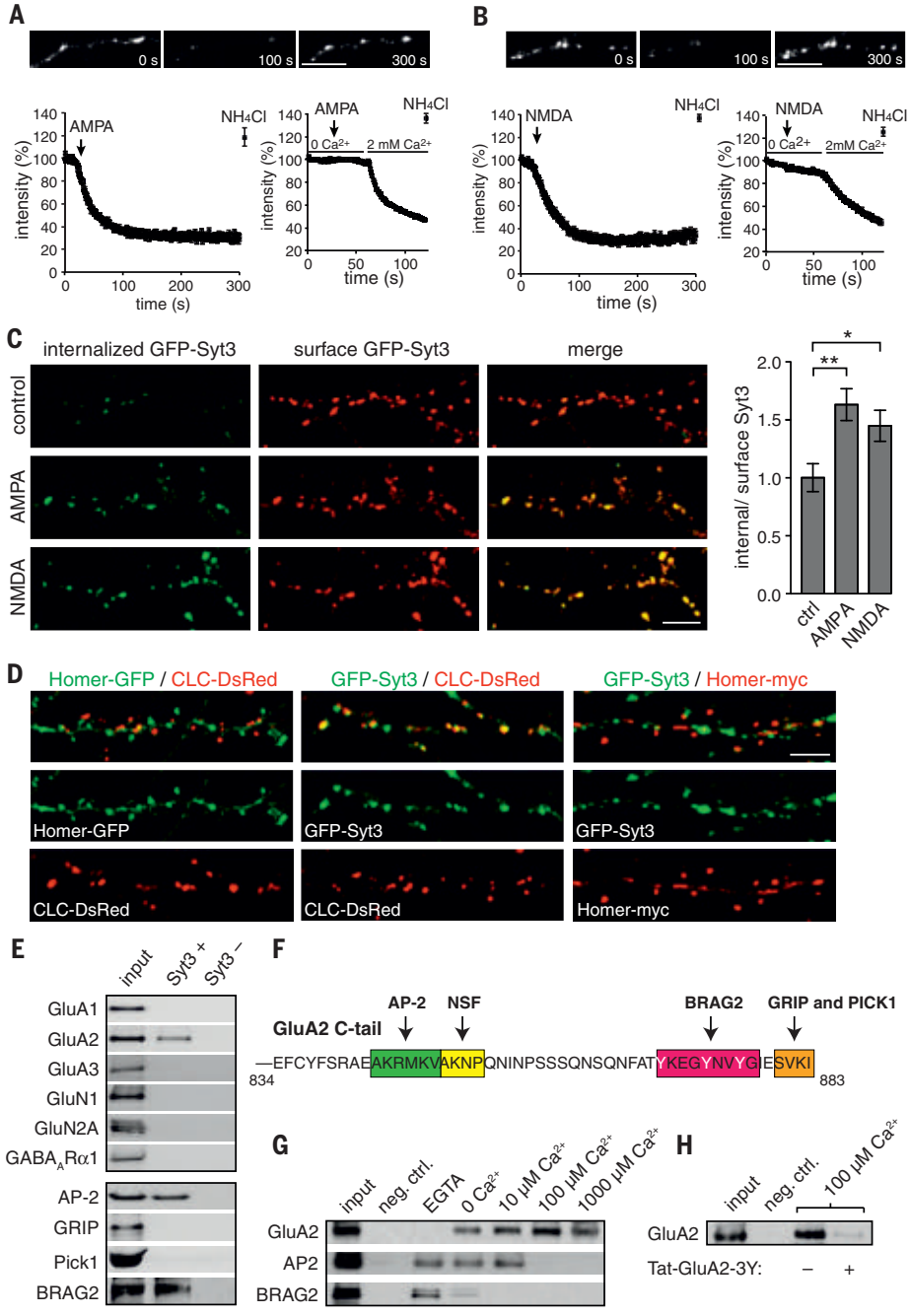


Fig. 2. Syt3 endocytoses in response to stimulation and binds GluA2, AP-2, and BRAG2. (**A**) pHluorin-Syt3 endocytoses in response to stimulation with AMPA or NMDA (**B**) in 2 mM calcium (left), but not in the absence of calcium (right) [*n*; ROI/neurons/cultures, AMPA 2 mM Ca^{2+} 37/7/3 (left); 0 and 2 mM Ca^{2+} 42/6/3 (right); NMDA, 2 mM Ca^{2+} 35/3/3 (left); 0 and 2 mM Ca^{2+} 40/3/3 (right)]. NH_4Cl dequenches internalized pHluorin-Syt3 fluorescence. Scale bars, 5 μm. (**C**) GFP-Syt3 expressed in Syt3 knockout hippocampal neurons internalizes in response to stimulation with AMPA and NMDA (*n* = 28, 25, and 26 neurons per 2 cultures for control, AMPA, and NMDA, respectively; one-way ANOVA, Tukey's test). (**D**) GFP-Syt3 colocalizes with CLC (clathrin light chain)–Dsred at postsynaptic endocytic zones, and not with Homer-myc at PSDs (*n*; ROI/neurons, 34/20 Homer-GFP/CLC-DsRed; 29/19 GFP-Syt3/CLC-Dsred; 34/18 GFP-Syt3/Homer-myc from three cultures). Scale bars, 5 μm. (**E**) Recombinant Syt3 C2AB pulls down GluA2, AP2, and BRAG2 from brain homogenates. Syt3– is beads only. (**F**). Binding sites, on the GluA2 C-terminal tail, of proteins implicated in receptor recycling and tested in pull downs. (**G**) Calcium-dependence of Syt3 C2AB pull down of GluA2, AP2, and BRAG2 from brain homogenate. In 0 Ca^{2+} conditions, brains were homogenized and solubilized in calcium-free buffer, but endogenous buffered calcium in internal stores remains. Addition of EGTA eliminates any potential effects of these additional sources of calcium. (**H**) Western blot of GluA2 pulled down from brain homogenate by Syt3 C2AB in 100 μM calcium with or without preincubation with 1 μM Tat-GluA2-3Y peptide. Negative controls are beads only.

receptors [Fig. 3, A to C, and fig. S4, A to D; Syt3 short hairpin RNA (shRNA) knockdown validation is provided in fig. S4, E and F]. Manipulation of Syt3 did not affect surface versus internal receptors in basal conditions. Syt3 knockout neurons yielded similar results; GluA2 receptors were internalized in response to AMPA and NMDA stimulation in wild-type neurons but not in Syt3 knockout neurons or in wild-type neurons treated with the Tat-GluA2-3Y peptide (Fig. 3, D and E, and fig. S4G).

Syt3 does not affect basal transmission

We found no change in miniature excitatory postsynaptic current (mEPSC) amplitude, frequency, or decay time in Syt3 overexpressing or knockdown neurons compared with GFP-transfected controls (decay time GFP, 2.3 ± 0.2 ms; GFP-Syt3, 2.1 ± 0.2 ms; Syt3 KD, 1.9 ± 0.2 ms) (fig. S5, A to D), further confirming that postsynaptic Syt3 does not affect surface receptor number in basal conditions. To exclude presynaptic effects, we also compared mEPSCs in wild-type and Syt3 knockout cultures and again found no difference in mEPSC amplitude or frequency (frequency $P = 0.2671$; t test, Welch's correction) (fig. S5, E to G). After stimulation, however, mEPSC amplitude decreased in wild-type but not in Syt3 knockout neurons, and this effect was rescued by reintroducing GFP-Syt3 postsynaptically into knockout neurons (fig. S5H), confirming a lack of stimulation-induced internalization of synaptic AMPA receptors in Syt3 knockouts.

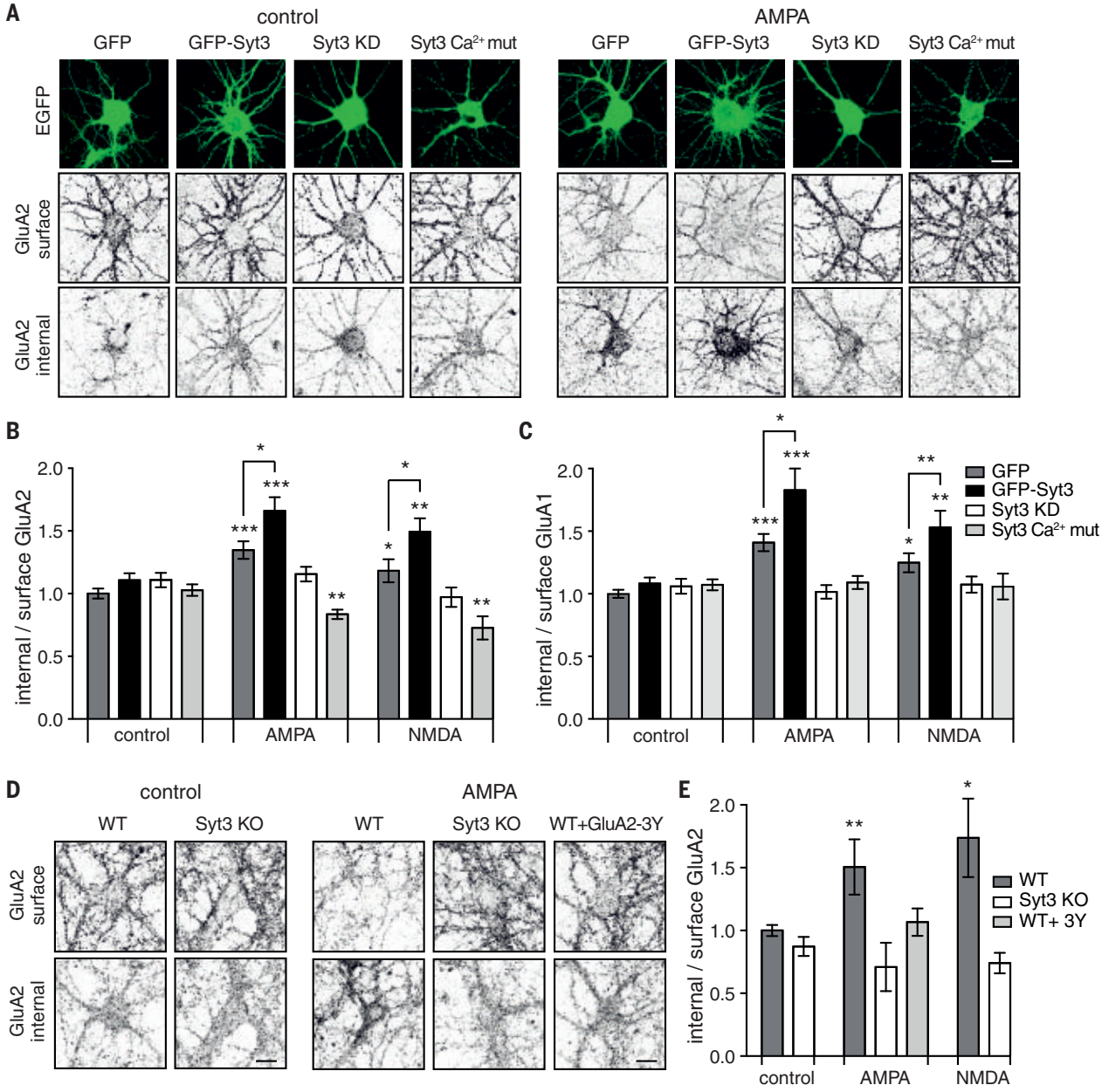


Fig. 3. Syt3 internalizes AMPA receptors in response to stimulation. (**A**) Hippocampal neurons transfected with GFP, GFP-Syt3, Syt3 knockdown (Syt3 KD), or Syt3 calcium-binding deficient mutant (Syt3 Ca^{2+} mut) constructs in control and AMPA-stimulated conditions, immunostained for surface and internalized GluA2 receptors. (**B**) Stimulation-induced GluA2 and GluA1 (**C**) internalization is increased by overexpression of GFP-Syt3 and blocked by Syt3 KD or Syt3 Ca^{2+} mut (*n*; neurons for GluA2/GluA1; GFP, 64/69 ctrl, 68/46 AMPA, 28/47 NMDA; GFP-Syt3, 46/65 ctrl, 46/47 AMPA, 23/43 NMDA; Syt3 KD, 51/41 ctrl, 69/38 AMPA, 23/24 NMDA; Syt3 Ca^{2+} mut, 42/37 ctrl, 28/36 AMPA, 27/31 NMDA from six cultures). (**D** and **E**) Stimulation-induced GluA2 internalization is abolished in Syt3 knockout neurons, and in WT neurons treated with the Tat-GluA2-3Y peptide (*n*; Syt3 KO/WT neurons, 31/29 ctrl, 24/22 AMPA, 18/17 NMDA; WT+GluA2-3Y, 18 AMPA, from four cultures). Scale bars, 10 μm; two-way ANOVA, Dunnett's test.

Basal synaptic transmission was also unchanged in Syt3 knockouts. Recordings of synaptic responses from CA1 neurons, elicited by Schaffer collateral stimulation in acute hippocampal slices, displayed no change in NMDA/AMPA, γ-aminobutyric acid (GABA)/AMPA, or GABA/NMDA receptor–mediated response ratio in Syt3 knockout compared with wild-type neurons (fig. S5, I to L). There was also no change in the AMPA EPSC decay time (wild type, 19.4 ± 4 ms; knockout, 21.9 ± 4.6 ms) or GABA inhibitory postsynaptic current (IPSC) decay time (wild type, 76 ± 4.9 ms; knockout, 58.8 ± 7 ms). In addition, we found no change in stimulus intensity versus fEPSP (field excitatory postsynaptic potential) slope or fiber volley amplitude (fig. S5, M and N) and no change in paired pulse ratio (fig. S5O), suggesting that Syt3 does not affect the number of synaptic inputs from CA3 onto CA1 neurons or short-term presynaptic plasticity.

Syt3 promotes LTD and decay of LTP

LTD—which requires activity-dependent endocytosis of AMPA receptors—was abolished in Syt3 knockouts (Fig. 4A), which is consistent with defective activity-dependent AMPA receptor endocytosis in the absence of Syt3. In wild-type slices, application of the GluA2-3Y peptide, which disrupts Syt3 binding to the GluA2 cytoplasmic tail, mimicked the Syt3 phenotype and abolished LTD (Fig. 4A) (*8*, *11*, *35*).

The decay of LTP induced by means of a 1XTET stimulation (a single tetanus of 16 pulses at 100 Hz) also depends on receptor internalization (*13*). LTP induced by means of a 1XTET stimulus decayed within 1 hour in wild-type slices (*36*) but was reinforced and remained potentiated in Syt3 knockouts (Fig. 4B); this form of LTP was NMDA receptor–dependent in both wild-type and Syt3 knockout slices (fig. S6, A and B). The GluA2-3Y peptide again mimicked the Syt3 knockout phenotype and reinforced decaying LTP (Fig. 4C). This reinforced decaying LTP was insensitive to ZIP (PKMζ inhibitory peptide) (Fig. 4C), which blocks the activity of atypical protein kinases and has the unusual property of reversing LTP after induction, through endocytosis of AMPA receptors (*12*, *13*, *37*, *38*). Reinforced decaying LTP in Syt3 knockouts was similarly ZIP-insensitive (Fig. 4D). The GluA2-3Y peptide had no effect on the reinforced decaying LTP in Syt3 knockouts, confirming that Syt3 decays LTP by acting on the GluA2-3Y region (Fig. 4D).

Nondecaying LTP, induced by means of 3XTET stimulation (three tetanizing trains of 100 pulses at 100 Hz), was unchanged in Syt3 knockout hippocampal slices compared with wild-type slices (Fig. 4E). However, ZIP promoted decay of 3XTET-induced LTP in wild-type slices but not in Syt3 knockout slices (Fig. 4F), further indicating a defect in activity-dependent receptor internalization in Syt3 knockouts. Because ZIP had no effect on LTP in Syt3 knockout slices, it did not appear to cause excitotoxicity or neural silencing (*39*, *40*). Analysis of stimulation frequency–potentiation dependence revealed increased potentiation in Syt3 knockouts compared with wild-type slices (Fig. 4G).

To test whether calcium-sensing by postsynaptic Syt3 is important for receptor internalization in these plasticity paradigms, we injected wild-type or calcium-binding–deficient mutant Syt3 AAV1/2 postsynaptically into the dorsal CA1 region of the hippocampus of mice in which Syt3 had been knocked out (Syt3 knockout mice) (Fig. 4H). We found that LTD (Fig. 4I), decaying 1XTET-induced LTP (Fig. 4J), and ZIP-mediated decay of 3XTET LTP (Fig. 4K) were rescued by expression of wild-type Syt3 but not calcium-binding–deficient Syt3.

Syt3 knockout mice learn normally but have impaired forgetting

Because Syt3 knockout mice had normal LTP but lacked LTD, we hypothesized that they would learn normally but have deficits in forgetting. To test this, we used the water maze spatial memory task that involves the CA1 hippocampal subregion, in which mice learn to navigate to a hidden platform position over sequential days of training using visual cues. Syt3 knockout mice showed no difference in anxiety or hyperactivity (fig. S7, A to C) but had a 15% decrease in body weight and faster swim speed compared with those of wild-type mice (fig. S7, D and E). We therefore plotted proximity to the platform in addition to the traditional escape latency parameter to control for possible effects of swim speed. The maximum average proximity difference—between closest (after training of all mice to a platform position) and farthest possible proximity (by using the same data but calculating proximity to a platform in the opposite quadrant)—was 12.5 cm. Syt3 knockout and wild-type mice learned the position of the hidden platform equally well; there was no significant difference in proximity to the platform (Fig. 5A) or escape latency (fig. S7F) during training. When the platform was removed in probe tests after training, to test how well the mice have learned the platform position, the percentage of time spent in the target quadrant was well above chance level for both Syt3 knockout and wild-type mice (Fig. 5B), and proximity to the platform position (Fig. 5C) and platform position crossings (Fig. 5D) were similar. Syt3 knockout mice in cohort 1 (of two cohorts tested) had significantly lower proximities and more platform crossings in the first probe test compared with those of wild-type mice but no difference in the second probe test. Thus, Syt3 knockout mice learned the task as well or slightly faster than did wild-type mice.

We observed an indication of a lack of forgetting in the second probe test after learning. Syt3 knockout mice exhibited a lack of "within-trial" extinction. Wild-type mice showed maximal searching (proximity to the platform) in the first 10 to 20 s of the probe test and then gradually shifted to a dispersed search pattern of other regions of the pool (*41*), whereas Syt3 knockout mice continued to persevere to the platform position even near the end of the trial (Fig. 5E). When the platform was shifted to the opposite quadrant in reversal training, Syt3 knockout mice learned the new platform position as well as did wild-type mice in the probe test (probe test 3) (Fig. 5, B, C, and D) but continued to persevere to the original platform position during reversal training (fig. S7G) and in the probe test after reversal (Fig. 5F), exhibiting a lack of forgetting of the previous platform position.

In a fourth cohort, we extended the training days after reversal (fig. S7H). Syt3 knockout mice persevered to the previous platform position significantly more than did wild-type mice, even after 7 days of reversal training (Fig. 5G). Injection of Syt3 AAV1/2 specifically into the dorsal CA1 (postsynaptic) region of the hippocampus of Syt3 knockout mice rescued this perseverence phenotype, compared with Syt3 knockout or wild-type mice injected with control GFP AAV1/2 (Fig. 5G).

To test whether the lack of forgetting exhibited by Syt3 knockouts is due to defective AMPA receptor internalization, we trained wild-type and Syt3 knockout mice to learn a platform position in the water maze, during habituation to daily saline intraperitoneal injections (fig. S7I). We then injected the Tat-GluA2-3Y peptide (5 μmol/kg, intraperitoneally)—which disrupts Syt3:GluA2 binding and blocks AMPA receptor internalization (*11*–*13*, *35*)—1 hour before the probe test after training to the initial platform position, and then daily, 1 hour before reversal training to a new platform position in the opposite quadrant. In the probe test after training to the initial platform position, peptide-injected wild-type and Syt3 knockout mice showed a similar lack of within-trial extinction and continued to persevere to the platform position, whereas saline-injected wild-type mice shifted to a dispersed search pattern near the end of the trial (Fig. 5H and fig. S7J). Similarly, in probe tests after reversal training to a new platform position, wild-type mice injected with the Tat-GluA2-3Y peptide persevered to the original platform position significantly more than did saline-injected wild-type mice [$P = 0.042$; one-way analysis of variance (ANOVA), Bonferroni's correction] or uninjected wild-type mice in previous cohorts ($P = 0.039$) and to a similar extent as did Tat-GluA2-3Y peptide–injected Syt3 knockout mice ($P = 0.096$) or uninjected Syt3 knockout mice in previous cohorts ($P = 0.105$) (Fig. 5I). Thus, Tat-GluA2-3Y peptide injection mimics the Syt3 knockout phenotype. There was no difference in perseverence between Tat-GluA2-3Y–injected and uninjected Syt3 knockouts in previous cohorts ($P = 0.184$), indicating that the effect of Tat-GluA2-3Y peptide injection is occluded in Syt3 knockout mice.

To further test whether the lack of forgetting phenotype of Syt3 knockouts is due to a lack of receptor internalization via Syt3:GluA2 interaction, we tested spatial memory in the Barnes maze, in which mice learn to navigate to 1 of 20 holes around the perimeter of a circular platform leading to an escape cage, using visual cues. Time spent in the target hole area during training and

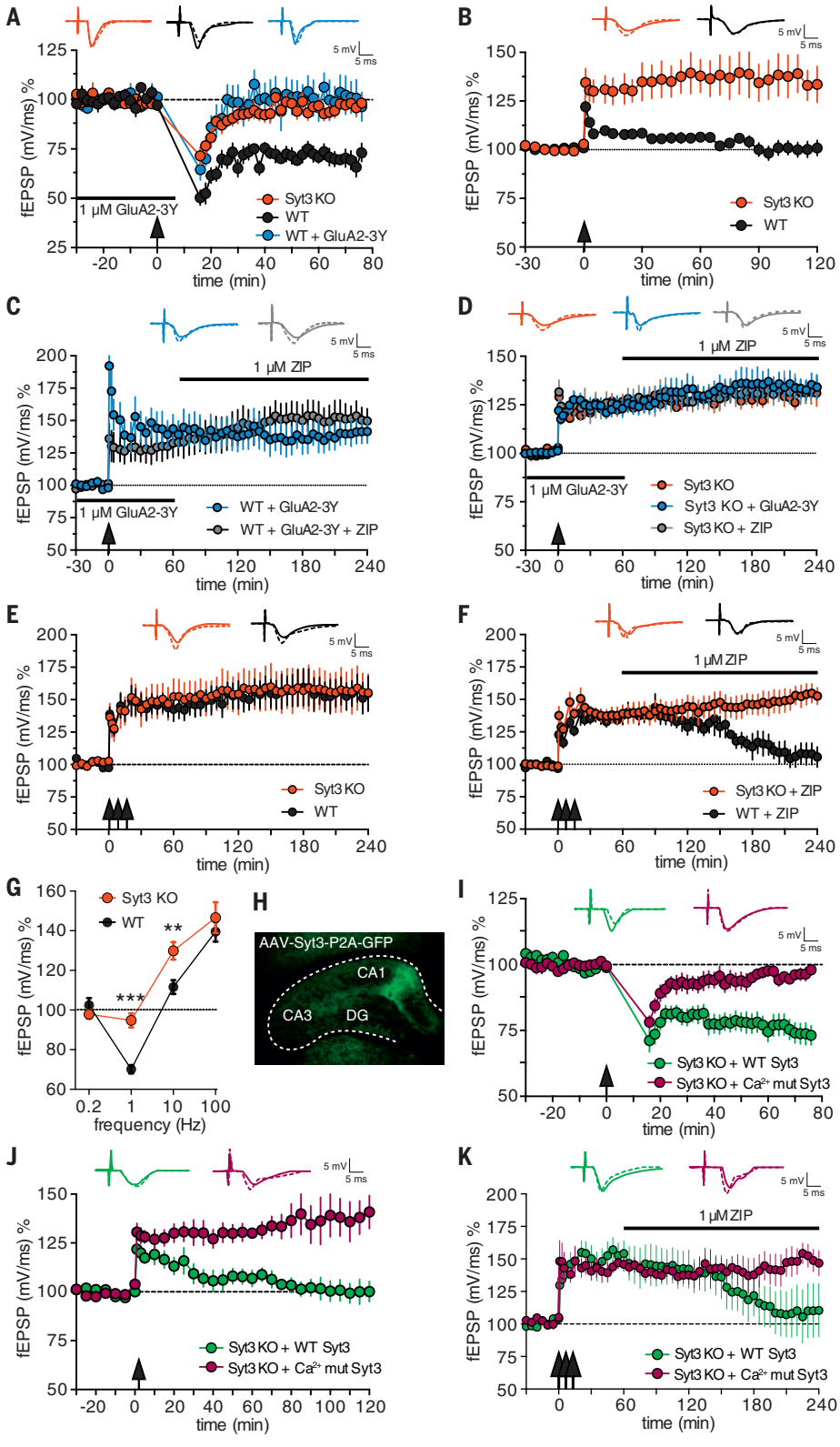


Fig. 4. LTD and decay of LTP is abolished in Syt3 knockouts. (**A**) LTD is abolished in Syt3 KO, and Tat-GluA2-3Y peptide treated WT hippocampal slices (*n*; slices/mice, 12/8 Syt3 KO, 8/6 WT, and 8/5 WT + Tat-GluA2-3Y); ***$P < 0.001$ for Syt3 KO/WT; **$P < 0.05$ for WT +GluA2-3Y/WT. (**B**) 1XTET-induced LTP is reinforced in Syt3 KO slices compared with WT ($n = 10/7$ Syt3 KO, 10/10 WT); **$P < 0.05$. (**C**) 1XTET-induced LTP in WT slices treated with the Tat-GluA2-3Y peptide is reinforced and ZIP-insensitive ($n = 7/7$). (**D**) Reinforced 1XTET-induced LTP in Syt3 KOs is unchanged by the Tat-GluA2-3Y peptide and is ZIP-insensitive ($n = 6/6$ Syt3 KO, 7/7 Syt3 KO + GluA2-3Y peptide, 7/7 Syt3 KO + ZIP). (**E**) 3XTET-induced LTP is unchanged in Syt3 KO compared with WT ($n = 6/6$). (**F**) 3XTET-induced LTP in Syt3 KOs is ZIP-insensitive ($n = 6/6$); **$P < 0.01$. (**G**) fEPSP changes in Syt3 KO and WT slices induced by different stimulation frequencies, recorded 30 min after stimulation [Syt3 KO/WT, $n = 9/12$ (0.2 Hz), 9/12 (1 Hz), 12/7 (10 Hz), and 14/14 (100 Hz)]. (**H**) GFP fluorescence in a hippocampal slice from a mouse injected with AAV1/2 Syt3-P2A-GFP in the dorsal CA1 region. (**I**) LTD in Syt3 KOs is rescued by hippocampal CA1 AAV1/2 injection of WT Syt3 ($n = 13/11$) but not calcium-binding deficient Syt3 (Ca^{2+} mut) ($n = 8/6$); **$P < 0.01$. (**J**) Decay of 1XTET-induced LTP in Syt3 KOs is rescued by WT ($n = 8/6$) but not Ca^{2+} mut Syt3 ($n = 7/6$ mice); **$P<0.01$. (**K**) ZIP-mediated decay of 3XTET-induced LTP in Syt3 KOs is rescued by WT ($n = 5/5$) but not Ca^{2+} mut Syt3 ($n = 6/6$); *$P < 0.05$; Mann-Whitney U test or Kruskal-Wallis test with Dunn's test for multiple comparisons; *n* = slices/mice.

in subsequent probe tests gives a readout of spatial memory. We tested four groups simultaneously: wild-type mice and Syt3 knockout mice, injected with saline or GluA2-3Y peptide, during "reversal" training to a new hole after initial learning of an original hole position. We predicted that (i) Syt3 knockout mice would exhibit a lack of forgetting—persevere more to the original hole after reversal as compared with wild-type; (ii) disruption of Syt3:GluA2 interaction through injection of the GluA2-3Y peptide in wild-type mice would mimic the lack of forgetting phenotype of Syt3 knockouts; and (iii) injection of the GluA2-3Y peptide in Syt3 knockout mice would have no effect on their lack of forgetting.

In initial training, all four groups (injected intraperitoneally daily with saline only, 1 hour before training) learned the target hole equally

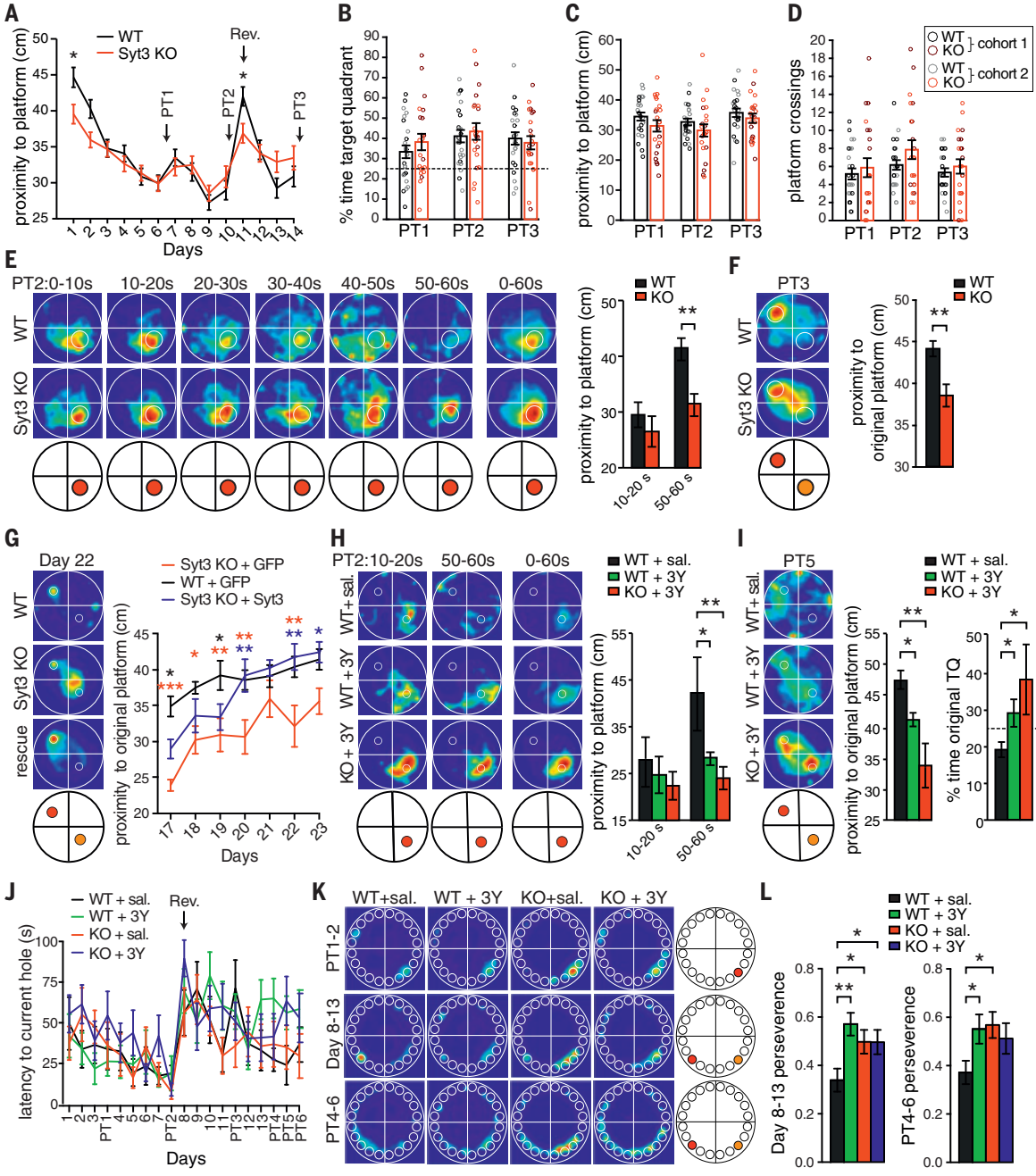


Fig. 5. Syt3 knockout mice learn as well as wild-type mice but have impaired forgetting. (**A**) Syt3 KO and WT mice had similar proximity to the platform during training (genotype effect, $P = 0.1052$; two-way ANOVA, Bonferroni's test) and in probe tests exhibited similar percent of time in (**B**) target quadrant, (**C**) proximity to platform, and (**D**) platform crossings. KOs in cohort 1 had lower proximities than those of WT in probe tests ($P = 0.027$; Student's t test). (**E**) Syt3 KO mice lack within-trial-extinction in time-binned average occupancy plots of all mice and proximity to the platform position (red in schematics) in probe test 2. (**F**) Syt3 KO mice persevere to the previous platform position (orange) in the probe test after reversal training (to the red platform position) ($n = 9/13$ Syt3 KO and 10/14 WT mice for cohort 1/cohort 2); Student's t test, Welch's correction in (B), (C), (E), and (F) and Mann-Whitney U test in (D). (**G**) Expression of Syt3 in the hippocampal CA1 region rescues the perseverence phenotype of Syt3 KOs in training after reversal; two-way ANOVA, Tukey's test. Black, red, and blue asterisks are significance between WT+GFP/Syt3 KO+Syt3, Syt3 KO+GFP/WT+GFP, and Syt3 KO+GFP/Syt3 KO+Syt3, respectively ($n = 10$ Syt3 KO+Syt3, 7 Syt3 KO+GFP, and 15 WT+GFP). (**H**) Injection of the Tat-GluA2-3Y peptide mimics the lack of within-trial-extinction in probe test 2 after initial training, and (**I**) perseverence to the previous platform position in probe tests after reversal training, exhibited by Syt3 knockouts ($n = 8$ WT saline, 7 WT+3Y, and 10 Syt3 KO+3Y); one-way ANOVA, Bonferroni's correction. (**J**) All mice learned (decreased latency to) the escape hole in the Barnes maze equally well during training (two-way ANOVA, Tukey's test) and (**K**) (top) in probe tests 1 and 2 after training. (Middle) In reversal training and (bottom) probe tests 4 to 6 after reversal training, WT saline mice learned the new hole position (red), but WT+3Y, KO saline, and KO+3Y mice persevered to the previous hole position (orange). (**L**) WT+3Y, KO saline, and KO+3Y mice had a higher perseverence ratio [time spent exploring original hole (orange) divided by total time spent exploring original and reversal (red) holes] as compared with that of WT saline mice ($n = 10$ WT saline, 11 WT+3Y, 11 Syt3 KO saline, 12 Syt3 KO+3Y; $P = 0.060$ for WT saline/Syt3 KO+3Y in probe tests 4 to 6); one-way ANOVA, Bonferroni's correction. All occupancy plots show average search path densities across all mice in a group.

well (Fig. 5J) and spent the majority of their time exploring the target hole region in probe tests (Fig. 5K, top). Beginning on the day of probe test 2 and during reversal training, in which the target hole was positioned in a different quadrant, mice were injected with saline or GluA2-3Y peptide 1 hour before training each day. Saline-injected Syt3 knockout mice and GluA2-3Y-injected wild-type and Syt3 knockout mice all showed a similar increased perseverence to the original hole region as compared with wild-type saline-injected mice throughout reversal training (Fig. 5K, middle) and in probe tests after reversal training (Fig. 5, K, bottom, and L), which is in agreement with our prediction.

Syt3 knockouts have impaired working memory owing to lack of forgetting

To further test deficits in forgetting exhibited by Syt3 knockout mice, we examined working memory in the delayed matching to place (DMP) water maze task, in which the platform position is moved to a new position each day (*42*). Each day, the mice have four trials to learn the new platform position and forget the previous platform position. We first trained the mice to a visible platform for 3 consecutive days, in which Syt3 knockout and wild-type mice performed equally well (fig. S8A). Because swimming speed of Syt3 knockout mice was again faster than that

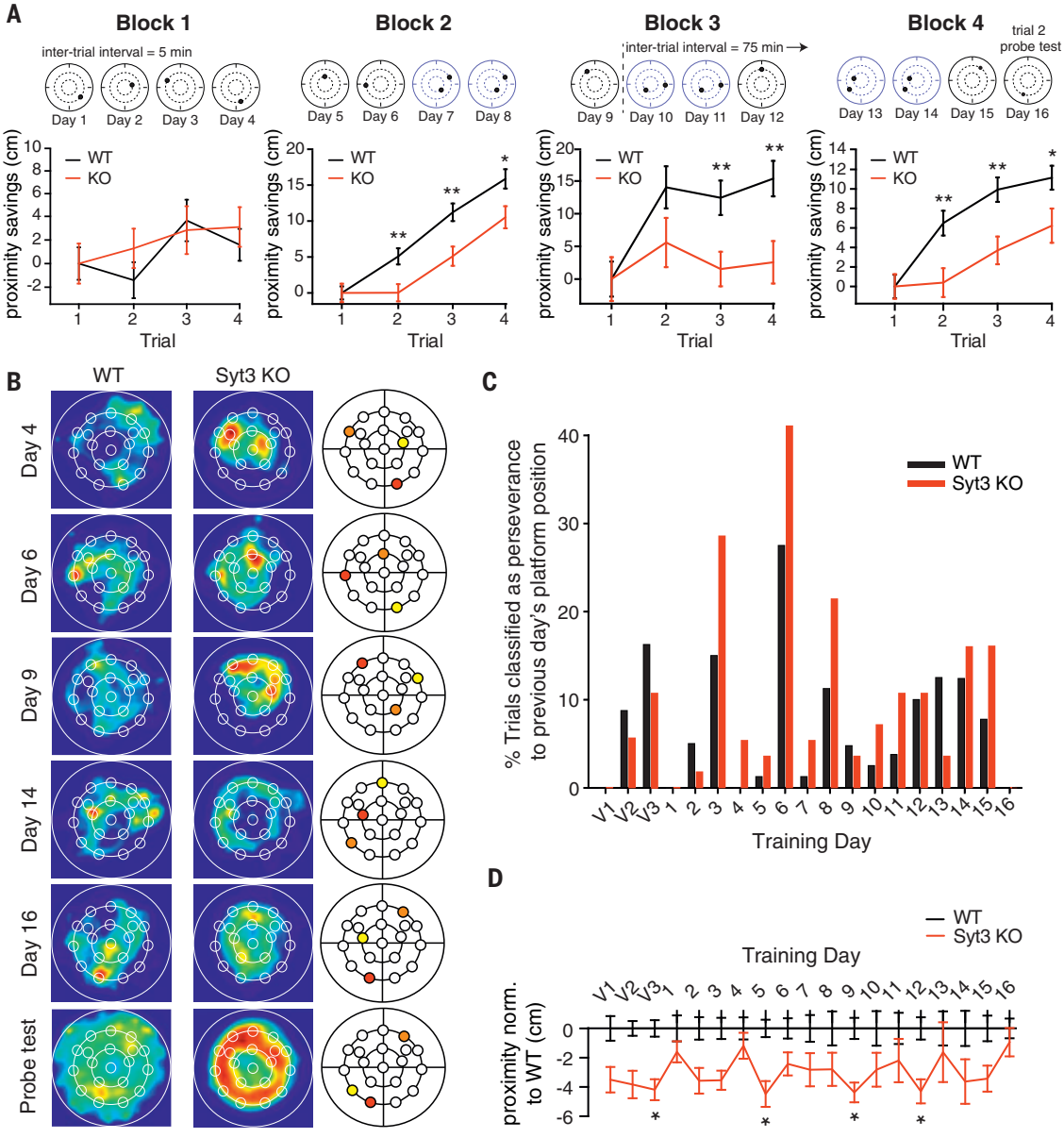


Fig. 6. Syt3 knockout mice show deficits in the delayed matching to place task because of impaired forgetting. (**A**) WT mice had a closer proximity to the platform in trials 2 to 4 relative to trial 1 (higher proximity savings) as compared with that of Syt3 KO mice, indicating that Syt3 KOs had deficits in finding the platform when it appeared in a new position each day. Hidden platform positions presented each day are indicated above the graphs. Blue-outlined mazes indicate counter-balancing, in which half the cohort is trained to one of the positions and the other half is trained to the other position, to avoid biased search of inner- or outer-platform positions (n = 14 Syt3 KO and 20 WT mice); Student's t test. (**B**) Occupancy plots of individual trials averaged across all mice on training days, and in the probe test after training, reveal impaired forgetting in Syt3 KO mice—higher perseverence to previous days' platform positions as compared with that of WT. In maze schematics, the current day's platform is red, the previous day's is orange, and the platform position 2 days previous is yellow. (**C**) Syt3 KOs have significantly more perseverence trials compared with those of WT in strategy analysis (P = 0.0383; unpaired t test, Welch's correction). (**D**) In the probe test on day 16, Syt3 KO mice persevere more (have closer proximity as compared with WT) to all previous positions. V1, V2, and V3 indicate visible platform training days; two-way ANOVA for genotype effect, $P < 0.0001$; Bonferroni's test, $^*P < 0.05$.

of wild-type mice (fig. S8B), we quantified proximity savings in hidden platform training (Fig. 6A). In the first 4 days of training, mice became accustomed to the task. In the second 4-day block, mice had learned the task, and wild-type mice performed significantly better than Syt3 knockouts. In the third block, from day 10 onward, the intertrial interval between trial 1 and trial 2 was increased from 5 to 75 min, and wild-type mice continued to perform better than Syt3 knockout mice; this difference was most pronounced in the fourth 4-day block of training (Fig. 6A). Quantitation of escape latency savings (fig. S8D) and path savings (fig. S8E) yielded similar results. To test whether the poor performance of Syt3 knockouts was due to difficulty forgetting previous platform positions, we examined perseverance to previous positions. Occupancy plots indicated that Syt3 knockouts indeed persevered to previous platform positions more than did wild-type mice throughout training and sampled most previous platform positions in the probe test, whereas wild-type mice adopted a more random search pattern (Fig. 6B). Syt3 knockout mice had significantly more trials classified as perseverence (*43*) to the previous day's platform position as compared with wild-type mice across all days of training (P = 0.0383, unpaired t test, Welch's correction) (Fig. 6C). Neither Syt3 knockout nor wild-type mice exhibited chaining (*43*), a nonspatial search strategy for platform positions at two fixed distances from the wall (fig. S8C). Syt3 knockout mice also had a significantly closer proximity to all previous platform positions in the probe test at the end of the 16-day training period, compared with that of wild-type mice (Fig. 6D).

Discussion

We found that activity-dependent AMPA receptor internalization, LTD, decay of LTP, and forgetting of spatial memories requires Syt3. In rescue experiments, calcium-sensing by postsynaptic Syt3 was required for LTD and the decay of LTP. The GluA2-3Y peptide competitively inhibited Syt3 binding to the 3Y region of the GluA2 receptor tail. Introduction of this peptide in a wild-type background mimicked the Syt3 knockout phenotypes of lack of activity-dependent AMPA receptor internalization, LTD, decay of LTP, and forgetting. Effects of the peptide were occluded in Syt3 knockout mice, implicating Syt3 in a GluA2-3Y–dependent mechanism of AMPA receptor internalization.

Our data give rise to a model in which Syt3 at postsynaptic endocytic zones is bound to AP-2 and BRAG2 in the absence of calcium. GluA2 could then accumulate at endocytic zones by binding Syt3 in response to increased calcium during neuronal activity. This would potentially bring GluA2 into close proximity to BRAG2, where a transient interaction could activate BRAG2 and Arf6, and promote endocytosis of receptors via clathrin and AP-2 (*10*, *32*). PICK1 is also important for AMPA receptor endocytosis, raising the question of the interplay of Syt3 and PICK1. Two possible mechanisms are conceivable. PICK1 could sequester receptors internally after endocytosis (*44*) and act downstream of Syt3. Alternatively, PICK1 could transiently bind GluA2 and then AP-2 after stimulation, to cluster AMPA receptors at endocytic zones and promote their subsequent internalization (*45*). Thus, it is also possible that PICK1 acts upstream of or in concert with Syt3 to bring GluA2 to endocytic zones.

Blockade of postsynaptic expression of Syt1/Syt7 was recently reported to abolish LTP but have no effect on LTD (*18*). Thus, distinct Syts may insert and remove receptors from the postsynaptic membrane to mediate LTP and LTD, respectively. Syts may regulate both exo- and endocytosis (*46*) but be predisposed to one or the other depending on their calcium affinity. Endocytosis occurs on slower time scales than does exocytosis. As calcium concentration declines, high-affinity Syts such as Syt3 (*22*) may remain active, whereas low-affinity Syts such as Syt1 inactivate. Alternatively, Syts predisposed to endocytosis may interact with protein complexes that extend association with Ca^{2+} or bind proteins in resting conditions that are released upon Ca^{2+} binding to cause endocytosis.

Although the ability to remember is often regarded as the most important aspect of memory, forgetting is equally important. Deficits in forgetting can have severe consequences and lead to posttraumatic stress disorder, for example. Our data identify Syt3 as a molecule important for a forgetting mechanism by which AMPA receptors are internalized to promote LTD and decay of LTP.

Materials and methods

Animals

Use of mice for experimentation was approved and performed according to the specifications of the Institutional Animal Care and Ethics Committees of Göttingen University (T10.31), and German animal welfare laws. Animal sample size was estimated by experimental literature and power analysis (G* Power version 3.1) to reduce animal number where possible while maintaining statistical power. The Syt3 and Syt6 knock-out and Syt5 and Syt10 knock-in quadruple targeted mutation mice (B6;129-Syt6^{tm1Sud} Syt5^{tm1Sud} Syt3^{tm1Sud} Syt10^{tm1Sud}/J, stock no. 008413) were obtained from The Jackson Laboratory (www.jax.org/strain/008413). These mice were then crossed with Black6/J mice, obtained from Charles River, to isolate mice with homozygous knock-out alleles of Syt3 but WT alleles of Syt5, Syt6 and Syt10. The genotypes of all breeder pairs and mice used for behavioral experiments were confirmed by PCR.

Dissociated hippocampal culture and neuronal transfection

Rat hippocampi were isolated from E18-19 Wistar rats as described previously (*47*). Hippocampi were treated with trypsin (Sigma) for 20 min at 37°C, triturated to dissociate cells, plated at 40,000 cells/cm^2 on poly-D-lysine (Sigma)–coated coverslips (Carolina Biologicals), and cultured in Neurobasal medium supplemented with 2% B-27 and 2 mM Glutamax (Gibco/Invitrogen). Cultures were fixed at 12 to 18 DIV with 4% paraformaldehyde and immunostained with primary antibodies, followed by secondary labeling with Alexa dyes for confocal imaging.

Hippocampi from P0 mouse brains were dissected in ice cold dissection medium (HBSS (Gibco), 20 mM HEPES (Gibco), 1.5 mM $CaCl_2$, 10 mM $MgCl_2$, pH adjusted to 7.4 with NaOH) and then incubated in a papain digestion solution (dissection medium, 0.2 mg/ml L-Cysteine, 0.5 mM NaEDTA, 1 mM $CaCl_2$, 3 mM NaOH, 1% Papain equilibriated in 37°C and 5% CO_2 + 0.1 mg/ml DNAseI) for 30 min at 37°C and 5% CO_2. Papain was inactivated with serum medium [DMEM, 2 mM Glutamax, 5% FBS, 1X Mito$^+$ supplements (VWR), 0.5X MEM vitamins (Gibco)] + 2.5 mg/ml BSA + 0.1 mg/ml DNAseI, followed by 3 washes with serum medium. After trituration in serum medium, the cell suspension was centrifuged for 5 min at 500 × g at room temperature. The cell pellet was resuspended in plating medium (Neurobasal medium, 100 U/ml Pen/Strep, 2% B27, 0.049 mM L-aspartate, 0.05 mM L-glutamate, 2 mM Glutamax, 10% serum medium). Cells were plated at 60,000 cells/cm^2 on 12 mm diameter acid-etched glass coverslips coated with 0.04% Polyethyleneimine (PEI; Sigma). Glial growth was blocked on DIV4 with 200 µM FUDR. 50% conditioned medium was replaced with feeding medium (Neurobasal, 2 mM Glutamax, 100 U/ml Pen/Strep, 2% B-27) on DIV7.

For transfection of neurons on 12 mm coverslips in 24-well plates, the culture medium from each well was exchanged with 400 µl Neurobasal medium; the original culture medium was stored at 37°C and 5% CO_2. 1 µg DNA/50 µl Neurobasal medium and 1 µl Lipofectamine 2000 (Gibco)/50 µl Neurobasal medium were incubated separately for 5 min, mixed, and incubated for 20 additional min, before adding the mixture to neurons in a well. Following incubation for 2 hours, neurons were washed once with Neurobasal, and conditioned media was replaced.

HEK cell culture, transfection, immunostaining, and Western blots for Syt3 knockdown validation

Human embryonic kidney (HEK) 293 cells (purchased from American Type Culture Collection CRL-3216; not tested for mycoplasma) cultured in DMEM supplemented with 10% FBS on 10 cm dishes were transfected at 60 to 80% confluence using the calcium-phosphate method: 20 µg plasmid DNA in 360 µl dH_2O was mixed with 40 µl 2.5 M $CaCl_2$, followed by addition of an equal volume (400 µl) of transfection buffer (280 mM NaCl, 1.5 mM Na_2HPO_4, 50 mM HEPES, pH 7.05). This mixture was incubated at room temperature for 20 min and 800 µl added per dish. 24 to 48 hours later cells were harvested for Western blots in lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM EDTA, 0.5% NP40, and protease inhibitors). For immunostaining, HEK cells growing on 12 mm coverslips were transfected with Lipofectamine 2000 (Invitrogen) by incubating 1 µg DNA/50 µl medium and 1 µl Lipofectamine 2000/50 µl medium separately for 5 min, mixing the two solutions together, and

then incubating for an additional 20 min, before adding the mixture to wells of a 24-well plate.

Immunohistochemistry

Acute hippocampal slices were prepared from 8-week-old mice anesthetized with isoflurane and decapitated. The hippocampus was removed and 200 μm thick slices were cut transversely using a tissue chopper (Stoelting) in ice-cold artificial cerebrospinal fluid (ACSF) containing, in mM 124 NaCl, 4.9 KCl, 1.2 KH_2PO_4, 2 $MgSO_4$, 2 $CaCl_2$, 24.6 $NaHCO_3$, and 10 D-glucose (saturated with 95% O_2 and 5% CO_2, pH 7.4, ~305 mOsm). Slices were fixed in 4% paraformaldehyde in PBS for 30 min and washed 3 X 20 min in PBS. After washing, slices were incubated in antibody buffer (2% donkey serum, 0.1% Triton X-100 and 0.05% NaN_3 in PBS) for 30 min at room temperature. Then slices were incubated with primary antibodies in antibody buffer overnight at 4°C. Slices were then washed with PBS 3 X 20 min and incubated with fluorescently tagged secondary antibodies for 2 hours at room temperature. Slices were washed 3 X 20 min in PBS and mounted on microscope slides with Fluoromount-G (Sigma) and sealed with nail polish. Images were collected using 10X air and 40X oil immersion objectives on a Zeiss A1 laser scanning confocal microscope with Zen software (Carl Zeiss). Digital images were processed using Adobe Photoshop software.

Antibodies and expression constructs

Antibodies used including species and catalog number were rabbit Syt3 105133 (Synaptic Systems) and mouse Syt3 N278/19 (NeuroMab), chick MAP2 C-1382-50 (Biosensis), rabbit GFP ab290 (Abcam), mouse Synaptophysin (Syp) 101011, rabbit Synaptobrevin2 (Syb2) 104202, mouse Syt1 105101, mouse SNAP25 111011, guinea pig Piccolo 142104, guinea pig Homer 160004, guinea pig Beta3-tubulin 302304, rabbit VGluT1 135303, mouse VGAT 131011, mouse Rab-GDI 130011, mouse Gephyrin 147011, rabbit GluA1 182003, mouse GluA2 182111, rabbit GRIP 151003 (Synaptic Systems), mouse $GABA_AR\alpha1$ 75136 (NeuroMab), rabbit GluA3 17311 (Epitomics), mouse PSD95 MA1-045 (Thermo Scientific), rabbit GluA1 PC246 (Calbiochem), mouse GluA2 MAB397, rabbit GluN1 AB9864, mouse GluN2A MAB5216 (Millipore), rabbit PICK1 PA1073 (Thermo Scientific), mouse Synapsin (Syn), mouse Rab3a, mouse Syntaxin1A, rabbit EEA1 (provided by Reinhard Jahn, Max Planck Institute for Biophysical Chemistry, Goettingen, Germany), rabbit Neurexin (Nrx), rabbit Neuroligin (Nlg) (provided by Nils Brose, Max Planck Institute of Experimental Medicine, Goettingen, Germany), mouse AP-2 (provided by Pietro De Camilli, Yale School of Medicine, New Haven, CT, USA), and rabbit BRAG2 (*10*) (provided by Hans-Christian Kornau, Charité University of Medicine, Berlin, Germany). Mammalian expression constructs used were pHluorin-Syt3, as previously described (*23*). GFP-Syt3 and mCherry-Syt3 were subcloned by replacing the pHluorin in pHluorin-Syt3 with GFP or mCherry, respectively. Homer-GFP, Homer-myc, and Clathrin light chain (CLC)-DsRed constructs were provided by Daniel Choquet (*30*) (University of Bordeaux). The calcium-binding deficient mutant of Syt3 was generated by Genscript by mutagenesis of D386, 388N and D520, 522 N, corresponding to the calcium-binding sites of syt1: D230, 232 N and D363, 365 N (*48*). Syt3 shRNA knockdown constructs, transfected in equal amounts, were KD1; TGCTGTTGACAGTGAGCGACAAGCTCATCGGTCAGATCAATAGTGAAGCCACAGATGTATTGATCTGACCGATGAGCTTGGTGCCTACTGCCTCGGA, KD2; TGCTGTTGACAGTGAGCGCAGGTGTCAAGAGTTCAACGAATAGTGAAGCCACAGATGTATTCGTTGAACTCTTGACACCTATGCCTACTGCCTCGGA and KD3; TGCTGTGACAGTGAGCCAGGATTGTCAGAGAAAGAGAATAGTGAAGCCACAGATGTATTCTCTTTCTCTGACAATCCTTTGCCTACTGCCTCGGA in a pGIPZ vector co-expressing turboGFP (Thermo Scientific Openbiosystems).

Receptor internalization assays

Neurons were labeled with primary extracellular antibodies against GluA1 (rabbit PC246 Calbiochem) or GluA2 (mouse MAB397 Millipore) in medium for 15 min at 37°C and 5% CO_2, washed for 2 min in medium, and then stimulated with 100 μM AMPA or NMDA for 2 min, followed by incubation in conditioned medium for an additional 8 min at 37°C and 5% CO_2, before fixing cells with 4% paraformaldehyde. Surface receptors were labeled with Alexa 647 secondary antibodies, then cells were permeabilized and internal receptors labeled with Alexa 546 secondary antibodies. The internalization index was calculated as the ratio of internal to surface receptor fluorescence. Cultures in which control conditions did not show an increase in internalization following stimulation were excluded from analysis.

pHluorin imaging

Neurons were transferred to a live imaging chamber (Warner Instruments) containing bath saline solution (140 mM NaCl, 5 mM KCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$, 5.5 mM glucose, 20 mM HEPES, 1 μM TTX, pH = 7.3). Transfected cells were selected, and images acquired at 1 s intervals and 500 ms exposure times, with 484/20nm excitation and 517/20-nm emission filters, on a Zeiss Axio Observer inverted microscope with a Photometric Evolve EMCCD camera, and Lambda DG-4 fast-switching light source interfaced with Metamorph software. A baseline of at least 30 images was collected before addition of high potassium buffer (100 mM NaCl, 45 mM KCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$, 5.5 mM glucose, 20 mM HEPES, 1 μM TTX, pH = 7.3), 100 μM AMPA, 100 μM NMDA, or field stimulation to depolarize neurons. For AMPA stimulation, 50 μM AP5 was included in the bath solution to block NMDA receptors; for NMDA stimulation, 20 μM CNQX was included to block AMPA receptors. Dendritic puncta were selected using Metamorph software (Molecular Devices) and fluorescence intensity normalized to baseline was plotted versus time.

Subcellular fractionation from whole brain

Rat brains were homogenized in ice-cold homogenization buffer (320 mM sucrose, 4 mM HEPES, pH 7.4 with NaOH) with 10 strokes at 900 rpm. Samples were then centrifuged at 1000 × g for 10 min. The supernatant (S1) was collected; the resulting pellet (P1) contains large cell fragments and nuclei. S1 was then centrifuged at 15,000 × g for 15 min. The supernatant (S2) contains soluble proteins and the pellet (P2) contains synaptosomes. The pellet was then carefully resuspended in 1 ml homogenization buffer, 9 ml ice-cold ddH_20 was added and three strokes at 2,000 rpm were performed. 50 μl 1M HEPES and protease inhibitors were added. The lysate was centrifuged at 17,000 × g for 25 min to separate synaptosomal membranes (LP2) from synaptosomal cytosol (LS2). The LP2 pellet was resuspended in 6 ml 40 mM sucrose and layered over a continuous sucrose gradient from 50 mM to 800 mM. The sucrose cushion was then centrifuged at 28,000 rpm for 2 hours. Following centrifugation, the region between approximately 200 mM and 400 mM sucrose was collected and separated by chromatography on controlled-pore glass beads (CPG column) and run overnight. The first peak (PI) contained larger membrane fragments and SVs were found in the second peak.

Immuno-organelle isolation of synaptic vesicles

Mouse monoclonal antibodies directed against Syb2 and Syt1 were coupled to Protein G magnetic Dynabeads (Invitrogen) in PBS for 2 hours at 4°C. Antibody-coated beads were added to whole brain S1 fractions and incubated overnight at 4°C. Magnetic beads were separated from immuno-depleted supernatant and washed 3 times with PBS. Bound vesicles were eluted in sample buffer and analyzed by SDS–polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblotting.

Syt3 pulldowns

Recombinant His-tagged Syt3 C2AB (provided by Edwin Chapman, University of Wisconsin, Madison) was expressed in *E. coli* and purified as previously described (*49*). Recombinant Syt3 was incubated with solubilized rat brain homogenate for 2 hours at 4°C. After incubation, Ni-beads were added and incubated for an additional 2 hours at 4°C. The mixture was then poured into a MT column from Biorad and washed 3 X with wash buffer (20 mM Tris pH 7.4, 500 mM NaCl, 20 mM imidazole). Proteins bound to recombinant Syt3 in the column were eluted with elution buffer (20 mM Tris pH 7.4, 500 mM NaCl, 400 mM imidazole). For experiments testing inhibition of Syt3:GluA2 binding, 1 μM Tat-GluA2-3Y peptide (sequence $YGRKKRRQRRR\text{-}_{869}YKEGYNVYG_{877}$, provided by Yu-Tian Wang, University of British Columbia, Vancouver, Canada) was pre-incubated with brain homogenate overnight prior to incubation with recombinant Syt3 and Ni-beads. Eluted proteins were loaded onto SDS-PAGE gels and analyzed by Western blot.

Synaptosome trypsin cleavage assay

Synaptosomes were prepared and treated with trypsin as described previously (*26*). Purified synaptosomes were centrifuged for 3 min at 8700 × g,

4°C. The pellet was resuspended in 320 mM sucrose, 5 mM HEPES, pH 8. For trypsin cleavage, a 0.1 mg/ml trypsin stock solution was added to yield a final protein-protease ratio of 100:1. Synaptosomes were incubated for 10, 20, 30, 60, or 90 min at 30°C with gentle agitation. This mixture was then centrifuged for 3 min at 8700 × g, and the resulting pellet resuspended in sucrose buffer containing 400 μM Pefabloc (Roche) to stop trypsin cleavage activity. Samples were then analyzed by SDS-PAGE and immunoblotting.

AAV preparation and hippocampal injections

AAV ESYN-GFP-P2A-Syt3 or calcium-binding deficient mutant Syt3 (D386, 388N, and D520, 522N) constructs were synthesized and subcloned by Genscript, and AAV1/2 viral particles prepared by transfecting 10 cm dishes of 70-80% confluent HEK293 cells with 12.5 μg of viral construct containing Syt3 or calcium-binding deficient mutant Syt3, with 25 μg pFdelta6, 6.25 μg pRV1 and 6.25 μg pH21 helper plasmids using calcium phosphate. Cell media was replaced after 6 hours and cells incubated for 48 hours prior to virus harvesting. Viral particles were released by adding 10% sodium deoxycholate and benzonase (Sigma-Aldrich), and purified in an Optiprep (Sigma-Aldrich) step gradient by ultracentrifugation for 90 min. The pure viral fraction was concentrated to approximately 500 μl through a 0.22 μm Amicon Ultra unit (Millipore/Merck), aliquoted and stored at −80°C. Virus was used at a titer of 10^7 particles/μl. Mice were given metamizol (3ml/L) in drinking water for 2 days prior to surgery and up to 3 days after. On the day of surgery, mice were anesthetized with a single intraperitoneal injection of ketamine/xylazine (80 and 10 mg/kg, respectively) and given a single subcutaneous injection of buprenorphine (0.1 mg/kg). Mice were fixed in a stereotaxic device (myNeuroLab Wetzlar, Germany), an incision was made to expose the bone and anteroposterior (−1.75mm) and mediolateral (±1.0 mm) coordinates from bregma were used for drilling holes bilaterally. A fine glass injection capillary filled with 1.2 μl virus was then slowly lowered to −1.5 (dorsoventral) and allowed to rest for 1 min. 0.9 μl was then injected over 3 min. The needle was allowed to rest in place for an additional minute after the injection and then slowly lifted above bone level. This procedure was repeated for both hippocampi. Mice were then removed from the stereotaxic device and tissue glue (Histoacryl, B Braun) was used to seal the wound. They were then allowed to recover on a heat blanket and transferred to individual cages for post-operation recovery. Mice were given buprenorphin injections up to 2 days after the operation and closely monitored for signs of distress or pain. Mice were only used for subsequent experiment after a minimum of 2 weeks post-operation.

Electrophysiological recordings from dissociated hippocampal neurons

Following transfection on DIV10, DIV13-20 rat neurons growing on coverslips were placed in a custom-made recording chamber in extracellular solution containing, in mM: 142 NaCl, 2.5 KCl, 10 HEPES, 10 D-Glucose, 2 $CaCl_2$, 1.3 $MgCl_2$ (295 mOsm, pH adjusted to 7.2 with NaOH). The temperature of the bath was maintained at 30 to 32°C. To record miniature excitatory postsynaptic currents (mEPSCs), 1 μM TTX (to block action potentials) and 50 μM picrotoxin (to inhibit $GABA_A$ receptors) was added to the extracellular solution. An Olympus upright BX51WI microscope equipped with a 40X water-immersion objective, fluorescent light source (Lumen 200Pro, Prior Scientific), and filters for GFP fluorescence imaging were used to visualize neurons transfected with GFP, GFP-Syt3 or Turbo GFP-expressing Syt3 knockdown constructs. Patch pipettes were pulled from borosilicate glass (1.5 mm OD, 0.86 mm ID, 3 to 6 megohms) using a P-97 micropipette puller (Sutter Instruments). The internal solution contained, in mM: 130 K-gluconate, 10 NaCl, 1 EGTA, 0.133 $CaCl_2$, 2 $MgCl_2$, 10 HEPES, 3.5 Na_2-ATP,1 Na-GTP (285 mOsm, pH adjusted to 7.4 with KOH). Whole-cell patch-clamp recordings were obtained using a HEKA EPC10 USB double patch clamp amplifier coupled to Patchmaster acquisition software. Signals were low pass filtered using a Bessel filter at 2.9 KHz and digitized at 5 KHz. mEPSCs were recorded while holding neurons at −60 mV in the voltage-clamp mode, with fast and slow capacitance, and series resistance compensated. The series resistance was monitored during recording to ensure it did not change by more than ± 3 megohms and neurons were recorded from only if uncompensated Rs < 20 megohms. MEPSCs were analyzed using Mini Analysis software v6.0.3 (Synaptosoft Inc.) with an amplitude threshold of 3.5 X average RMS noise.

DIV13-19 mouse neurons were recorded from in the same conditions except the bath was at room temperature, and 100 μM picrotoxin was added. For AMPA stimulation, coverslips were transferred to 250 μl prewarmed medium containing 100 μM S-AMPA (Abcam ab120005) and incubated at 37°C and 5% CO_2 for 2 min, followed by 8 min incubation in conditioned medium without AMPA, as for receptor internalization assays. Whole cell recordings were performed on an upright microscope (Zeiss Examiner D1) equipped with a 40X water immersion objective with a fluorescent light source (Zeiss Colibri) using an ELC-03XS patch clamp amplifier (NPI Electronics, Germany) with custom written data acquisition scripts for IgorPro 6.12A software (Wavemetrics), from Oliver Schlüter (European Neuroscience Institute, Goettingen). Signals were digitized at 10 KHz using an InstruTECH ITC-18 data acquisition interface (HEKA). The fast capacitance was compensated and slow capacitance and series resistance were not compensated. The series resistance did not change by more than 10%, monitored every 10 s. The experimenter was blinded to file names during analysis.

Whole-cell electrophysiology in acute hippocampal slices

Patch pipettes with a resistance of 2.5 to 5 megohms were prepared from glass capillaries (Harvard Apparatus, cat. no. 300060, 1.5 mm OD, 0.86 mm ID) using a P-97 puller (Sutter Inst, Novato, CA). P12-P17 male and female mouse pups were anesthetized with isoflurane (Abbott, Wiesbaden, Germany), decapitated, and the brain extracted and transferred to cold NMDG buffer containing (in mM) 45 NMDG, 0.33 KCl, 0.4 KH_2PO_4, 0.5 $MgCl_2$, 0.16 $CaCl_2$, 20 choline bicarbonate, 12.95 glucose. 300 μm coronal hippocampal slices were made with a Leica VT1200 vibratome (Wetzlar, Germany) and a stainless steel blade (Feather) in ice cold NMDG buffer. Slices were transferred to a preincubation chamber with a mesh bottom filled with ACSF containing (in mM) 124 NaCl, 4.4 KCl, 1 NaH_2PO_4, 26.2 $NaHCO_3$, 1.3 $MgSO_4$, 2.5 $CaCl_2$, and 10 D-Glucose, bubbled with 95% O_2/5% CO_2 (carbogen) and incubated at 35°C for 0.5 hours followed by another 0.5 hours at room temperature. Hippocampi were then dissected using a Zeiss Stemi 2000 stereoscopic microscope. A cut perpendicular to the CA3 pyramidal cell layer was made in the CA3 region and another cut perpendicular to the CA1 field at the medial end of the slice was made to prevent recurrent activity. Slices were submerged in a chamber perfused with carbogen-bubbled ACSF maintained at 30°C-32°C and CA1 neurons visualized with an upright Zeiss Examiner D1 microscope equipped with a 40X water-immersion objective. The intracellular pipette solution contained (in millimolar) 130 $CsMeSO_3$, 2.67 CsCl, 10 HEPES, 1 EGTA, 4 Mg-ATP, 0.3 Na-GTP, 3 QX-314 Cl, 5 TEA-Cl, 15 Phosphocreatine disodium, and 5 U/ml Creatine-phosphokinase (mOsm, 303; pH, 7.44). Whole-cell patch-clamp recordings were obtained using a NPI ELC-03XS patch clamp amplifier and digitized using an InstruTECH ITC-18 data acquisition interface (HEKA). A custom written procedure (provided by Oliver Schlüter, European Neuroscience Institute Goettingen) in Igor Pro 6.12A was used to visualize and store recorded data. Signals were low pass filtered using a Bessel filter at 3 KHz and digitized at 10 KHz. Schaffer collaterals were stimulated at 0.1 Hz using a bipolar glass electrode filled with ACSF at the distal dendritic region of the stratum radiatum close to the border of the lacunosum-moleculare, and a minimum of 30 EPSCs were averaged. The fast capacitance was compensated and slow capacitance and series resistance were not compensated. Synaptic responses were first recorded at a holding potential of −56 mV, the measured average reversal potential for Cl^- in wild-type slices. $GABA_A$ receptor-mediated currents were then recorded at 0 mV, the measured average reversal potential of NMDA and AMPA receptors. 0.1 mM picrotoxin was then perfused while holding at −56 mV after which the elimination of GABA IPSCs at 0 mV was confirmed. The residual AMPA+NMDA EPSC at 0 mV was then subtracted from the GABA IPSC. The AMPA EPSCs subsequently recorded at −56 mV, which were free of any $GABA_A$ receptor IPSC component, were used for analysis. The AMPA+NMDA compound EPSC was then recorded at +40 mV in the presence of 0.1 mM picrotoxin, where the amplitude of the EPSC approximately 60 ms after the peak of the AMPA

EPSC is considered to be purely NMDA receptor-mediated, since the measured AMPA receptor EPSC decay time constant $\tau \approx 20$ ms. The input and series resistance (Rs) were monitored every 10 s. Recordings where Rs was > 25 megohms or changed by more than ± 20% during recording were excluded from analysis. Input resistances ranged from 100 to 400 megohms and did not change by more than ± 20% during the course of the recording. Matlab was used to calculate current ratios and decay times.

Extracellular recordings from acute hippocampal slices

Acute hippocampal slices were prepared from 8 to 12-week-old male mice anesthetized with isoflurane and decapitated. The hippocampus was removed and 400 μm thick dorsal hippocampal slices were cut transversely using a tissue chopper (Stoelting) in ice-cold artificial cerebrospinal fluid (ACSF) containing, in mM: 124 NaCl, 4.9 KCl, 1.2 KH_2PO_4, 2 $MgSO_4$, 2 $CaCl_2$, 24.6 $NaHCO_3$, and 10 D-glucose (saturated with 95% O_2 and 5% CO_2, pH 7.4, ~305 mOsm). Within 5 min of decapitation, slices were transferred to a custom-built interface chamber and pre-incubated in ACSF for at least 3 hours. Following pre-incubation, the stimulation strength was set to elicit 40% of the maximal field EPSP (fEPSP) slope. The slope of the rising phase of the fEPSP was used to determine potentiation of synaptic responses. For stimulation, biphasic constant current pulses were used. A baseline was recorded for at least 30 min before LTD or LTP induction. Four averaged 0.2 Hz biphasic, constant-current pulses (0.1 ms per polarity) were used to test responses post-tetanus for up to 4 hours.

LTD was induced by a single tetanus of 900 pulses at 1 Hz. Nondecaying LTP was induced by a 3XTET high frequency stimulation with three stimulus trains of 100 pulses at 100 Hz, stimulus duration 0.2 ms per polarity with 10 min intertrain intervals (*50*). Decaying LTP was induced with a 1XTET single tetanus of 16 pulses at 100 Hz, stimulus duration 0.2 ms per polarity, modified for mouse hippocampal slices from the protocol consisting of 21 pulses used for rat hippocampal slices (*36*). The Tat-GluA2-3Y peptide (sequence YGRKKRRQRRR-$_{869}$YKEGYNVYG$_{877}$, provided by Yu-Tian Wang, University of British Columbia, Vancouver, Canada) was used at a concentration of 1 μM and ZIP (Tocris cat. no. 2549) was dissolved in water, as recommended by the manufacturer, and used at a concentration of 1 μM (*13*). Control and experimental recordings were interleaved as much as possible for comparison of conditions. For each condition, the corresponding control was repeated before and intermittently throughout experiments, to ensure consistent recording conditions.

Behavioral experiments

Behavioral experiments were performed on 3- to 9-month-old male mice by a male experimenter, except for the Barnes maze, where a male experimenter performed intraperitoneal injections and a female experimenter performed experiments. All experimenters were blinded to genotype. Mice were individually housed at the European Neuroscience Institute Göttingen, Germany animal facility in standard environmental conditions (temperature, humidity), with ad libitum access to food and water and a 12 hours light/dark cycle. Mice were allowed to habituate to different holding rooms for behavioral experiments for two weeks before testing, and to experimenters for at least three days before experiments . All experiments were performed during the light cycle. Video tracking was done with the TSE monitoring system Videomot2. All behavioral experiments were approved under project number G15.1794 by the Niedersächsisches Landesamt für Verbraucherschutz und Lebensmittelsicherheit (LAVES, Lower Saxony, Germany).

Open field, elevated plus maze

In the open field test, mice were introduced near the wall of an empty, opaque square plexiglas box and allowed to freely explore the arena for 5 min. Before every trial, urine was removed with Kimwipes, and the arena was cleaned with water and 70% EtOH. Time spent in the center (2 × 2 grid in the middle of a 4 × 4 grid arena) relative to near the walls, and the total path travelled was recorded.

For the elevated plus maze, mice were introduced into the center quadrant of a 4-arm maze with two open and two closed arms Before every trial, urine was removed with Kimwipes, and the maze was cleaned with water and 70% EtOH. Time spent in the open arms was analyzed.

Reference memory water maze

Näive mice from four independent cohorts were trained to find a 13 × 13 cm square hidden platform (uninjected mice) or a 10 cm diameter circular platform (injected mice) submerged 1.5 cm below the surface in a 1.1 m diameter circular pool filled with white opaque water at 19 ± 1°C. Mice were trained in 4 trials per day in succession where each trial lasted 1 min during which the mice searched for the platform. A trial ended when the mouse spent at least 2 s on the platform, after which they were left on the platform for 15 s prior to the start of the next trial. Four shapes around the pool (star, square, triangle, circle) served as visual cues. Mice that failed to find the platform after 1 min were guided to it and left on the platform for 15 s. Mice that failed to swim (floaters) were excluded from analysis. Mice were placed into the pool facing the wall at the beginning of each trial, and the position of pool entry from four different directions was randomly shuffled daily. For probe tests, the platform was removed and mice were placed into the pool near the wall in the quadrant opposite to that of the platform location and allowed to search for the platform for 1 min. Two probe tests were performed to monitor learning of the first platform position and additional probe test(s) performed after reversal, i.e., switching the platform position to the opposite quadrant. Only data from coincident days of training from the first 2 cohorts (Fig. 5, A to F) was pooled. Video tracking data was analyzed using Matlab to extract time-tagged xy-coordinate information and quantify escape latency, platform crossings, proximity [mean distance of all tracked path points to platform center, which is reported to be the most reliable parameter to quantify spatial specificity of water maze search patterns (*41*)], percent time spent in target quadrant, and average swimming speed. Occupancy plots were generated by super-imposing all path points of all mice in a group. Densities of path points (normalized to total number of path points in a group of mice) within a grid size of 2.1 cm × 2.1 cm (*43*) were calculated. Density data was smoothened and interpolated to plot heat maps (Scattercloud function, Matlab File ID 6037). Occupancy plot color maps were linearly scaled using the global minimum and maximum densities for all groups, to allow comparison between them. The last 0.5 s of the trial, when mice were on the platform and not searching, was excluded in occupancy plots for Fig. 5G. To generate occupancy plots of probe trials from two cohorts, in which the platform was in different positions (Fig. 5, E and F), the tracking data from one cohort was transposed and mirror imaged with respect to the center pool line and superimposed on data from the second cohort. Circular areas encompassing platforms of both cohorts in both original and reversal quadrants were defined as target areas.

Because forgetting cannot be analyzed in mice that did not learn, mice in i.p. injected cohorts (which learned more slowly) were excluded from analysis if (i) their swim speed was less than 6.2 cm/s for more than three consecutive days, (ii) their proximity or escape latency failed to decrease between the first day of training and the average of the last three days of training before platform reversal, or (iii) they spent 0% time in the target quadrant in the second probe test. These criteria were also met by all other cohorts tested. Daily i.p. injections (maximum volume injected, 10 μl/g body weight) were performed 1 hour before training (*13*). Mice were injected with saline (0.9% NaCl) during training to the first platform position and then injected with saline or with Tat-GluA2-3Y peptide (5 μmol/kg body weight) one hour before the second probe test and daily after reversal. Probe tests following reversal training of i.p. injected cohorts were performed on three consecutive days.

Delayed matching to place water maze

The delayed matching-to-place task protocol was performed as previously described (*42*) with modifications. Mice were first habituated to the task by training to a visible platform (marked by a graduated cylinder coated with multi-colored paper on top of a submerged platform) placed at a unique position every day for 3 days in a 1.1 m diameter circular pool filled with white opaque water at 19 ± 1°C. The circular platform was submerged 1.5 cm below the surface and had a radius of 5 cm. Four shapes around the pool (star, square, triangle, circle) served as visual cues. After habituation, mice were trained to 16 unique hidden platform positions at two fixed distances

from the pool center (11 in an outer ring and 5 in an inner ring) over 16 days. Mice underwent 4 trials of 2 min every day. The platform was shifted to a new random position every day in a different quadrant alternating between inner and outer rings as much as possible. A trial ended when the mouse spent at least 2 s on the platform, after which they were left on the platform for 15 s and then returned to their home cage. Mice were warmed with infrared lamps after every trial. The drop-off points with respect to the platform position of that day were in a random order. Mice were divided into two groups for counterbalancing in which each group experienced different alternations of platform positions between inner and outer rings, to prevent non-spatial chaining search strategies in which mice search for platforms within a certain distance of the pool wall. Inter-trial intervals were 5 min; from day 10 onwards, the inter-trial interval between trial 1 and trial 2 was increased to 1 hour 15 min. Trial 2 on day 16 was a probe trial, i.e., the platform was removed and the exploratory behavior of mice recorded for 2 min. Because the platform in this trial was missing, only the search path data before the first crossing of the platform position from trial 1 was used to calculate savings of proximity, escape latency and path length. All mice were trained at the same time of the day as far as possible, i.e., some mice were always trained in the morning while others were always trained in the evening. In all occupancy plots, the last 1.5 s of the trial were excluded to enhance contrast in the pool and avoid high occupancy densities on the platform, since all mice spent the last 2 s of every trial on the platform.

Quantitation of search strategy was modified from a previously described method (*43*). A Matlab script extracted time-tagged xy-coordinate information from video tracking data of all trials and classified individual trials as search strategies. The following parameters were used for classification of the indicated search strategies. Direct swimming strategy: total path length ≤1.4 X distance between drop-off point and platform center, ≥80% path points within a 30° goal corridor angle; Focal search strategy: mean proximity to platform center ≤0.35 X pool radius, mean proximity to path centroid ≤0.4 X pool radius; Directed search strategy: ≥70% path points within a 30° goal corridor angle; Perseverance strategy: mean proximity to previous platform center ≤0.5 X pool radius, mean proximity to path centroid allowed for perseverance ≤0.6 X pool radius; Chaining strategy: ≥70% path points inside annulus around the ring of platforms (inner or outer) containing the current day's platform, by 0.01 to 0.02 X pool radius, annulus for platforms in outer ring between 0.52 X pool radius and 0.73 X pool radius, annulus for platforms in inner ring between 0.22 X pool radius and 0.42 X pool radius; Scanning strategy: ≥80% of path points inside scanning radius (a circular area enclosing all platforms, 0.73 X pool radius), total % pool area scanned ≥10% and ≤50%; Thigmotaxis: ≥30% of path points inside closer wall zone starting 0.85 X pool radius and ≥50% of path points inside wider wall zone starting 0.75 X pool radius; Random Search strategy: % pool area scanned ≥50%.

Barnes Maze

The Barnes maze consisted of a white circular platform made of plexiglass, 92cm in diameter, with 20 equally spaced holes (5 cm in diameter) along the perimeter. The platform was elevated 40 cm above the ground. An escape chamber ($15.5 \times 9 \times 6$ cm^3) was placed under one of the holes, defined as the target hole. To encourage mice to enter the escape chamber, it contained a plastic ramp to enable the mice to climb into it and an odorless paper towel, resembling nesting material. Before every trial, urine was removed with Kimwipes, and the arena was cleaned with water and 70% EtOH to eliminate olfactory cues. For each trial a mouse was placed onto the middle of the maze, such that the target hole was not distinguishable from any other hole and the mouse had to rely on three visual cues (26×26 cm^2 in size) placed 5 cm from the edge of the platform to identify the location of the target hole. The platform was uniformly illuminated by bright LED light which served as a mildly aversive stimulus to motivate the mice to search for the target hole.

Daily i.p. injections (maximum volume injected, 10 µl/g body weight) were performed one hour before training (*13*). Mice were injected with saline (0.9% NaCl) during training to the first hole position and then injected with saline or with Tat-GluA2-3Y peptide (5 µmol/kg body weight) one hour before the second probe test and daily after reversal.

On the day before training, mice were habituated to the maze by placing them in a clear plastic cylinder (15 cm in diameter) in the middle of the platform for 30 s, before gently guiding them to the target hole, where they were given 3 min to enter the escape chamber. If mice did not enter the escape chamber, they were gently nudged with the cylinder into the chamber. All mice spent 1 min in the escape chamber before being returned to their home cage. On subsequent training days, mice were placed into an opaque plastic cylinder (15 cm in diameter) covered by an opaque lid for 15 s to randomize starting orientation. Each trial, lasting 2 min, was initiated after lifting the cylinder. If the mouse climbed into the escape chamber, the trial was stopped and the mouse returned to its home cage after remaining inside the escape chamber for 1 min. If the mouse did not climb into the escape chamber within 2 min, it was gently guided to the target hole with the clear plastic cylinder and given 3 min to climb into the escape chamber, where it remained for 1 min, before being returned to its home cage.

In probe tests, the escape chamber was removed, and mice were allowed to explore the maze for 2 min. The "reversal" hole was located in an adjacent quadrant, since some mice showed a bias toward exploring the opposite quadrant during initial training. Mice experienced one trial per day, and were given a one-day "break" after the first three probe tests, during which they did not perform in the Barnes maze, but still received saline (maximum volume injected, 10 µl 0.9% NaCl/g body weight) or GluA2-3Y (5 µmol/kg body weight) i.p. injections.

Three-point (head, center, tail) tracking data was exported into Matlab for further analysis. Occupancy plots were generated using Matlab as described above for the water maze, except that head instead of center coordinates were used. Video tracking errors in which head and tail coordinates were erroneously swapped (mostly close to hole regions) were detected by a distance threshold and excluded. Mice that did not find the original target hole in either of the first two probe tests (did not learn the target hole) were excluded from analysis of perseverance after reversal. Target area was defined as the area of the target hole and its two neighboring holes using center tracked coordinates, for calculation of the perseverance ratio (time spent in reversal target area divided by total time spent in original and reversal target areas).

Statistical analysis

All statistical analysis was done using GraphPad Prism or Microsoft Excel. All reported values in statistical analysis represent the mean, error bars indicate SEM, and all Student's t tests are two-tailed type 2, unless otherwise indicated. For all statistical tests, data met the assumptions of the test. All n numbers listed in Figure Legends refer to biological replicates.

REFERENCES AND NOTES

1. J. M. Henley, K. A. Wilkinson, Synaptic AMPA receptor composition in development, plasticity and disease. *Nat. Rev. Neurosci.* **17**, 337–350 (2016). doi: 10.1038/nrn.2016.37; pmid: 27080385
2. J. D. Shepherd, R. L. Huganir, The cell biology of synaptic plasticity: AMPA receptor trafficking. *Annu. Rev. Cell Dev. Biol.* **23**, 613–643 (2007). doi: 10.1146/annurev.cellbio.23.090506.123516; pmid: 17506699
3. G. L. Collingridge, J. T. Isaac, Y. T. Wang, Receptor trafficking and synaptic plasticity. *Nat. Rev. Neurosci.* **5**, 952–962 (2004). doi: 10.1038/nrn1556; pmid: 15550950
4. R. C. Malenka, M. F. Bear, LTP and LTD: An embarrassment of riches. *Neuron* **44**, 5–21 (2004). doi: 10.1016/j.neuron.2004.09.012; pmid: 15450156
5. S. Nabavi *et al*., Engineering a memory with LTD and LTP. *Nature* **511**, 348–352 (2014). doi: 10.1038/nature13294; pmid: 24896183
6. T. J. Ryan, D. S. Roy, M. Pignatelli, A. Arons, S. Tonegawa, Memory. Engram cells retain memory under retrograde amnesia. *Science* **348**, 1007–1013 (2015). doi: 10.1126/science.aaa5542; pmid: 26023136
7. J. H. Choi *et al*., Interregional synaptic maps among engram cells underlie memory formation. *Science* **360**, 430–435 (2018). doi: 10.1126/science.aas9204; pmid: 29700265
8. P. V. Migues *et al*., Blocking synaptic removal of GluA2-containing AMPA receptors prevents the natural forgetting of long-term memories. *J. Neurosci.* **36**, 3481–3494 (2016). doi: 10.1523/JNEUROSCI.3333-15.2016; pmid: 27013677
9. A. Hayashi-Takagi *et al*., Labelling and optical erasure of synaptic memory traces in the motor cortex. *Nature* **525**, 333–338 (2015). doi: 10.1038/nature15257; pmid: 26352471
10. R. Scholz *et al*., AMPA receptor signaling through BRAG2 and Arf6 critical for long-term synaptic depression. *Neuron* **66**, 768–780 (2010). doi: 10.1016/j.neuron.2010.05.003; pmid: 20547133
11. G. Ahmadian *et al*., Tyrosine phosphorylation of GluR2 is required for insulin-stimulated AMPA receptor endocytosis and LTD. *EMBO J.* **23**, 1040–1050 (2004). doi: 10.1038/sj.emboj.7600126; pmid: 14976558

12. P. V. Migues *et al*., PKMzeta maintains memories by regulating GluR2-dependent AMPA receptor trafficking. *Nat. Neurosci.* **13**, 630–634 (2010). doi: 10.1038/nn.2531; pmid: 20383136
13. Z. Dong *et al*., Long-term potentiation decay and memory loss are mediated by AMPAR endocytosis. *J. Clin. Invest.* **125**, 234–247 (2015). doi: 10.1172/JCI77888; pmid: 25437879
14. M. Y. Xiao, Y. P. Niu, H. Wigström, Activity-dependent decay of early LTP revealed by dual EPSP recording in hippocampal slices from young rats. *Eur. J. Neurosci.* **8**, 1916–1923 (1996). doi: 10.1111/j.1460-9568.1996.tb01335.x; pmid: 8921282
15. D. M. Villarreal, V. Do, E. Haddad, B. E. Derrick, NMDA receptor antagonists sustain LTP and spatial memory: Active processes mediate LTP decay. *Nat. Neurosci.* **5**, 48–52 (2002). doi: 10.1038/nn776; pmid: 11740500
16. M. D. Ehlers, Reinsertion or degradation of AMPA receptors determined by activity-dependent endocytic sorting. *Neuron* **28**, 511–525 (2000). doi: 10.1016/S0896-6273(00)00129-X; pmid: 11144360
17. E. C. Beattie *et al*., Regulation of AMPA receptor endocytosis by a signaling mechanism shared with LTD. *Nat. Neurosci.* **3**, 1291–1300 (2000). doi: 10.1038/81823; pmid: 11100150
18. D. Wu *et al*., Postsynaptic synaptotagmins mediate AMPA receptor exocytosis during LTP. *Nature* **544**, 316–321 (2017). doi: 10.1038/nature21720; pmid: 28355182
19. R. B. Sutton, B. A. Davletov, A. M. Berghuis, T. C. Südhof, S. R. Sprang, Structure of the first C2 domain of synaptotagmin I: A novel Ca^{2+}/phospholipid-binding fold. *Cell* **80**, 929–938 (1995). doi: 10.1016/0092-8674(95)90296-1; pmid: 7697723
20. S. Butz, R. Fernandez-Chacon, F. Schmitz, R. Jahn, T. C. Südhof, The subcellular localizations of atypical synaptotagmins III and VI. Synaptotagmin III is enriched in synapses and synaptic plasma membranes but not in synaptic vesicles. *J. Biol. Chem.* **274**, 18290–18296 (1999). doi: 10.1074/jbc.274.26.18290; pmid: 10373432
21. T. C. Südhof, Synaptotagmins: Why so many? *J. Biol. Chem.* **277**, 7629–7632 (2002). doi: 10.1074/jbc.R100052200; pmid: 11739399
22. S. Sugita, O. H. Shin, W. Han, Y. Lao, T. C. Südhof, Synaptotagmins form a hierarchy of exocytotic Ca(2+) sensors with distinct Ca(2+) affinities. *EMBO J.* **21**, 270–280 (2002). doi: 10.1093/emboj/21.3.270; pmid: 11823420
23. C. Dean *et al*., Axonal and dendritic synaptotagmin isoforms revealed by a pHluorin-syt functional screen. *Mol. Biol. Cell* **23**, 1715–1727 (2012). doi: 10.1091/mbc.e11-08-0707; pmid: 22398727
24. L. W. Gong, P. De Camilli, Regulation of postsynaptic AMPA responses by synaptojanin 1. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17561–17566 (2008). doi: 10.1073/pnas.0809221105; pmid: 18987319
25. D. T. Lin, R. L. Huganir, PICK1 and phosphorylation of the glutamate receptor 2 (GluR2) AMPA receptor subunit regulates GluR2 recycling after NMDA receptor-induced internalization. *J. Neurosci.* **27**, 13903–13908 (2007). doi: 10.1523/JNEUROSCI.1750-07.2007; pmid: 18077702
26. J. Boyken *et al*., Molecular profiling of synaptic vesicle docking sites reveals novel proteins but few differences between glutamatergic and GABAergic synapses. *Neuron* **78**, 285–297 (2013). doi: 10.1016/j.neuron.2013.02.027; pmid: 23622064
27. M. Rathje *et al*., AMPA receptor pHluorin-GluA2 reports NMDA receptor-induced intracellular acidification in hippocampal neurons. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 14426–14431 (2013). doi: 10.1073/pnas.1312982110; pmid: 23940334
28. T. A. Blanpied, D. B. Scott, M. D. Ehlers, Dynamics and regulation of clathrin coats at specialized endocytic zones of dendrites and spines. *Neuron* **36**, 435–449 (2002). doi: 10.1016/S0896-6273(02)00979-0; pmid: 12408846
29. J. Lu *et al*., Postsynaptic positioning of endocytic zones and AMPA receptor cycling by physical coupling of dynamin-3 to Homer. *Neuron* **55**, 874–889 (2007). doi: 10.1016/j.neuron.2007.06.041; pmid: 17880892
30. E. M. Petrini *et al*., Endocytic trafficking and recycling maintain a pool of mobile surface AMPA receptors required for synaptic potentiation. *Neuron* **63**, 92–105 (2009). doi: 10.1016/j.neuron.2009.05.025; pmid: 19607795
31. M. Rosendale, D. Jullié, D. Choquet, D. Perrais, Spatial and temporal regulation of receptor endocytosis in neuronal dendrites revealed by imaging of single vesicle formation. *Cell Reports* **18**, 1840–1847 (2017). doi: 10.1016/j.celrep.2017.01.081; pmid: 28228251
32. S. H. Lee, L. Liu, Y. T. Wang, M. Sheng, Clathrin adaptor AP2 and NSF interact with overlapping sites of GluR2 and play distinct roles in AMPA receptor trafficking and hippocampal LTD. *Neuron* **36**, 661–674 (2002). doi: 10.1016/S0896-6273(02)01024-3; pmid: 12441055
33. W. Lu *et al*., Subunit composition of synaptic AMPA receptors revealed by a single-cell genetic approach. *Neuron* **62**, 254–268 (2009). doi: 10.1016/j.neuron.2009.02.027; pmid: 19409270
34. J. W. Lin *et al*., Distinct molecular mechanisms and divergent endocytotic pathways of AMPA receptor internalization. *Nat. Neurosci.* **3**, 1282–1290 (2000). doi: 10.1038/81814; pmid: 11100149
35. Z. Dong *et al*., Hippocampal long-term depression mediates spatial reversal learning in the Morris water maze. *Neuropharmacology* **64**, 65–73 (2013). doi: 10.1016/j.neuropharm.2012.06.027; pmid: 22732443
36. S. Sajikumar, S. Navakkode, J. U. Frey, Identification of compartment- and process-specific molecules required for "synaptic tagging" during long-term potentiation and long-term depression in hippocampal CA1. *J. Neurosci.* **27**, 5068–5080 (2007). doi: 10.1523/JNEUROSCI.4940-06.2007; pmid: 17494693
37. P. Tsokas *et al*., Compensation for PKMζ in long-term potentiation and spatial long-term memory in mutant mice. *eLife* **5**, e14846 (2016). doi: 10.7554/eLife.14846; pmid: 27187150
38. Y. Yao *et al*., PKM zeta maintains late long-term potentiation by N-ethylmaleimide-sensitive factor/GluR2-dependent trafficking of postsynaptic AMPA receptors. *J. Neurosci.* **28**, 7820–7827 (2008). doi: 10.1523/JNEUROSCI.0223-08.2008; pmid: 18667614
39. N. Sadeh, S. Verbitsky, Y. Dudai, M. Segal, Zeta Inhibitory Peptide, a candidate inhibitor of protein kinase Mζ, is excitotoxic to cultured hippocampal neurons. *J. Neurosci.* **35**, 12404–12411 (2015). doi: 10.1523/JNEUROSCI.0976-15.2015; pmid: 26354909
40. M. J. LeBlancq, T. L. McKinney, C. T. Dickson, ZIP It: Neural silencing is an additional effect of the PKM-ζ inhibitor zeta-inhibitory peptide. *J. Neurosci.* **36**, 6193–6198 (2016). doi: 10.1523/JNEUROSCI.4563-14.2016; pmid: 27277798
41. H. R. Maei, K. Zaslavsky, C. M. Teixeira, P. W. Frankland, What is the most sensitive measure of water maze probe test performance? *Front. Integr. Nuerosci.* **3**, 4 (2009). doi: 10.3389/neuro.07.004.2009; pmid: 19404412
42. K. Nakazawa *et al*., Hippocampal CA3 NMDA receptors are crucial for memory acquisition of one-time experience. *Neuron* **38**, 305–315 (2003). doi: 10.1016/S0896-6273(03)00165-X; pmid: 12718863
43. A. Garthe, J. Behr, G. Kempermann, Adult-generated hippocampal neurons allow the flexible use of spatially precise learning strategies. *PLOS ONE* **4**, e5464 (2009). doi: 10.1371/journal.pone.0005464; pmid: 19421325
44. A. Citri *et al*., Calcium binding to PICK1 is essential for the intracellular retention of AMPA receptors underlying long-term depression. *J. Neurosci.* **30**, 16437–16452 (2010). doi: 10.1523/JNEUROSCI.4478-10.2010; pmid: 21147983
45. M. Fiuza *et al*., PICK1 regulates AMPA receptor endocytosis via direct interactions with AP2 α-appendage and dynamin. *J. Cell Biol.* **216**, 3323–3338 (2017). doi: 10.1083/jcb.201701034; pmid: 28855251
46. J. Yao, S. E. Kwon, J. D. Gaffaney, F. M. Dunning, E. R. Chapman, Uncoupling the roles of synaptotagmin I during endo- and exocytosis of synaptic vesicles. *Nat. Neurosci.* **15**, 243–249 (2011). doi: 10.1038/nn.3013; pmid: 22197832
47. G. A. Banker, W. M. Cowan, Rat hippocampal neurons in dispersed cell culture. *Brain Res.* **126**, 397–42 (1977). doi: 10.1016/0006-8993(77)90594-7; pmid: 861729
48. J. Rizo, T. C. Südhof, C2-domains, structure and function of a universal Ca2+-binding domain. *J. Biol. Chem.* **273**, 15879–15882 (1998). doi: 10.1074/jbc.273.26.15879; pmid: 9632630
49. A. Bhalla, M. C. Chicka, E. R. Chapman, Analysis of the synaptotagmin family during reconstituted membrane fusion. Uncovering a class of inhibitory isoforms. *J. Biol. Chem.* **283**, 21799–21807 (2008). doi: 10.1074/jbc.M709628200; pmid: 18508778
50. U. Frey, R. G. Morris, Weak before strong: Dissociating synaptic tagging and plasticity-factor accounts of late-LTP. *Neuropharmacology* **37**, 545–552 (1998). doi: 10.1016/S0028-3908(98)00040-9; pmid: 9704995

ACKNOWLEDGMENTS

We thank R. Morris for recommendations of intertrial intervals in the DMP task, H. Urbanke for the Barnes maze protocol, and J. Schrader for technical assistance. **Funding:** This work was supported by a Sofja Kovalevskaja grant from the Alexander von Humboldt Foundation, European Research Council (ERC) starting grant SytActivity FP7 260916, and Deutsche Forschungsgemeinschaft (DFG) grants DE1951/1 and -3 and the Center for Nanoscale Microscopy and Molecular Physiology of the Brain (CNMPB) to C.D.; DFG research group KFO241/PsyCourse Fi981-4/Fi981 11-1, DFG project 179/1-1/2013, ERC consolidator grant DEPICODE 648898, BMBF projects ENERGI (01GQ1421A), and Intergrament 01ZX1314D, and funds from the German Center for Neurodegenerative Diseases to A.F.; Canadian Institute for Health Research grant FDN-154286 to Y.T.W.; and a Dorothea Schloezer fellowship to K.B. **Author contributions:** A.A. coordinated the project; designed, conducted, and analyzed all behavior experiments; designed and conducted receptor and Syt3 internalization assays and immunocytochemistry experiments of Syt3/endocytic zones and PSDs; conducted and analyzed pHluorin-Syt3 experiments; conducted and trained and supervised A.H. in mEPSC recordings; trained and supervised N.N. in whole-cell recordings from hippocampal slices and S.R. in Barnes maze behavior experiments; and revised the manuscript. B.R. designed, conducted, and analyzed all LTP, LTD, and field recording experiments and conducted all immunohistochemistry experiments. S.A. designed, conducted, and analyzed all biochemistry experiments. E.B. designed behavior experiments, trained A.A. in behavior experiments, and performed all hippocampal viral injections. Y.S. designed, conducted, and analyzed immunocytochemistry experiments of Syt3 localization and all pHluorin-Syt3 experiments. N.N. conducted and analyzed whole-cell recordings from hippocampal slices. A.H. conducted and analyzed mEPSC recordings. S.R. conducted and analyzed Barnes maze behavior experiments. H.M. designed, developed, and validated the Syt3 antibody. J.B. trained A.A. in and helped plan behavior experiments. K.B. performed biochemistry experiments. Y.T.W. generated and validated the GluA2-3Y peptide and helped plan behavior experiments. A.F. provided funding for J.B., and E.B. and helped plan behavior experiments. C.D. conceived and coordinated the project; provided funding for A.A., B.R., S.A., Y.S., and K.B.; conducted and analyzed receptor and Syt3 internalization assays and immunocytochemistry of Syt3, endocytic zones, and PSDs; and wrote and revised the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/363/6422/eaav1483/suppl/DC1
Figs. S1 to S8

18 August 2018; accepted 1 November 2018
Published online 13 December 2018
10.1126/science.aav1483

RESEARCH ARTICLE SUMMARY

PLANT SCIENCE

Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field

Paul F. South, Amanda P. Cavanagh, Helen W. Liu, Donald R. Ort*

INTRODUCTION: Meeting food demands for the growing global human population requires improving crop productivity, and large gains are possible through enhancing photosynthetic efficiency. Photosynthesis requires the carboxylation of ribulose-1,5-bisphosphate (RuBP) by ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO), but photorespiration occurs in most plants such as soybean, rice, and wheat (known as C_3 crops) when RuBisCO oxygenates RuBP instead, requiring costly processing of toxic byproducts such as glycolate. Photorespiration can reduce C_3 crop photosynthetic efficiency by 20 to 50%. Although various strategies exist for lowering the costs of photorespiration, chamber- and greenhouse-grown plants with altered photorespiratory pathways within the chloroplast have shown promising results, including increased photosynthetic rates and plant size.

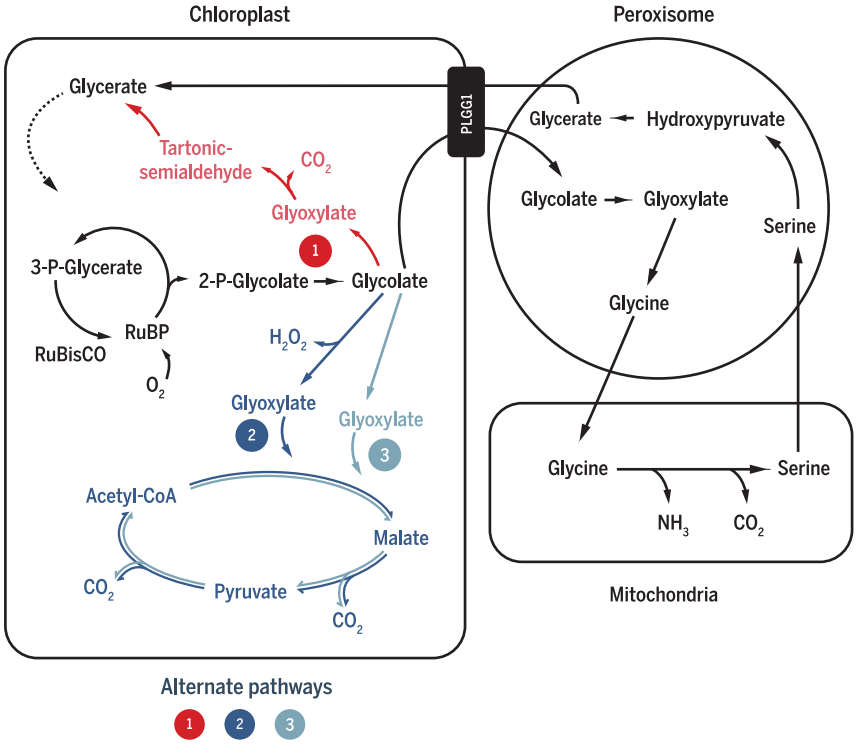
RATIONALE: To determine if alternative photorespiratory pathways could effectively improve C_3 field crop productivity, we tested the performance of three alternative photorespiratory pathways in field-grown tobacco. One pathway used five genes from the *Escherichia coli* glycolate oxidation pathway; a second pathway used glycolate oxidase and malate synthase from plants and catalase from *E. coli*; and the third pathway used plant malate synthase and a green algal glycolate dehydrogenase. All enzymes in the alternative pathway designs were directed to the chloroplast. RNA interference (RNAi) was also used to down-regulate a native chloroplast glycolate transporter in the photorespiratory pathway, thereby limiting metabolite flux through the native pathway. The three pathways were introduced with and without the transporter RNAi construct into tobacco, which is an ideal model field crop because it is easily transformed, has a short life cycle, produces large quantities of seed, and develops a robust canopy similar to that of other field crops.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aat9077

RESULTS: Using a synthetic biology approach to vary promoter gene combinations, we generated a total of 17 construct designs of the three pathways with and without the transporter RNAi construct. Initial screens for photoprotection by alternative pathway function under high–photorespiratory stress conditions identified three to five independent transformants of each design for further analysis. Gene and protein expression analyses confirmed expression of the introduced genes and suppression of the native transporter in RNAi plants. In greenhouse screens, pathway 1 increased biomass by nearly 13%. Pathway 2 showed no benefit compared to wild type. Introduction of pathway 3 increased biomass by 18% without RNAi and 24% with RNAi, which were consistent with changes in photorespiratory metabolism and higher photosynthetic rates. Ultimately, field testing across two different growing seasons showed >25% increase in biomass of pathway 3 plants compared to wild type, and with RNAi productivity increased by >40%. In addition, this pathway increased the light-use efficiency of photosynthesis by 17% in the field.

CONCLUSION: Engineering more efficient photorespiratory pathways into tobacco while inhibiting the native pathway markedly increased both photosynthetic efficiency and vegetative biomass. We are optimistic that similar gains may be achieved and translated into increased yield in C_3 grain crops because photorespiration is common to all C_3 plants and higher photosynthetic rates under elevated CO_2, which suppresses photorespiration and increases harvestable yield in C_3 crops. ■



Alternative photorespiratory pathways in tobacco. Three alternative pathways [1 (red), 2 (dark blue), and 3(light blue)] introduced into tobacco chloroplasts for more efficient recycling of glycolate. RNAi suppresses the native glycolate/glycerate transporter PLGG1 to prevent glycolate from leaving the chloroplast and entering the native pathway (gray).

The list of author affiliations is available in the full article online.
***Corresponding author. Email: d-ort@illinois.edu**

Cite this article as P. F. South *et al.*, *Science* 363, eaat9077 (2019). DOI: 10.1126/science.aat9077

RESEARCH ARTICLE

PLANT SCIENCE

Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field

Paul F. South[1,2], Amanda P. Cavanagh[2], Helen W. Liu[3]*, Donald R. Ort[1,2,3,4]†

Photorespiration is required in C_3 plants to metabolize toxic glycolate formed when ribulose-1,5-bisphosphate carboxylase-oxygenase oxygenates rather than carboxylates ribulose-1,5-bisphosphate. Depending on growing temperatures, photorespiration can reduce yields by 20 to 50% in C_3 crops. Inspired by earlier work, we installed into tobacco chloroplasts synthetic glycolate metabolic pathways that are thought to be more efficient than the native pathway. Flux through the synthetic pathways was maximized by inhibiting glycolate export from the chloroplast. The synthetic pathways tested improved photosynthetic quantum yield by 20%. Numerous homozygous transgenic lines increased biomass productivity by >40% in replicated field trials. These results show that engineering alternative glycolate metabolic pathways into crop chloroplasts while inhibiting glycolate export into the native pathway can drive increases in C_3 crop yield under agricultural field conditions.

Population growth, increasing global affluence, and an expanding bioeconomy are conspiring to increase mid-century agricultural demand by 60 to 120% over 2005 levels, a challenge that current rates of crop productivity improvement averaging <2% per year cannot meet (*1*–*3*). In the 45 years after 1960, global crop productivity increased 135% from 1.84 to 3.96 metric tons per hectare (*4*). The increased use of pesticides, fertilizers and irrigation, and mechanization, along with the adoption of higher-yielding crop varieties that drove this remarkable global increase in productivity, are now largely optimized for major crops and are unlikely to generate sufficient yield increases to meet mid-century agricultural demand. However, photosynthetic efficiency remains standing as a determinant of yield potential with the improvement capacity to double crop productivity (*1*–*3*, *5*, *6*). In C_3 crops such as wheat, rice, and soybeans, photorespiration reduces the photosynthetic conversion efficiency of light into biomass by 20 to 50%, with the largest losses occurring in hot dry climates where yield increases are sorely needed. Whereas ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) carboxylates ribulose-1,5-bisphosphate (RuBP) during photosynthesis, the unproductive and energy-intensive process of photorespiration results from oxygenation of RuBP by BuBisCO, which becomes more prevalent at higher temperatures and under drought conditions (*6*, *7*). Toxic by-products of the BuBisCO oxygenation reaction (2-phosphoglycolate and glycolate) and of the glycine decarboxylation reaction (ammonia) are recycled by photorespiration into nontoxic products but at the expense of energy and net loss of fixed carbon (*6*, *7*). Some photosynthetic algae, bacteria, and plants have evolved mechanisms to reduce the oxygenation reaction by BuBisCO via carbon-concentrating mechanisms (CCMs), including C_4 photosynthesis (*8*, *9*), inspiring efforts to introduce CCMs into C_3 plants (*8*–*12*). Here we have taken an alternative approach of introducing non-native and synthetic metabolic pathways to recycle the products of BuBisCO oxygenation more efficiently (*13*). Previously, two alternative photorespiratory pathways implemented in *Arabidopsis* improved photosynthesis and plant size in chamber and greenhouse experiments (*14*, *15*). These results inspired us to optimize these alternative photorespiratory pathways in tobacco, a useful agricultural model crop, for field trials. Computer modeling of these alternative pathways revealed the importance of optimized expression of non-native genes to achieve maximum flux through the alternative pathway and thus maximize the benefits for crop plants under field conditions (*16*). Additionally, we sought to minimize flux through the native photorespiratory pathway and maximize flux through the introduced pathways by inhibiting glycolate export from the chloroplast.

[1]Global Change and Photosynthesis Research Unit, United States Department of Agriculture–Agricultural Research Service, Urbana, IL 61801, USA. [2]Carl R. Woese Institute for Genomic Biology, University of Illinois, Urbana, IL 61801, USA. [3]Department of Crop Sciences, University of Illinois, Urbana, IL 61801, USA. [4]Department of Plant Biology, University of Illinois, Urbana, IL 61801, USA.
*Present address: Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA.
†Corresponding author. Email: d-ort@illinois.edu

Results

Transgene assembly

We transformed *Nicotiana tabacum* cv. Petite Havana (tobacco) with three different photorespiratory alternative pathway (AP) designs, expressing as many as seven genes in single constructs (Fig. 1A and table S1). Tobacco is an ideal model crop for these studies because of its completely sequenced genome, short life cycle (3 months from seed to seed), well established high-efficiency transformation protocols, and the ability to form a fully closed canopy like other crops in the field. The AP1 construct targets the five genes of the *Escherichia coli* glycolate oxidation pathway to the chloroplast (Fig. 1A) (*14*). AP2 includes *Arabidopsis* glycolate oxidase (GO) and *Cucurbita maxima* (pumpkin) malate synthase (MS), along with a catalase (CAT) from *E. coli* (Fig. 1A) (*15*). AP3 also contains *C. maxima* MS sequence but replaces the plant GO used in AP2 with *Chlamydomonas reinhardtii* glycolate dehydrogenase (CrGDH) to avoid hydrogen peroxide production when glycolate is converted to glyoxylate (Fig. 1A). With this modification, expression of *E. coli* CAT in the chloroplast is unnecessary (*17*). Using multigene constructs assembled from modular parts by Golden Gate cloning, we generated multiple promoter gene combinations and within-construct position effects to optimize AP performance. We generated five iterations of AP1, three iterations of AP2, and a single design of AP3 for testing (table S1). In addition to the expression of the AP genes, we designed a long hairpin RNA interference (RNAi) construct and added it to the library of multigene constructs to reduce the expression of the chloroplast glycolate-glycerate transporter *PLGG1* with the goal of minimizing glycolate flux out of the chloroplast and into the native pathway (Fig. 1 and table S1) (*18*, *19*). In total, we successfully transformed 17 different constructs of the three AP designs into tobacco with and without the inclusion of an RNAi module targeting the PLGG1 transporter.

Gene and protein analysis confirm chloroplast-localized transgene expression

Transgene expression analysis conducted on three independent transformants of each AP design selected for further analysis confirmed strong expression of the transgenes along with ~80% RNAi suppression of *PLGG1* expression (Fig. 1B and fig. S1). Immunoblot analysis of whole-cell extract was normalized on the basis of total protein content and verified using antibodies against the RuBisCO large subunit and actin (fig. S2). Immunoblot analysis of isolated intact chloroplasts from AP3 plants (Fig. 1C) verified that the construct design of AP3 directs CrGDH and MS protein to the chloroplast and that RNAi suppresses expression of the PLGG1 transporter protein. The cytoplasmic marker protein actin was undetectable in the isolated chloroplast fraction, ensuring that the AP3 proteins in the chloroplast fraction was not a result of cytoplasmic contamination (Fig. 1C). Moreover, the chloroplast marker PGL35 was only faintly detectable in the whole-leaf extracts but was greatly enriched in the isolated chloroplast fraction (Fig. 1C). Whereas MS was also greatly enriched in the

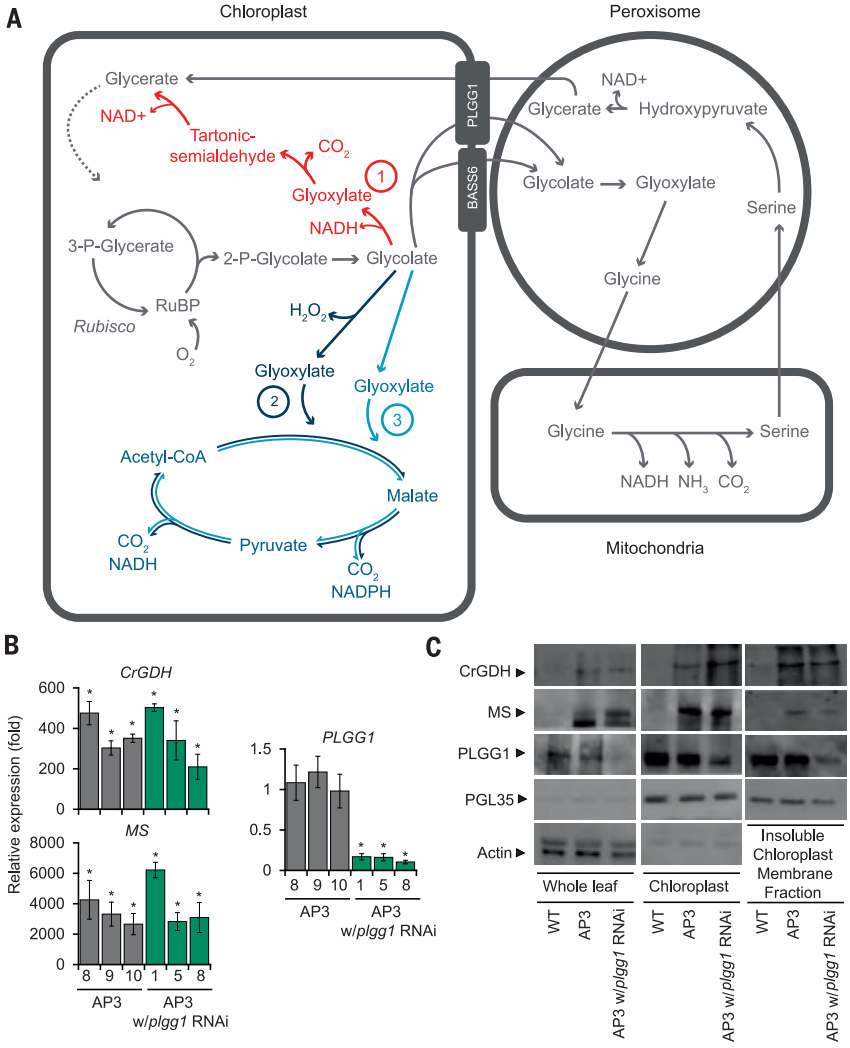


Fig. 1. Alternative photorespiratory pathways. (**A**) Model of three alternative photorespiration pathway designs. AP1 (red) converts glycolate to glycerate using five genes from the *E. coli* glycolate pathway encoding the enzymes glycolate dehydrogenase, glyoxylate carboligase, and tartonic semialdehyde reductase. AP2 (dark blue) requires three introduced genes encoding glycolate oxidase, malate synthase, and catalase (to remove hydrogen peroxide generated by glycolate oxidase). AP3 (blue) relies on two introduced genes: *Chlamydomonas reinhardtii* glycolate dehydrogenase and *Cucurbita maxima* malate synthase. (**B**) qRT-PCR analysis of the two transgenes in AP3 and the target gene *PLGG1* of the RNAi construct. Results for three independent transformation events are shown with (1, 5, and 8) and without (8, 9, and 10) *PLGG1* RNAi. Error bars indicate SEM. * indicates statistical difference at $P < 0.05$ compared to WT based on one-way ANOVA. Actual *P* values are shown in supplementary data set 15. (**C**) Immunoblot analysis from whole leaves and isolated chloroplasts, including the insoluble membrane fraction, using custom antibodies raised against the indicated target genes, cytosolic marker actin, and chloroplast-specific marker platoglobulin 35 (PGL35). Five micrograms of protein was loaded per lane. Arrows indicate detected protein based on molecular weight. The kinetic properties of CrGDH, as well as numerous malate synthase enzymes, have been previously characterized (table S3) (*17*).

chloroplast fraction, CrGDH appeared to be much less enriched in this fraction (Fig. 1C). Glycolate dehydrogenases have be shown to strongly associated with membranes in both chlorophytes and bacteria (*20*, *21*) and thus may have been inefficiently extracted from our chloroplast preparation (*17*). Isolation of the insoluble membrane fraction from the chloroplast extraction showed that a large fraction CrGDH in tobacco chloroplasts was associated with the membranes (Fig. 1C) and that CrGDH was enriched relative to PGL35 in the membrane fraction.

AP plants are resistant to photorespiration stress

Following selection for construct expression by selectable marker screening [BASTA resistance (bar) gene added to all constructs] (table S1) and genotyping selection for single-insert homozygous transgenic plants, all independent constructs of the three AP designs were assessed for resistance to photorespiration stress in a high-throughput chlorophyll fluorescence assay. Photorespiratory mutants typically display impaired growth and photosynthesis when transferred from elevated CO_2 concentrations ($[CO_2]$) to ambient air, which is accompanied by the onset of photoinhibition that can be diagnosed by monitoring chlorophyll fluorescence (*19*, *22*–*24*). We hypothesized that AP function would be photoprotective under high photorespiratory stress, thus protecting photosystem II operating efficiency (i.e., Fv′/Fm′) from photodamage (*19*, *22*). Previously, this method of monitoring Fv′/Fm′ after illumination in low $[CO_2]$ enabled identification of photorespiration mutants that cause photoinhibition (*19*, *22*, *24*). Using this protocol to monitor AP function, we exposed thousands of single-insert homozygous T2 seedling plants to 24 hours of high light intensity (1200 μmol m^{-2} s^{-1}) and very low $[CO_2]$ (1 to 38 μbar CO_2) and then compared Fv′/Fm′ in the transformants with azygous wild-type (WT) and empty vector (EV) controls (fig. S3). Many independent transformants (66% of AP1, 54% of AP2, and 84% of AP3 plants) were significantly more photoprotective under this severe photorespiratory stress. Versions of AP1 and AP3 sustained 33 to 48% higher Fv′/Fm′ values compared to WT and EV controls (Fig. 2, A and B, and data set S1). Under ambient $[CO_2]$, there were no observed differences in Fv′/Fm′ between the AP and control lines. However, *PLGG1* RNAi inhibition of glycolate efflux from the chloroplast reduced Fv′/Fm′ when these plants were shifted from elevated $[CO_2]$ to ambient (fig. S4). This photoinhibited phenotype of the *PLGG1* RNAi plants was not only rescued by transgenic complementation with AP1 or AP3 constructs, but was also substantially more resistant to photoinhibition than WT and EV controls (Fig. 2C and dataset S1).

AP plants show enhanced biomass accumulation in greenhouse growth studies

Following the initial photoprotection screen and expression analysis, we determined the impact of the three APs on plant growth in greenhouse growth studies. Both the AP1 and AP3 designs significantly increased dry-weight biomass relative to the WT plants. Overall, AP1 plants increased dry weight biomass by 13%, but the benefit was lost when the *PLGG1* RNAi module was present (Fig. 3B). AP2 introduction did not significantly alter dry weight (Fig. 3B). Three AP3 lines that sustained much higher Fv′/Fm′ values (200-8,9,10) compared to WT and EV were taller (Fig. 3A) and showed the largest increases in biomass in greenhouse studies, with a 24% increase with and 18% increase without the *PLGG1* RNAi module compared to WT (Fig. 3B). We also tested an AP3 line that had the same Fv′/Fm′ as WT and EV (200-4), which showed no increase in biomass, and one line that had an intermediate Fv′/Fm′ (200-6) that showed a small but statistically significant biomass increase in greenhouse studies (fig. S5, A and C). Transcript expression analysis of AP3

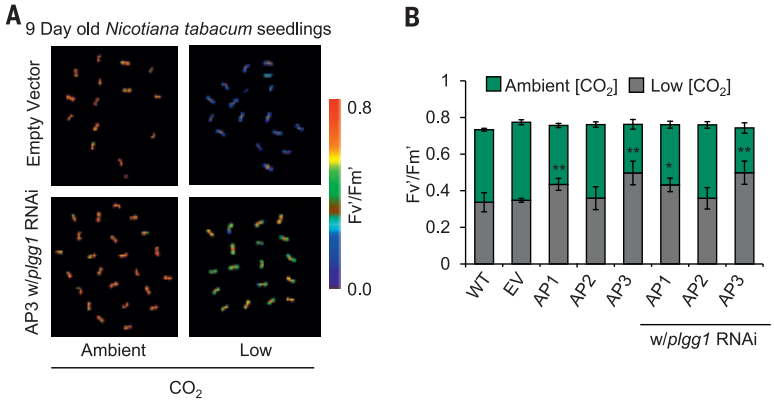


Fig. 2. AP plant lines are more photoprotective under photorespiration stress. (**A**) Representative photos of 9-day-old T2 transgenic tobacco lines during the chlorophyll fluorescence photoprotection screen for AP pathway function showing AP3 protecting photosystem II from photodamage under severe photorespiratory conditions. (**B**) Combined values of the three AP construct designs with and without RNAi targeting the glycolate-glycerate transporter PLGG1. Error bars indicate SEM. * indicates statistical difference compared to WT based on one-way ANOVA at $P \leq 0.05$, ** $P \leq 0.001$. Fv′/Fm′ for individual lines is described in supplementary data set 1. Actual significant P values are shown in supplementary data set 15.

events 200-4 and 200-6 revealed that CrGDH and MS expression was greatly reduced compared to transgenic events 200-8,9,10 (fig. S5B).

AP3 plants have an altered photorespiratory metabolite profile

We further investigated the AP3 plants that showed the greatest growth stimulation and gene expression to determine the effect of AP3 enzymes on the leaf photorespiratory metabolite profile. We performed gas chromatography followed by mass spectrometry on leaf samples from greenhouse-grown WT and AP3 plants to analyze the photorespiratory intermediates glycolate, glyoxylate, glycine, and serine and the AP3-specific intermediate pyruvate (Fig. 4). AP3 introduction with and without the RNAi module increased glyoxylate and pyruvate concentration compared to WT, suggesting altered native photorespiration and possibly flux through the alternative pathway (Fig. 4, B and F). AP3 plants with and without the RNAi module also had decreased concentrations of the photorespiratory intermediates serine, for which photorespiration is a major source (*25*), and glycerate, possibly due to a diversion of carbon away from the native photorespiratory pathway (Fig. 4, D and E). Glycine concentrations were similar in AP3 and WT plants (Fig. 4C). AP3 with the RNAi module targeting the glycolate-glycerate transporter *PLGG1* had increased glycolate accumulation compared to WT in a manner similar to the *Arabidopsis* T-DNA insertion mutant *plgg1-1* (Fig. 4A) (*18*, *19*).

AP3 plants exhibit increased photosynthetic rate and chloroplast [CO_2]

To test if altered photorespiration due to introduction of the AP3 design in plants affects rates of photosynthesis, we compared CO_2 assimilation rates (A) as a function of intercellular CO_2 concentrations (C_i) under saturating light in AP3 and WT plants. AP3 plant lines with and without the *PLGG1* RNAi module had increased rates of photosynthesis compared to WT (Fig. 5A). Modeling of the A/C_i curves showed increases in the maximum rate of BuBisCO carboxylation (V_{cmax}) visualized in the initial slope of the A/C_i curve in AP3 lines (Fig. 5, A and C). We observed no statistical differences in the maximum rate of electron transport (J_{max}) in any AP design (fig. S6). Increases in V_{cmax}, which is a property of BuBisCO enzymatic activity, could be due to increased BuBisCO protein content or increased availability of CO_2 as a substrate for BuBisCO. Immunoblot analysis shows no difference in BuBisCO content on a per microgram protein basis (fig. S2), suggesting that the observed difference is based on increased availability of CO_2 at the site of carboxylation in the chloroplast. Increases in CO_2 availability for BuBisCO carboxylation could arise from increased mesophyll conductance (g_m; i.e., the diffusion of CO_2 into mesophyll cell chloroplasts) or from the direct release of photorespiratory CO_2 in the chloroplast by the decarboxylation of malate and pyruvate in the plastid (Fig. 1A), both of which would result in observed increase in V_{cmax} determined from A/C_i curves (*26*). There is no apparent reason to expect that the introduction of these alternative pathways would decrease the resistance for the movement of CO_2 from the mesophyll intercellular air space to the

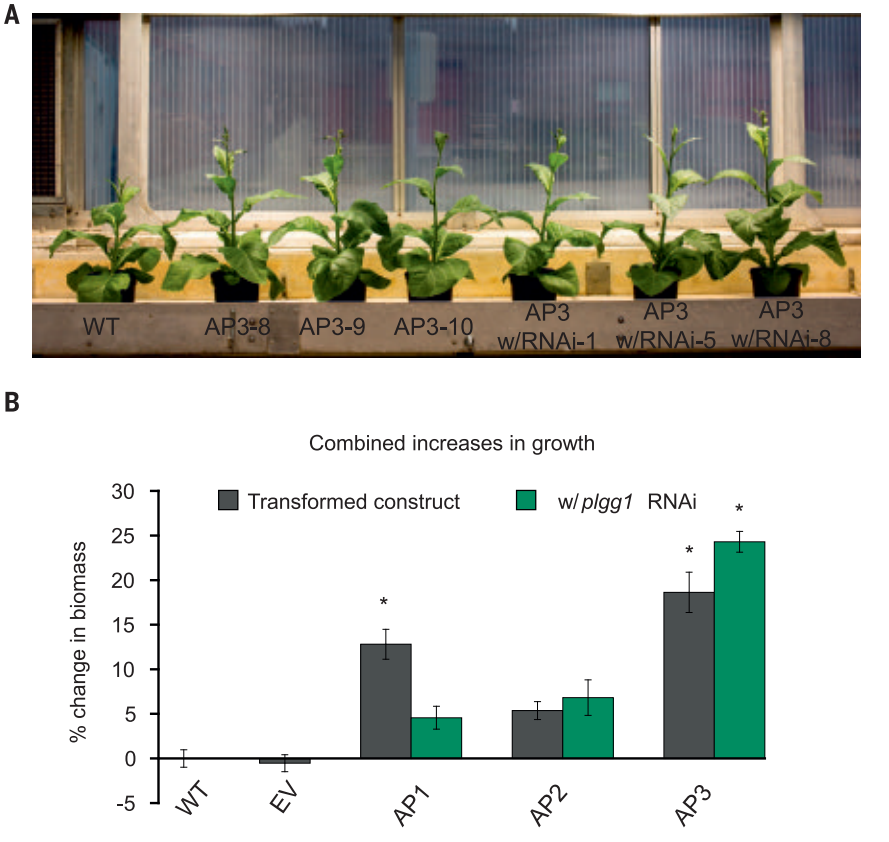


Fig. 3. Photorespiration AP lines increase biomass under greenhouse conditions. (**A**) Photos of 6-week-old AP3 and WT plants grown in the greenhouse. Individual plant lines are indicated in the labels below the plant. (**B**) Percent difference in total dry weight biomass of the indicated combined plant lines. * indicate statistical difference based on one-way ANOVA. Error bars are SEM, $n = 7$ (plants measured), $P < 0.05$ values listed in data set 15.

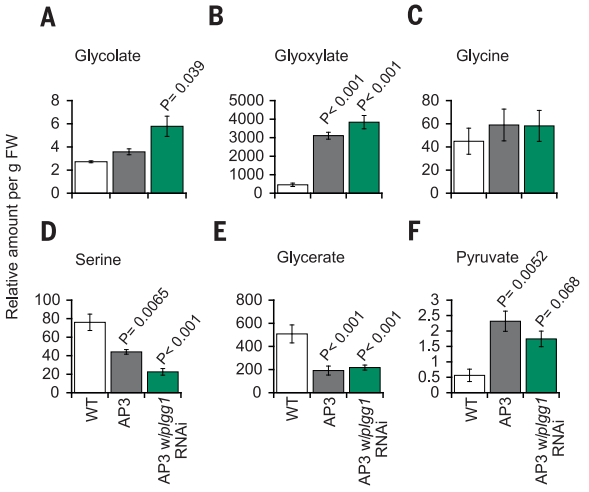


Fig. 4. Photorespiratory and AP3 metabolic intermediates. (**A** to **F**) Relative amount of the indicated metabolite detected from ~40 mg of leaf tissue (fresh weight; FW) sampled in the late morning. Metabolite concentrations were reported as concentrations relative to the internal standard, which is the target compound peak area divided by peak area of hentriacontanoic acid: N_i (relative concentration) = X_i (target compound peak area) * X^{-1}IS (peak area of hentriacontanoic acid) per gram fresh weight. Error bars indicate SEM, n = 4 leaf samples. Statistical differences between AP3 designs and WT based on one-way ANOVA, with P values indicated. All P values are listed in dataset 15.

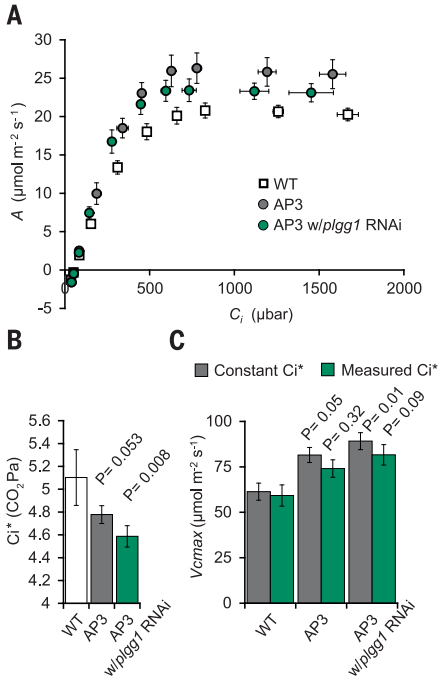


Fig. 5. Photosynthetic efficiency of greenhouse-grown plants. Data are the combined result of three independent transformants (hereafter referred to as combined) with and without *PLGG1* RNAi. (**A**) CO_2 assimilation based on intercellular [CO_2] (C_i). (**B**) Combined apparent CO_2 compensation point: C_i* calculated using the common intercept method and slope regression (*29*). (**C**) Combined maximum rate of BuBisCO carboxylation (V_{cmax}). V_{cmax} values are presented at 25°C and modeled from photosynthetic response under changing CO_2 concentration. Gray bars indicate constant C_i*; green bars indicate derived values based on measured C_i*. Error bars indicate SEM. P values for statistical comparison to WT based on one-way ANOVA are given.

chloroplast stroma or from the mitochondria to the chloroplast stroma. However, an increase in g_m contributing to increased CO_2 availability within the AP3 plant chloroplast is difficult to rule out, largely because modeling of g_m requires knowledge of, or assumptions about, the conductance of CO_2 released from the mitochondria during the conversion of glycine to serine to the chloroplast, which is directly affected by the introduction of the alternative pathway.

Theory predicts that the release of photorespiratory CO_2 in the chloroplast by the AP pathways, instead of in the mitochondria through the native photorespiratory pathway, would lower C_i*, the intercellular [CO_2] at which the chloroplast [CO_2] reaches Γ*, the [CO_2] in the chloroplast at which the rates of BuBisCO oxygenation and carboxylation are equal (*27*–*29*). To determine C_i*, we measured the internal [CO_2] at which CO_2 response curves measured at different subsaturating illumination intensities intersect (*29*). AP3 plants with the RNAi module targeting *PLGG1* showed a significant reduction of 10% in C_i* and AP3 plants without the RNAi module showed a significant reduction of 6.4% in C_i* compared to WT (Fig. 5B). The observed decreases in C_i*, coupled with unaltered respiration (fig. S7), are consistent with elevated chloroplastic [CO_2] due to decarboxylation of malate and pyruvate within the introduced pathway (Figs. 1A and 5B and fig. S6), which would also explain the significantly higher values of V_{cmax} in AP3 plants compared to WT plants (Fig. 5, A and C, and fig. S6). Accounting for the observed C_i* in A/C_i curve analysis reduces the apparent change in V_{cmax} further indicating that the difference in V_{cmax} was not due to changes in BuBisCO content or activity but rather by increased chloroplastic [CO_2] (Fig. 5C).

AP plants show increased photosynthetic rates, quantum efficiency, and biomass accumulation in replicated field trials

In the 2016 growing season, we tested four independent events of AP1, two independent events of AP2, and five independent transformation events of AP3, along with two WT and two EV controls in the field, using a randomized single block design experiment (fig. S8). Biomass increased by 16% in AP1 lines and 10% in one of the AP2 lines tested (fig. S9). The three AP3 lines that showed the largest biomass increases in the greenhouse consistently showed the largest increases in dry-weight biomass, with total biomass increasing by as much as 23% relative to WT (fig. S9). Independent AP3 events 200-4 and 200-6, in which CrGDH and MS expression was significantly lower compared to other transgenic events (fig. S5B) and showed less or no improvement in greenhouse biomass (fig. S5C), also showed no increases in total biomass in the 2016 field season (fig. S9). We anticipated, owing to their lower energetic requirements, that the AP pathways would improve the maximum quantum efficiency of net CO_2 assimilation (Φa) relative to the native pathway. Φa increased in lines of all AP pathways, in many cases by >20%, including those containing the RNAi module targeting the PLGG1 transporter (fig. S10) but not in AP3 events 200-4 and 200-6 (fig. S11A). The high–biomass-producing AP3 plant lines exhibited an increased light-saturated rate of assimilation compared to WT, to several AP1 lines, and to all AP2 plant lines (fig. S10C) and to AP3 events 200-4 and 200-6 (fig. S11).

To validate the 2016 field results and improve the statistical power of comparisons with AP3 plants under agricultural conditions, we tested five randomized replicate blocks of three AP3 independent transformed lines with and without the RNAi module targeting *PLGG1* in comparison to WT during the 2017 growing season (fig. S12). The AP3 plant lines showed a 25% increase in total dry-weight biomass (22% leaf, 44% stem), and the inclusion of the *PLGG1* RNAi module in AP3 designs further increased leaf dry biomass to 33%, stem dry biomass to 50%, and total dry biomass to 41% compared to WT (Fig. 6A). That AP3 plant lines with the RNAi module showed a significant leaf and total dry weight biomass increase (12% and 17%, respectively) over the AP3-only plants supports our hypothesis that forcing greater glycolate flux through the synthetic pathway by inhibiting flux through the native photorespiratory pathway drove the increased productivity. Total mid-day starch content in AP3 plants increased by ~70% and in AP3 with *PLGG1* RNAi by ~40% compared to the WT control (Fig. 6B). The apparent quantum

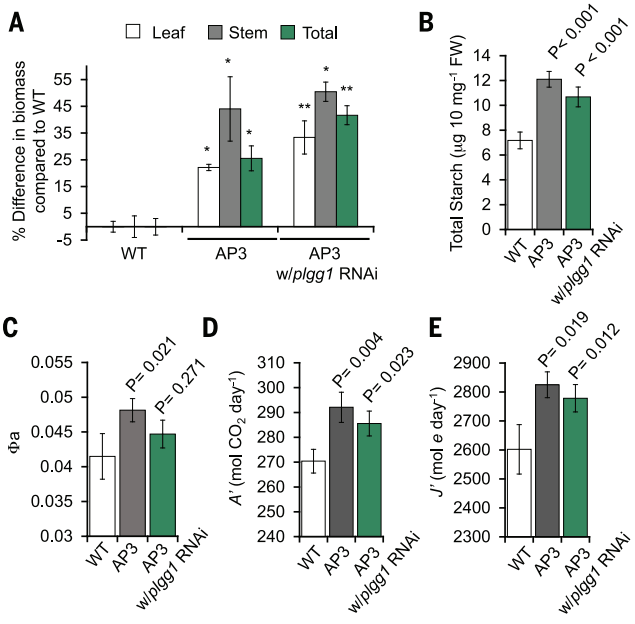


Fig. 6. Plant productivity and photosynthetic performance in 2017 field trials. (**A**) Percent difference from WT for stem, leaf, and total biomass of AP3 with and without the *PLGG1* RNAi module. Data are the combined result of three independent transformants with and without *PLGG1* RNAi. * indicates significance compared to WT and ** indicates significance between WT and AP3-only lines compared to AP3 with RNAi lines. *P* values are shown in supplementary data set 15. (**B**) Total combined accumulated leaf starch for indicated lines extracted from 10 mg of fresh weight leaf material. (**C**) Combined apparent quantum efficiency of photosynthesis (Φa) determined by linear regression of assimilation based on available light-response curves. (**D**) Combined accumulated assimilation of CO_2 (*A′*) based on diurnal analysis of photosynthesis. (**E**) Combined accumulated electrons used in electron transport determined from assimilation based on diurnal photosynthesis. Error bars indicate SD, and *P* values are indicated based on two-way ANOVA.

efficiency of photosynthesis increased in both AP3 plant pathways; by 7% with and 17% without *PLGG1* RNAi for the 2017 field season (Fig. 6C and fig. S13). Because plants with both AP3 designs exhibited increases in the quantum efficiency of photosynthesis and decreases in C_i*, we hypothesized that total daily net carbon gain through photosynthesis would be higher compared to WT, resulting in the observed increases in biomass over the growing season (Figs. 5B and 6, A and C). Indeed, modeled daily net carbon gain from measurements of photosynthesis over a diurnal time course in plants containing AP3 showed an increase of 5 to 8% in CO_2 assimilation (*A′*) and increases in electron use in photosynthesis (*J′*) compared to WT (Fig. 6, D and E).

Discussion

We showed that installing synthetic glycolate metabolic pathways into tobacco chloroplasts drove large increases in biomass accumulation in both greenhouse conditions and in the field under agricultural conditions. Because AP3 plants exhibited the greatest growth stimulation, we selected this pathway for more in-depth characterization. In summary, the AP3 transgene products CrGDH and MS localized to the chloroplast (Fig. 1C). Evidence that these transgene products function in the chloroplast to catalyze the reactions depicted in Fig. 1 include the stimulation of the rate of photosynthesis (Fig. 5) and improvement of photosynthetic quantum yield (Fig. 6 and fig. S10), the lowering of C_i* (Fig. 5) and increase in the initial slope of an A/C_i relationship (Fig. 5) that both indicate increased [CO_2] in the chloroplast, and the altered photorespiratory metabolite profile (Fig. 4). In addition, the direct role of this pathway in chloroplast glycolate metabolism is supported by its ability to prevent photoinhibition and rescue the reduced growth phenotype of *PLGG1* RNAi tobacco lines (fig. S4). Moreover, AP3 lines that contain the full transgene construct but with reduced transgene expression showed less, or no improvement in greenhouse (fig. S5C) or field biomass (fig. S9), and Φa values similar to WT (fig. S11) provide evidence that the amount of expression of the introduced alternative photorespiratory pathway drove the extent of improved growth and increased photosynthetic efficiency.

Of the two alternative pathways to photorespiration that inspired our designs (*13*, *14*), AP2 showed limited improvements in plant productivity, and 24% of the independent transgenic AP2 lines resulted in stunted growth and yellow leaves (fig. S14C). The AP1 design improved productivity in tobacco, but the enhancement associated with AP1 was eliminated in both greenhouse and field settings when the *PLGG1* RNAi module was added (Fig. 3 and fig. S9). Modeling (*16*) predicted that directing the complete flux of glycolate through the AP1 pathway by inhibiting glycolate export from the chloroplast would result in the largest increase in energy savings, photosynthetic efficiency, and growth among all designs. Elimination of the AP1 enhancements by the *PLGG1* RNAi module implies that this introduced pathway may not have had sufficient kinetic capacity to handle the full glycolate flux under high rates of BuBisCO oxygenation. Further optimization of expression of AP1 genes and/or use of AP1 genes of different origins and kinetic properties may lead to achieving the full benefits that modeling predicts for this design. The AP3 design containing *C. maxima* MS and CrGDH reliably increased plant biomass and improved photosynthetic efficiency (Figs. 3, 5, and 6), and the phenotype is dependent on the level of expression of the transgenes in the independent transformation events (fig. S5). The inclusion of an RNAi module to reduce expression of the *PLGG1* chloroplast glycolate-glycerate transporter in numerous independent transformant plant lines increased postharvest dry-weight biomass compared to AP3 introduction alone by 17% ($P < 0.001$) (Figs. 3 and 6A and fig. S13). Without an alternative photorespiration pathway in place, inhibition of *PLGG1* expression by RNAi decreased plant growth and led to photoinhibition (i.e., reduced Fv′/Fm′) when these plants were transferred from elevated [CO_2] to ambient air (fig. S4), as was reported previously for the *plgg1-1* T-DNA knockout line in *Arabidopsis* (*23*). Thus, the genetic complementation of the low-growth photoinhibited phenotype and the significant increase in biomass in AP3 lines with RNAi over AP3 alone are consistent with the expected benefit of directing a greater proportion of the glycolate flux through the synthetic pathway in the chloroplast and away from the native photorespiratory pathway outside of the chloroplast (*18*, *19*). Indeed, forcing greater glycolate flux through the synthetic pathway by inhibiting glycolate transport out of the chloroplast through PLGG1 into the native photorespiratory pathway resulted in growth stimulation in field trials of >40% for the AP3 plants with RNAi. The glycolate-glycerate exchange transporter PLGG1 works in tandem with a second glycolate exporter BASS6 to stoichiometrically balance the export of two glycolate molecules with the import of one glycerate molecule during photorespiration (*18*). Thus, targeting the expression of both transporters may further test AP3 kinetic capacity and may drive even greater growth stimulation. Recognizing that these alternative pathways are intervening in the central metabolism of photosynthetic cells, it will be important to validate the biochemistry that is occurring as AP pathway intermediates may well have destinations that are different from those depicted in Fig. 1A.

Although inhibiting photorespiration under normal oxygen-containing atmospheres invariably results in inhibited photosynthesis and growth (*7*), some evidence indicates that stimulating

photorespiratory flux can enhance photosynthetic rate and plant growth. Overexpression of the H-protein in the glycine decarboxylase complex or overexpression of plant glycolate oxidase (GO) can lead to increased photosynthesis and biomass production (*30*, *31*). In both of these reports, the overexpression of these photorespiration genes was accompanied by an increase in stomatal conductance that itself would be expected to increase photosynthesis and growth under water-replete conditions. Conversely, four different photorespiration mutants (*pglp1, shm1, hpr1,* and *glyk1*) partially lost stomatal responsiveness to altered CO_2 availability, possibly indicating that alternative pathways could influence plant adaptation through stomatal signaling (*32*). We saw no statistical differences in stomatal conductance (fig. S15, A and B) or the expression of GO (fig. S15C) in AP3 tobacco plants, indicating that neither of these contributed to the stimulations observed in AP3 plant lines. Whether the installation of these alternative pathways may affect global changes in the transcriptome and the proteome and if that may have secondary impacts on plant growth outside of changes to primary metabolism remain to be determined. Energy demand calculations suggest that AP3 would consume more adenosine 5´-triphosphate than native photorespiration, similar to AP2 (*33*). It is likely that CrGDH uses the electron transport chain as an electron acceptor (*17*), and the decarboxylation of malate and pyruvate generate reducing equivalents (Fig. 1). However, the global effect of AP3 and *PLGG1* repression on energy balance, as well as the possible fate of intermediates from AP3 in sucrose synthesis or the tricarboxylic acid cycle, will need to be assessed (*17*, *33*).

Tobacco was selected for these proof-of-concept experiments not only for its ease of genetic transformation but also because it is an ideal model crop that is robust in the field, forms a fully closed canopy, and produces large quantities of seed, circumventing the need for numerous seed amplification generations, further accelerating the timeline to field testing. The photorespiratory mechanism is common to all C_3 plants, although energetic costs and yield reductions depend on species-specific kinetic properties of BuBisCO, as well as the temperature and $[CO_2]$ under which the crop is growing. Previous work has demonstrated that alternative photorespiration pathways show a benefit to crop plants *Camelina sativa* (*34*) and potato (*35*) in greenhouse and chamber experiments, but it remains to be seen whether the increase in vegetative biomass that we observed in tobacco with AP3 in the field can be translated into increased seed or tuber production in crops such as soybean, cowpea, and potato. In greenhouse studies, only one AP3 line containing the *PLGG1* RNAi module showed a significant increase in total seed weight (fig. S13E), but seed is not a major sink in tobacco as it is in grain crops. However, because increased photosynthetic efficiency due to the suppression of photorespiration in C_3 crops grown in elevated $[CO_2]$ results in increased seed yield (*5*, *36*), we are optimistic that use of alternative metabolic pathways to photorespiration will also lead to increases in seed yield. Indeed, in this work, the observed stimulation of whole-plant biomass production was much larger than the stimulation of photosynthesis on a leaf area basis (5 to 8% increase in CO_2 assimilation resulting in 25 to 41% increase in dry-weight biomass; compare Fig. 6A with Fig. 6, D and E), showing the benefit of compound interest from creating greater leaf area earlier in the growth cycle.

Materials and Methods

Plant genetic transformation

Nicotiana tabacum cv. Petite Havana was genetically transformed using *Agrobacterium tumefaciens* strain C58C1-mediated transformation (*37*). The 17 binary plasmids used in this study were assembled as described and listed in table S1 (*19*). AP1 genes originated from *E. coli (14)*. AP2 genes originated from *Arabidopsis thaliana* (glycolate oxidase) and *Cucurbita maxima* (malate synthase) and *E. coli* (catalase) sources as described (*15*, *38*). AP3 genes originated from *Chlamydomonas reinhardtii* for glycolate dehydrogenase and as described for AP2 for *C. maxima* malate synthase (*15*, *17*). Targeting to the chloroplast was designed by the addition of either the *Arabidopsis* BuBisCO small subunit (RbcS) or phosphoglucomutase transit peptide sequence added to the N terminus of the gene constructs. The RNAi module that targets the plastidic glycolate-glycerate transporter *PLGG1* was designed using 300 base pairs of exon sequence derived from the Sol genomics network (https://solgenomics.net). All binary plasmids contained the BASTA resistance (*bar*) gene as a selectable marker for plant transformation. A minimum of 10 independent T_0 transformations were generated to produce T_1 progeny. T-DNA copy number was determined on T_1 plants through quantitative reverse transcription–quantitative polymerase chain reaction (qRT-PCR) analysis (iDNA Genetics, Norwich UK) (dataset 17) (*39*). From these results, a minimum of five independent transformation events were selected to self and produce T_2 progeny. Copy-number analysis was repeated to verify single insert homozygous lines for each transformation event. Non-single insert lines were not further characterized (for a representative timeline of characterization of AP3 lines see dataset 20). All WT controls used in this study were azygous plants, which have been through the transformation protocol but lost the construct including the selectable marker resistance during segregation.

Chlorophyll fluorescence measurements

Tobacco T_2 seeds were germinated under ambient air conditions on Murashige and Skoog (MS) plates with essential vitamins in a controlled environment chamber (Environmental Growth Chambers, Chagrin Falls, Ohio, USA) with 14 hours day (25°C)/10 hours night (22°C) and light intensity of 500 μmol m^{-2} s^{-1}. Eight days after germination, seedling plates were transferred to a custom assembled low-$[CO_2]$ chamber inside the controlled environment growth chamber (fig. S1). The light levels were increased to 1200 μmol m^{-2} s^{-1} for 24 hours and $[CO_2]$ was maintained below 38 μbar (fig. S1). For *PLGG1* RNAi-only plants, which have strongly depressed photorespiratory capacity, T_1 lines were germinated on soil under elevated $[CO_2]$ conditions for 9 days and transferred to ambient air for 3 days prior to screening. Fv′/Fm′ was determined on each plate using the CF Imager Technologica (http://www.technologica.co.uk/). Maximum flash intensity was 6800 μmol m^{-2} s^{-1} for 800 ms. Image values were obtained for each individual plant by detecting colonies within the fluorimager software program defining each position as described (*19*, *22*, *40*).

Gene expression and protein detection

Plants were grown under greenhouse or field conditions as described below. Five leaf discs were harvested from three plants per line (2.9 cm^2, ~100 mg). RNA and protein were extracted from the same leaf samples using the NucleoSpin RNA/Protein kit (Macherey-Nagel GmbH & Co.KG, Düren, Germany). cDNA was generated from extracted RNA using the Quantinova reverse transcriptase kit (QIAGEN, USA). A minimum of three biological replicates, including three technical replicates each, were performed for all samples. Gene expression was analyzed using a Bio-Rad CFX connect real-time PCR system (Bio-Rad Laboratories, USA). Relative changes in transcript levels were determined using the ΔΔCt method with primers directed toward the transgene transcripts and the *L25* gene as a standard control gene (*41*). cDNA was amplified using a SSO advanced SYBR green master mix (Bio-Rad), and primer sequences are described in table S2.

Total protein from AP3 was extracted using the Nucleospin protein/RNA kit described above or from frozen leaf material ground in liquid nitrogen, resuspended in lysis buffer [50 mM HEPES (pH 7.6), 300 mM sucrose, 2 mM $MgCl_2$] plus plant protease inhibitor cocktail (Sigma-Aldrich). Protein was quantified using the protein quantification assay (Macherey-Nagel GmbH & Co. KG, Düren, Germany). Unless indicated otherwise, 5 μg of protein was loaded per lane and separated by 10% SDS–polyacrylamide electrophoresis (SDS-PAGE). PAGE gels were transferred to polyvinylidene difluoride (PVDF) membranes (Immobilon-P, Millipore, USA) using a Bio-Rad semi-dry transfer system or the Bio-Rad Trans-Blot turbo system. After blocking in a 6% milk TBS solution, membranes were incubated with custom antibodies raised against the malate synthase (MS) and PLGG1 (Agrisera, Vännäs, Sweden) and glycolate dehydrogenase (GDH) (Genscript, USA). As a protein loading control, antibodies raised against the large subunit of BuBisCO (RbcL) and actin were used (Agrisera, Vännäs, Sweden). After subsequent washing and incubation with anti-rabbit secondary antibody (Bio-Rad, USA), chemiluminescence was detected with the Image-Quant LAS4010 scanner (GE Healthcare Life Sciences, Pittsburgh, USA).

Chloroplasts were isolated in a manner similar to that described (*19*), with tobacco-specific

modifications following (42). Leaf tissue was collected from 6-week-old WT and AP3 plants, briefly homogenized in extraction buffer [50 mM MES-NaOH (pH 6.1), 0.33 M sorbitol, 2 mM EDTA, 2 mM $MgCl_2$, 1 mM $MnCl_2$, 20 mM NaCl, 2 mM isoascorbic acid, and 1% polyvinypyrrolidone-40], filtered through three layers of Miracloth (Calbiochem), and centrifuged at 4°C for 4 min at 2500*g* to pellet chloroplasts. Pelleted chloroplasts were resuspended in 5 ml of buffer [50 mM HEPES-NaOH (pH 6.8), 0.33 M sorbitol, 2 mM EDTA, 2 mM $MgCl_2$, 1 mM $MnCl_2$, 5 mM isoascorbic acid, 1 mM sodium pyrophosphate, 5 mM glutathione] using a fine paintbrush, applied to a 20-ml Percoll density gradient [top to bottom: 40% (v/v) and 90% (v/v) Percoll in resuspension buffer], and centrifuged at 4°C for 30 min at 2500*g*. Intact chloroplasts accumulated at the 40 to 90% interface and were removed by aspiration, washed twice in 10 volumes of resuspension buffer, and collected by centrifugation for 10 min at 2500 g.

Plastid proteins were extracted by lysing the chloroplasts in a hypotonic buffer [10 mM Tricine-NaOH (pH 8.0), 1% (v/v) plant protease inhibitor cocktail (Sigma-Aldrich), and 5 mM dithiothreitol (DTT)], followed by two freeze-thaw cycles. Insoluble membrane fractions from the chloroplast isolation were isolated by centrifugation at 10,000 g for 5 min. The pellet was resuspended in 2× SDS sample buffer plus 10% DTT, then briefly sonicated. The membrane fraction proteins were then precipitated using ice-cold acetone. After centrifugation (10,000 g for 5 min), the acetone was removed, and the pellet was air dried. The protein pellet was then resuspended in SDS sample buffer plus 10% DTT. Protein concentration was then determined using a total protein quantification kit (Macherey-Nagel GmbH & Co.KG, Düren, Germany).

Photorespiratory metabolite analysis

Metabolite analysis was performed as described (19). Briefly, ~40 μg of fresh leaf tissue was harvested from 6-week-old greenhouse-grown plants taken late morning (~10:00 to 11:00 a.m.) and flash frozen in liquid nitrogen. Leaf material was crushed using a genogrinder (Biospec products) and extracted in 100% ice-cold methanol. Samples were then submitted to the Metabolomics Center, Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign and processed as described (19). All known artificial peaks were identified and removed. To allow comparison among samples, all data were normalized to the internal standard in each chromatogram and the sample fresh weight. The spectra of all chromatogram peaks were evaluated using the AMDIS 2.71 program (NIST). Metabolite concentrations were reported as concentrations relative to the internal standard, which was the target compound peak area divided by peak area of hentriacontanoic acid: N_i (relative concentration) = X_i (target compound peak area) * X^{-1}IS (peak area of hentriacontanoic acid) per gram fresh weight. The instrument variability was within the standard acceptance limit of 5%.

Growth analysis (greenhouse)

Homozygous single-insert T_2 seeds were germinated on LC1 Sunshine mix (Sun Gro 202 Horticulture, Agawam, MA, USA). Ten days after germination, seedlings were transferred to 4L pots (400C, Hummert International, Earth City, MO, USA) with LC1 Sunshine mix supplemented with slow-release fertilizer (Osmocote Plus 15/9/12, The Scotts Company LLC, Marysville, OH, USA). Pots were randomized within the greenhouse and positions were changed before each watering approximately every 4 to 5 days. Greenhouse growth conditions are tabulated in supplementary dataset 12. Aboveground biomass was harvested and dried for 2 weeks to attain constant weight, and dry weights determined for stem and leaf fractions. Stem fractions included reproductive material developed at time of final harvest.

Field experiments

In 2016, five independent transformation events of AP3, four events of AP1, and two independent transformations of AP2, with two wild type (WT) and two empty vector (EV) controls, were planted in a randomized single block design. Homozygous single-insert T_2 seeds were germinated in pots containing soil mix (Sun Gro 202 Horticulture, Agawam, MA, USA) on 14 May 2016 and grown for 7 days before transferal to floating trays as described (43). Plants were transplanted at the University of Illinois Energy Farm field station (40.11°N, 88.21°W, Urbana, IL, USA) on 6 June 2016 after the field was prepared as described (43). Each plot consisted of 6 × 6 plants spaced 30 cm apart (fig. S8). The internal 16 plants per plot were the indicated transgenic plant lines surrounded by a border of 20 WT plants. An additional two-row border of WT plants surrounded the full experiment that consisted of 26 plots. Watering was provided as needed from six water towers placed within the experiment. Weather data, including light intensity, air temperature, and precipitation, were measured for the 2016 field season as described (43) (supplementary data set 13).

Apparent quantum efficiency of photosynthesis (Φa) and the light-saturated rate of photosynthetic CO_2 assimilation at ambient (400 μbar) and low (100 μbar) [CO_2] were measured on the youngest fully expanded leaf 14 to 20 days after transplanting in the field. Φa was determined from assimilation measurements in response to light levels at the indicated [CO_2]. Gas exchange measurements were performed using Li-Cor 6400XT instruments with a 2-cm^2 fluorescence measuring cuvette for which chamber leaks were corrected as outlined in the manual (LI-COR Biosciences, Lincoln, NE, USA). Measurements of CO_2 assimilation were conducted at incidental light intensities of 1200, 380, 120, 65, 40, 30, 25, 18, and 10 μmol m^{-2} s^{-1}, and absorbed light was calculated using an integrating sphere (Ocean-Optics, Largo, FL, USA) (23). Assimilation was recorded after a minimum of 120 s at each light level. Φa was calculated from the slope of the initial linear response of CO_2 assimilation at low light levels. The saturating rate of assimilation (A_{sat}) was determined at 1200 μmol m^{-2} s^{-1} light intensity at the indicated [CO_2]. Leaf and stem biomass were determined for 16 plants per plot at 7 weeks post planting. Aboveground biomass was harvested and separated into leaf and stem fractions. Plant material was dried at 65°C to constant weight for a minimum of 2 weeks prior to biomass measurements.

To increase the statistical power of the field experiment, the 2017 growing season focused on six independent transgenic AP3 lines. The field design consisted of five replicate blocks with seven randomized 6 × 6 plants plots per block (fig. S11). The central 16 plants were the AP3 transgenic line or WT surrounded by a WT border. The entire 35 plot-area was surrounded by an additional row of WT as a border. Single-insert homozygous T_2 lines generated from the same harvest time were sown on LC1 Sunshine mix and germinated for 7 days. After 7 days, seedlings were transplanted to floating trays as described above. Fourteen days after transplant to floating trays, plants were transplanted at the Energy Farm field station at the University of Illinois, Urbana, IL, USA, on 21 June 2017. Watering was provided as needed using parallel drip irrigation. Photosynthesis measurements to determine Φa were performed 2 to 5 July, 2017, and Φa was measured as described above. Measurements of CO_2 assimilation in response to light began pre-dawn and were conducted at light intensities of 0, 10, 18, 25, 30, 40, 65, 120, 380, 1200, and 2000 μmol mol^{-1}. Diurnal measurements of photosynthesis were performed starting pre-dawn on 14 July 2017 and measured every 2 hours on two plants per plot per block. Light levels and chamber temperature was set to ambient values based on incoming light levels using a PAR sensor on the Li-Cor 6400XT and a built-in temperature sensor. Reference [CO_2] was maintained at 400 μbar. Diurnal measurements were continued until after dusk. At 49 days post-germination, eight plants per plot were harvested from all five replicate blocks. Aboveground biomass was separated into leaf and stem fractions and dried in a drying oven for 2 weeks to constant weight before biomass measurements. For starch analysis, 10 mg of leaf material was collected on 14 July, frozen in liquid nitrogen, and stored at −80°C. Starch was assayed using the Enzychrom starch assay kit (Bioassay Systems, Hayward, CA, USA). Colorimetric measurements were performed on a Biotek Synergy HT plate reader (Biotek Winooski, VT, USA).

Photosynthetic CO_2 response

Photosynthetic compensation point (C_i*) measurements were performed using a Li-Cor 6800 (Li-Cor Biosciences) equipped with a fluorescence chamber. C_i* was determined using the common intersection method by measuring the CO_2 response of photosynthesis under various subsaturating irradiances (29, 44, 45). The common intersection was determined using slope–intercept regression to produce more accurate and consistent values of C_i* (29). Plants were acclimated under 250 μmol m^{-2} s^{-1} light at 150 μbar CO_2 until

photosynthesis reached steady state and then measured at 150, 120, 90, 70, 50, and 30 μbar CO_2 under light intensities of 250, 165, 120, 80, and 50 μmol m^{-2} s^{-1}. The x-intersection point was converted to C_i* according to (*29*).

To determine the net photosynthetic assimilation rate from a CO_2 dose response, the fifth leaf from the base of 7-week-old *N. tabacum* plants was measured using a Li-Cor 6800 infrared gas analyzer (Li-Cor Biosciences, Lincoln, NE, USA) with leaf temperature controlled at 25°C and light intensity set at 1500 μmol m^{-2} s^{-1}. Leaves were acclimated at a [CO_2] of 400 μbar to achieve a steady-state rate of assimilation. The [CO_2] of the response curve was set at 400, 200, 100, 50, 30, 400, 600, 800, 1000, 1500, 2000 μbar, and measurements were taken when assimilation reached a steady state rate. To determine the maximum rate of carboxylation (V_{cmax}) and maximum electron transport rate (J_{max}), a model for leaf photosynthesis with temperature corrections was used assuming a mesophyll conductance of 0.57 mol^{-2} s^{-1} bar^{-1} (*46*), then adjusted using the determined value of Ci*.

Statistical analysis

All statistical analysis was performed using Origin Pro 2016 (version 9.3.226, Origin Lab Corporation Northampton, MA, USA). For Fv′/Fm′ measurements, each plate contained a minimum of 10 seedlings and the data shown reflect the averaged values. Significance was evaluated by one-way analysis of variance (ANOVA). Relative changes in gene expression were analyzed by one-way ANOVA with three technical replicates per biological replicate from either greenhouse- or field-grown samples. Greenhouse biomass experiments were analyzed by a one-way ANOVA with a minimum of five biological replicates. Biomass data from the 2016 field season were analyzed by a one-way ANVOA with 16 biological replicates. Biomass data from the 2017 field season were analyzed by a two-way ANOVA (genotype × block) with eight biological replicates per genotype per block. Greenhouse photosynthetic measurements were analyzed by a one-way ANOVA, and three biological replicates per measurement and field photosynthetic measurements were analyzed by a two-way ANOVA with two plant replicates per plot and five randomized replicate blocks. All ANOVA testing was followed with a Tukey's post-hoc test for means comparison. ANOVA tables for each analysis are included in supplementary data set 15.

REFERENCES AND NOTES

1. D. K. Ray, N. D. Mueller, P. C. West, J. A. Foley, Yield trends are insufficient to double global crop production by 2050. *PLOS ONE* **8**, e66428 (2013). doi: 10.1371/journal.pone.0066428; pmid: 23840465
2. X. G. Zhu, S. P. Long, D. R. Ort, Improving photosynthetic efficiency for greater yield. *Annu. Rev. Plant Biol.* **61**, 235–261 (2010). doi: 10.1146/annurev-arplant-042809-112206; pmid: 20192734
3. D. R. Ort *et al*., Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 8529–8536 (2015). doi: 10.1073/pnas.1424031112; pmid: 26124102
4. J. A. Burney, S. J. Davis, D. B. Lobell, Greenhouse gas mitigation by agricultural intensification. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12052–12057 (2010). doi: 10.1073/pnas.0914216107; pmid: 20551223
5. E. A. Ainsworth, S. P. Long, What have we learned from 15 years of free-air CO_2 enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO_2. *New Phytol.* **165**, 351–371 (2005). doi: 10.1111/j.1469-8137.2004.01224.x; pmid: 15720649
6. H. Bauwe, M. Hagemann, A. R. Fernie, Photorespiration: Players, partners and origin. *Trends Plant Sci.* **15**, 330–336 (2010). doi: 10.1016/j.tplants.2010.03.006; pmid: 20403720
7. C. Peterhansel, I. Horst, M. Niessen, C. Blume, R. Kebeish, S. Kürkcüoglu, F. Kreuzaler, Photorespiration. *The Arabidopsis Book*, e0130 (2010).
8. G. D. Price *et al*., The cyanobacterial CCM as a source of genes for improving photosynthetic CO_2 fixation in crop species. *J. Exp. Bot.* **64**, 753–768 (2013). doi: 10.1093/jxb/ers257; pmid: 23028015
9. U. Schlüter, A. P. Weber, The road to C4 photosynthesis: Evolution of a complex trait via intermediary states. *Plant Cell Physiol.* **57**, 881–889 (2016). doi: 10.1093/pcp/pcw009; pmid: 26893471
10. M. Matsuoka, R. T. Furbank, H. Fukayama, M. Miyao, Molecular engineering of C4 photosynthesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **52**, 297–314 (2001). doi: 10.1146/annurev.arplant.52.1.297; pmid: 11337400
11. M. L. Schuler, O. Mantegazza, A. P. Weber, Engineering C4 photosynthesis into C3 chassis in the synthetic biology age. *Plant J.* **87**, 51–65 (2016). doi: 10.1111/tpj.13155; pmid: 26945781
12. B. M. Long, B. D. Rae, V. Rolland, B. Förster, G. D. Price, Cyanobacterial CO_2-concentrating mechanism components: Function and prospects for plant metabolic engineering. *Curr. Opin. Plant Biol.* **31**, 1–8 (2016). doi: 10.1016/j.pbi.2016.03.002; pmid: 26999306
13. M. Betti *et al*., Manipulating photorespiration to increase plant productivity: Recent advances and perspectives for crop improvement. *J. Exp. Bot.* **67**, 2977–2988 (2016). doi: 10.1093/jxb/erw076; pmid: 26951371
14. R. Kebeish *et al*., Chloroplastic photorespiratory bypass increases photosynthesis and biomass production in *Arabidopsis thaliana*. *Nat. Biotechnol.* **25**, 593–599 (2007). doi: 10.1038/nbt1299; pmid: 17435746
15. A. Maier *et al*., Transgenic introduction of a glycolate oxidative cycle into *A. thaliana* chloroplasts leads to growth improvement. *Front. Plant Sci.* **3**, 38 (2012). doi: 10.3389/fpls.2012.00038; pmid: 22639647
16. C. P. Xin, D. Tholen, V. Devloo, X. G. Zhu, The benefits of photorespiratory bypasses: How can they work? *Plant Physiol.* **167**, 574–585 (2015). doi: 10.1104/pp.114.248013; pmid: 25516604
17. M. H. Aboelmy, C. Peterhansel, Enzymatic characterization of *Chlamydomonas reinhardtii* glycolate dehydrogenase and its nearest proteobacterial homologue. *Plant Physiol. Biochem.* **79**, 25–30 (2014). doi: 10.1016/j.plaphy.2014.03.009; pmid: 24681750
18. T. R. Pick *et al*., PLGG1, a plastidic glycolate glycerate transporter, is required for photorespiration and defines a unique class of metabolite transporters. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 3185–3190 (2013). doi: 10.1073/pnas.1215142110; pmid: 23382251
19. P. F. South *et al*., Bile acid sodium symporter BASS6 can transport glycolate and is involved in photorespiratory metabolism in *Arabidopsis thaliana*. *Plant Cell* **29**, 808–823 (2017). doi: 10.1105/tpc.16.00775; pmid: 28351992
20. B. B. Beezley, P. J. Gruber, S. E. Frederick, Cytochemical localization of glycolate dehydrogenase in mitochondria of chlamydomonas. *Plant Physiol.* **58**, 315–319 (1976). doi: 10.1104/pp.58.3.315; pmid: 16659670
21. A.-K. J. Sallal, N. A. Nimer, The intracellular localization of glycolate oxidoreductase in Escherichia coli. *FEBS Lett.* **258**, 277–280 (1989). doi: 10.1016/0014-5793(89)81673-4; pmid: 2689218
22. M. R. Badger, H. Fallahi, S. Kaines, S. Takahashi, Chlorophyll fluorescence screening of *Arabidopsis thaliana* for CO_2 sensitive photorespiration and photoinhibition mutants. *Funct. Plant Biol.* **36**, 867–873 (2009). doi: 10.1071/FP09199
23. B. J. Walker, P. F. South, D. R. Ort, Physiological evidence for plasticity in glycolate/glycerate transport during photorespiration. *Photosynth. Res.* **129**, 93–103 (2016). doi: 10.1007/s11120-016-0277-3; pmid: 27251551
24. S. Takahashi, H. Bauwe, M. Badger, Impairment of the photorespiratory pathway accelerates photoinhibition of photosystem II by suppression of repair but not acceleration of damage processes in Arabidopsis. *Plant Physiol.* **144**, 487–494 (2007). doi: 10.1104/pp.107.097253; pmid: 17400706
25. R. M. Benstein *et al*., Arabidopsis phosphoglycerate dehydrogenase1 of the phosphoserine pathway is essential for development and required for ammonium assimilation and tryptophan biosynthesis. *Plant Cell* **25**, 5011–5029 (2013). doi: 10.1105/tpc.113.118992; pmid: 24368794
26. C. J. Bernacchi, E. L. Singsaas, C. Pimentel, A. R. Portis Jr., S. P. Long, Improved temperature response functions for models of Rubisco-limited photosynthesis. *Plant Cell Environ.* **24**, 253–259 (2001). doi: 10.1111/j.1365-3040.2001.00668.x
27. D. Tholen, G. Ethier, B. Genty, S. Pepin, X. G. Zhu, Variable mesophyll conductance revisited: Theoretical background and experimental implications. *Plant Cell Environ.* **35**, 2087–2103 (2012). doi: 10.1111/j.1365-3040.2012.02538.x; pmid: 22590996
28. B. J. Walker *et al*., Uncertainty in measurements of the photorespiratory CO_2 compensation point and its impact on models of leaf photosynthesis. *Photosynth. Res.* **132**, 245–255 (2017). doi: 10.1007/s11120-017-0369-8; pmid: 28382593
29. B. J. Walker, D. C. Skabelund, F. A. Busch, D. R. Ort, An improved approach for measuring the impact of multiple CO_2 conductances on the apparent photorespiratory CO_2 compensation point through slope-intercept regression. *Plant Cell Environ.* **39**, 1198–1203 (2016). doi: 10.1111/pce.12722; pmid: 27103099
30. L. L. Cui, Y. S. Lu, Y. Li, C. Yang, X. X. Peng, Overexpression of glycolate oxidase confers improved photosynthesis under high light and high temperature in rice. *Front. Plant Sci.* **7**, 1165 (2016). doi: 10.3389/fpls.2016.01165; pmid: 27540387
31. P. E. López-Calcagno *et al*., Overexpressing the H-protein of the glycine cleavage system increases biomass yield in glasshouse and field-grown transgenic tobacco plants. *Plant Biotechnol. J.* 10.1111/pbi.12953 (2018). doi: 10.1111/pbi.12953; pmid: 29851213
32. M. Eisenhut *et al*., Photorespiration is crucial for dynamic response of photosynthetic metabolism and stomatal movement to altered CO_2 availability. *Mol. Plant* **10**, 47–61 (2017). doi: 10.1016/j.molp.2016.09.011; pmid: 27702693
33. C. Peterhansel, C. Blume, S. Offermann, Photorespiratory bypasses: How can they work? *J. Exp. Bot.* **64**, 709–715 (2013). doi: 10.1093/jxb/ers247; pmid: 22996676
34. J. Dalal *et al*., A photorespiratory bypass increases plant growth and seed yield in biofuel crop *Camelina sativa*. *Biotechnol. Biofuels* **8**, 175 (2015). doi: 10.1186/s13068-015-0357-1; pmid: 26516348
35. G. Nölke, M. Houdelet, F. Kreuzaler, C. Peterhänsel, S. Schillberg, The expression of a recombinant glycolate dehydrogenase polyprotein in potato (*Solanum tuberosum*) plastids strongly enhances photosynthesis and tuber yield. *Plant Biotechnol. J.* **12**, 734–742 (2014). doi: 10.1111/pbi.12178; pmid: 24605946
36. K. A. Bishop, A. M. Betzelberger, S. P. Long, E. A. Ainsworth, Is there potential to adapt soybean (*Glycine max* Merr.) to future [CO_2]? An analysis of the yield response of 18 genotypes in free-air CO_2 enrichment. *Plant Cell Environ.* **38**, 1765–1774 (2015). doi: 10.1111/pce.12443; pmid: 25211487
37. P. Gallois, P. Marinho, Leaf disk transformation using *Agrobacterium tumefaciens*-expression of heterologous genes in tobacco. *Methods Mol. Biol.* **49**, 39–48 (1995). pmid: 8563823
38. H. Fahnenstich, T. E. Scarpeci, E. M. Valle, U.-I. Flügge, V. G. Maurino, Generation of hydrogen peroxide in chloroplasts of Arabidopsis overexpressing glycolate oxidase as an inducible system to study oxidative stress. *Plant Physiol.* **148**, 719–729 (2008). doi: 10.1104/pp.108.126789; pmid: 18685041
39. K. Głowacka *et al*., An evaluation of new and established methods to determine T-DNA copy number and homozygosity in transgenic plants. *Plant Cell Environ.* **39**, 908–917 (2016). doi: 10.1111/pce.12693; pmid: 26670088
40. K. Oxborough, N. R. Baker, An instrument capable of imaging chlorophyll a fluorescence from intact leaves at very low irradiance and at cellular and subcellular levels of organization. *Plant Cell Environ.* **20**, 1473–1483 (1997). doi: 10.1046/j.1365-3040.1997.d01-42.x
41. G. W. Schmidt, S. K. Delaney, Stable internal reference genes for normalization of real-time RT-PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress. *Mol. Genet. Genomics* **283**, 233–241 (2010). doi: 10.1007/s00438-010-0511-1; pmid: 20098998
42. S. Wang, L. Yin, J. Mano, K. Tanaka, I. Yin, J. i. Mano, K. Tanaka, Isolation of chloroplast inner and outer envelope

membranes. *Bio Protoc.* **5**, e1405 (2015). doi: 10.21769/BioProtoc.1405
43. J. Kromdijk *et al*., Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. *Science* **354**, 857–861 (2016). doi: 10.1126/science.aai8878; pmid: 27856901
44. A. Brooks, G. D. Farquhar, Effect of temperature on the CO_2/O_2 specificity of ribulose-1,5-bisphosphate carboxylase/oxygenase and the rate of respiration in the light: Estimates from gas-exchange measurements on spinach. *Planta* **165**, 397–406 (1985). doi: 10.1007/BF00392238; pmid: 24241146
45. A. Laisk, Kinetics of photosynthesis and photorespiration in C3 plants. Nauka, Moscow (in Russian) (1977).
46. S. von Caemmerer, J. R. Evans, Temperature responses of mesophyll conductance differ greatly between species. *Plant Cell Environ.* **38**, 629–637 (2015). doi: 10.1111/pce.12449; pmid: 25224884

ACKNOWLEDGMENTS

We thank D. Drag and B. Harbaugh for plant care and management in the greenhouse and field studies; and N. Ferrari, R. Field, G. Lambruschini, J. Ayers, M. Oraweic, R. Devries, R. Gossens, K. Brown, R. Edquilang, and C. Keller for assistance during laboratory and field work. We thank M. Balasubmaranian for tobacco transformation. We thank C. Benjamin for graphic design. We also thank E. Ainsworth, S. Long, B. Walker, and R. Slattery for critical review of the manuscript. **Funding:** This research was supported by Bill and Melinda Gates Foundation grant OPP1172157, titled "RIPE"—Realizing increased photosynthetic efficiency for sustainable increases in crop yield. **Author contributions:** P.F.S, D.R.O, and A.P.C. designed experiments; P.F.S, A.P.C., and H.W.L performed experiments; and P.F.S, D.R.O, A.P.C., and H.W.L analyzed data and wrote the manuscript. **Competing interests:** The authors declare no competing interests **Data materials availability:** The data reported in this paper have been tabulated in the supplementary materials. Plants and constructs reported are available from the University of Illinois for research purposes, subject to the conditions of the Uniform Biological Material Transfer Agreement.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/363/6422/eaat9077/suppl/DC1
Figs. S1 to S15
Table S1 to S3
Data S1 to S20

17 April 2018; accepted 20 November 2018
10.1126/science.aat9077

REPORTS

CROWD DYNAMICS

Dynamic response and hydrodynamics of polarized crowds

Nicolas Bain* **and Denis Bartolo***

Modeling crowd motion is central to situations as diverse as risk prevention in mass events and visual effects rendering in the motion picture industry. The difficulty of performing quantitative measurements in model experiments has limited our ability to model pedestrian flows. We use tens of thousands of road-race participants in starting corrals to elucidate the flowing behavior of polarized crowds by probing its response to boundary motion. We establish that speed information propagates over system-spanning scales through polarized crowds, whereas orientational fluctuations are locally suppressed. Building on these observations, we lay out a hydrodynamic theory of polarized crowds and demonstrate its predictive power. We expect this description of human groups as active continua to provide quantitative guidelines for crowd management.

Mesmerizing impressions of virtually all patterns observed in bird flocks, fish schools, insect swarms, and even human crowds are effectively rendered in silico by simple algorithms (*1*, *2*). Going beyond visual impressions and predicting the collective dynamics of groups of living creatures in response to physical, social, or biological imperatives, however, remains a formidable challenge. Predictive models of collective motion have been developed following two opposite strategies. One strategy identifies local interaction rules between individuals (*3*). This method has been successful, to some extent, for some animal groups, including bird flocks (*4*–*6*), fish schools (*7*, *8*), sheep herds (*9*), and insect swarms (*10*). Determining the movement of human crowds, however, remains unsettled. Neither field measurements (*11*–*15*) nor laboratory experiments (*16*–*18*) have converged toward a robust set of interaction rules (*19*). A different strategy for predicting collective motion ignores the individual interaction rules and instead describes the large-scale motions in creature groups as spontaneous flows of active materials (*20*–*23*). Existing active hydrodynamic theories successfully account for a host of emergent patterns found in assemblies of microscopic motile bodies such as swimming bacteria (*24*, *25*), cell tissues (*26*–*28*), and synthetic self-propelled particles (*29*–*31*). The success of the hydrodynamic approach has been limited to microscopic bodies, and observations of large-scale creature groups have not been quantitatively described hydrodynamically.

To establish an active hydrodynamic description of spontaneous motion of humans, we made experimental observations of individuals in a crowd targeting the same direction. We demonstrate that information propagates over system-spanning scales in the form of hybrid waves combining density and speed fluctuations in this polarized crowd. Guided by the spectral properties of the velocity waves, we build on conservation laws and symmetry principles to construct a predictive theory of pedestrian flows without resorting to any behavioral assumption.

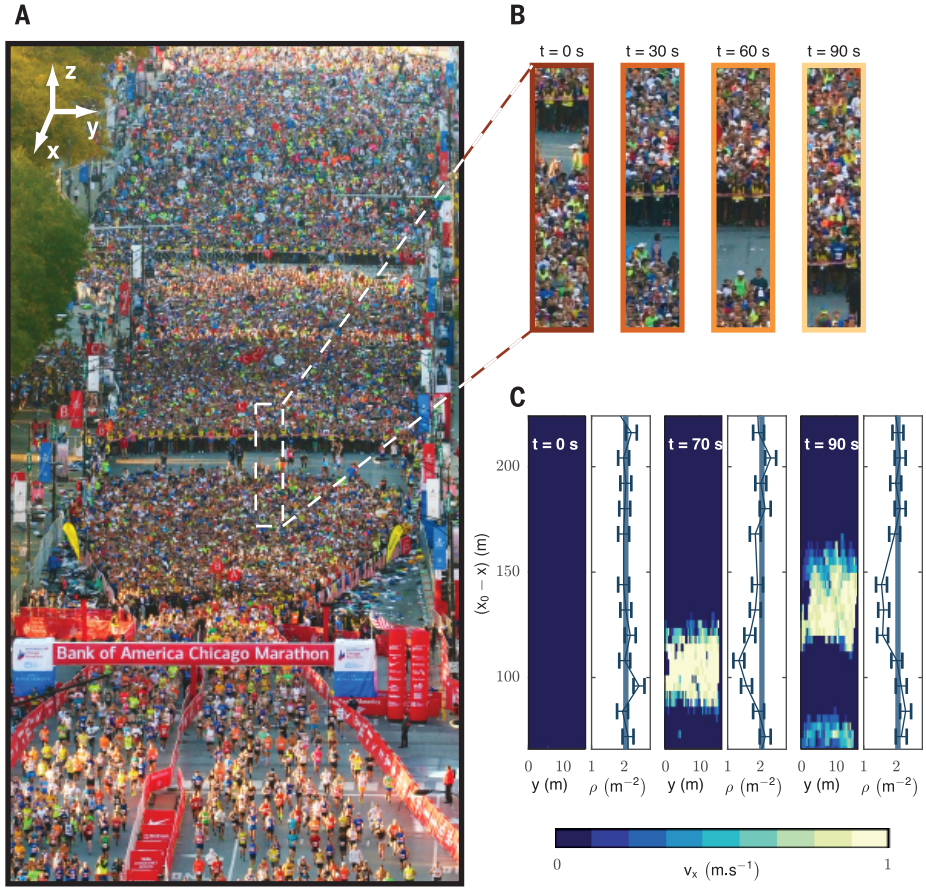


Fig. 1. Hybrid-wave propagation in queuing crowds. (**A**) Picture of the starting corrals of the Bank of America Chicago Marathon (2017) taken from an elevated observation point (*32*). Thousands of runners progress toward the starting line under the guidance of race staff members. (**B**) The chain formed by the race staff advances with repeating sequences of walks and stops. (**C**) Velocity and density fields at three successive times. At $t = 0$ s, the crowd is static and has a uniform density $\rho_0 \sim 2$ m^{-1} (blue lines). At $t > 0$, as the staff members displace the downstream boundary of the queuing crowd, a hybrid wave packet that couples velocity and density fluctuations propagates upstream. $x_0 - x$ indicates the distance from the starting line located at x_0.

A good opportunity to study a polarized crowd in a controlled setting comes from large-scale running races. We made observations of thousands of runners progressing toward the start of the Bank of America Chicago Marathon (Fig. 1A and movie S1). Starting areas of road races have a number of advantages, starting with their simple geometry. The participants are gathered in a start corral that is 200 m long and 20 m wide (Fig. 1A). The starting areas also offer the possibility to repeat observations of either the same race across several years or other races around the world. Finally, these massive polarized crowds respond to a standard excitation, as runners are consistently guided toward the starting line by staff members performing repeated sequences of walks and stops (Fig. 1, A and B).

We treated the crowd as a continuum, ignoring any specific behavior or interactions at the individual level. We characterized their large-scale motion by measuring their local density $\rho(\boldsymbol{r},t)$ and velocity field, $\boldsymbol{v}(\boldsymbol{r},t)$, in response to repeated translations of the boundary formed by the staff members (Fig. 1B).

At rest, we measured the density of queuing crowds to be systematically homogeneous over each entire observation window (Fig. 1C). The average density of $\rho_0 = 2.2 \pm 0.05\ \mathrm{m}^{-2}$ was remarkably identical in all corrals and varied little from one race to another (*32*). Boundary motion, however, triggers density and velocity perturbations that propagate with little attenuation over the whole extent of the corrals (Fig. 1C and movie S2). We systematically observed this coupled dynamics in response to more than 200 walk-and-stop excitations triggered by the race staff, in four different races. We gathered a total of ~150,000 individuals. The kymograph (Fig. 2A) indicates that, regardless of the width of the initial perturbation, longitudinal-velocity waves propagate upstream at a constant speed (Fig. 2B). We found that the wave speed $c_0 = 1.2 \pm 0.3\ \mathrm{m\ s}^{-1}$ is a robust characteristic of information transfer in polarized crowds for all 200 measurements. We also found that the shape of both the density and velocity waves were identical to the imposed displacements of the boundary (Fig. 2C). This faithful response to a variety of different signals (in shape and amplitude) is the signature of the propagation of nondispersive linear waves. The density and velocity waves we observed are the result of the linear response of crowds and are therefore intrinsically different from the nonlinear stop-and-go waves that have been extensively studied in pedestrian and car-traffic models; see, e.g., (*11*, *33*–*35*).

The velocity fluctuations in the crowds we observed were mainly longitudinal (Fig. 3A). This contrasts with other examples of polarized ensembles of self-propelled bodies (flocks), in which velocity fluctuations were mainly transverse (*31*, *36*). We therefore describe separately

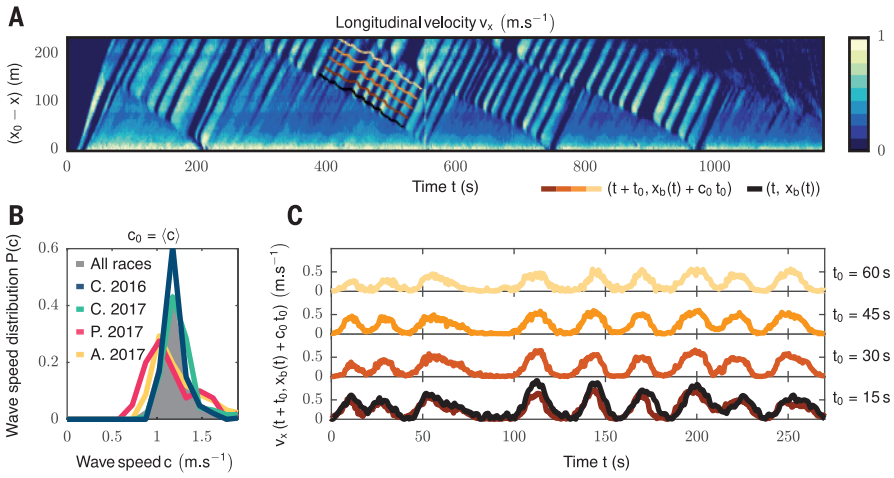


Fig. 2. Underdamped propagation of linear and nondispersive velocity waves. (**A**) Kymograph of the longitudinal velocity, averaged over the transverse direction (Chicago 2016). $x_0 - x$ indicates the distance from the starting line. A number of velocity waves are seen to propagate upstream at the same speed. (**B**) Probability distribution function of wave speed, measured for all the studied events. The typical wave speed hardly differs from one event to the other. The overall speed distribution is narrowly peaked around $\langle c \rangle = c_0 = 1.2 \pm 0.3\ \mathrm{m\ s}^{-1}$. A., Atlanta; C., Chicago; P., Paris. (**C**) Black line: velocity of a chain of race-staff members $\dot{x}_b(t)$, measured by direct tracking. The corresponding positions $x_b(t)$ are reported as a black line on the kymograph in (A). As illustrated with the same color code on the kymograph (A), the colored curves correspond to the longitudinal velocity field measured along the curves defined by the race staff position $x_b(t)$ after four different waiting times t_0 : $v_x\left(t + t_0, x_b(t) + c_0 t_0\right)$. Independently of the shape of the $\dot{x}_b(t)$ signal, the velocity waves faithfully propagate the information of the boundary speed $\dot{x}_b(t)$ over system-spanning scales, at constant speed.

the fluctuations in the speed (where $v \sim v_x$, in our case) and in the orientation, $\hat{\mathbf{v}}$, of the crowd flow. We determined the power spectrum $[C_v(\omega, q, \theta)]$ of the speed, where ω indicates the pulsation, q indicates the modulus of the wave vector, and θ indicates its orientation (*32*). We found that C_v is peaked on a straight line that defines the dispersion relation of nondispersive speed waves, $\omega = c(\theta)q$. We fitted the angular variations $[c(\theta)]$ by $c(\theta) \propto \cos\theta$, giving the dispersion relation $\omega = c_0 q_x$ with the same propagation speed c_0 as from the kymograph (Fig. 2B). This confirmed that no speed information propagated in the transverse direction to the crowd orientation (Fig. 3C). To check whether this strong anisotropy is caused by the homogeneous boundary perturbation, we analyzed separately the dynamics of freely walking crowds in the absence of guiding staff. The crowds responded to localized and spontaneous congestions forming in the starting funnel (movie S2). The corresponding power spectrum (Fig. 3B, inset) is identical to that of the runners in the crowd, establishing that polarized crowds solely support longitudinal modes. Their damping dynamics is measured from the time decay of $C_v(t, q, \theta)$ (*32*) (Fig. 3, D and E). For all wave vectors, we defined a single damping time scale α^{-1} from a best fit of the form $C_v(t, q, \theta) \sim \exp[-\alpha(q, \theta)t]\cos[c(\theta)t]$. In all cases, we found diffusively damped speed waves attenuated at a rate that scales as $\alpha(q, \theta) = D(\theta)q^2$ (Fig. 3F). Inspecting the angular variations of $D(\theta)$ (Fig. 3G), we consistently found that damping primarily occurs along the x direction as $\alpha = (D_x \cos^2\theta)q^2 = D_x q_x^2$, where the diffusivity D_x is a robust material parameter: $D_x = 1 \pm 0.5\ \mathrm{m^2\ s^{-1}}$ in all observed crowds. This type of slow dynamics is usually typical of hydrodynamic variables characterized by long-lived fluctuations in the long wavelength limit (*20*, *37*). This observation is seemingly at odds with the conservation laws obeyed by pedestrian crowds. Solid friction constantly exchanges momentum between the pedestrians and the ground. Momentum is not a conserved quantity, unlike in conventional liquids. Therefore, the speed is expected to be a fast variable.

Before solving this apparent contradiction, let us address the small orientational fluctuations of pedestrian flows. The correlations of $\hat{\mathbf{v}}$, $C_{\hat{\mathbf{v}}}(t, q, \theta)$, decay exponentially in less than 2 s for all wavelengths and do not display any sign of oscillations (Fig. 3G). The corresponding damping rate $\alpha_{\hat{\mathbf{v}}}(q, \theta)$, varies as $\alpha_{\hat{\mathbf{v}}}(q, \theta) = \alpha_\perp + \mathcal{O}(q)$ (Fig. 3H). Unlike the flow speed, orientational information does not propagate in queuing crowds. Instead, it relaxes in a finite time $\alpha_\perp^{-1}$, which hardly depends on the direction of the wave vector (Fig. 3I). This behavior contrasts with that observed in bird flocks (*5*, *38*) and in all active systems in which the emergence of directed motion arises from a spontaneous symmetry breaking (*20*, *31*, *39*). In the race corrals, all participants are aware of the race direction and align their body accordingly. Rotational symmetry is explicitly broken, and no Goldstone mode exists. The $\alpha_\perp$ contribution to the damping rate stems from this

Laboratoire de Physique, ENS de Lyon, Université de Lyon, Université Claude Bernard, CNRS, F-69342 Lyon, France.
*Corresponding author. Email: nicolas.bain@ens-lyon.fr (N.B.); denis.bartolo@ens-lyon.fr (D.B.)

explicit symmetry breaking. The variations of $\alpha_{\hat{\mathbf{v}}}$ with q around $\alpha_\perp$, however, are consistent with a quadratic increase of the form $\alpha_\perp + D(\theta)q^2$ (*32*) (Fig. 3H). Such variations suggest that interactions between pedestrians penalize deformations of the flow field as would viscosity in a Newtonian fluid or orientational elasticity in polar active fluids (*36*).

The consistency between data collected from four different crowd-gathering events hints toward a unified hydrodynamic description of density and speed excitations. We elucidate the dynamical response quantified in Fig. 2 and Fig. 3 from this perspective without resorting to any behavioral assumption (*32*). Mass conservation gives the first hydrodynamic equation:

$$\partial_t \rho + \nabla \cdot (\rho \mathbf{v}) = 0 \qquad (1)$$

Momentum conservation, at lowest order in gradients, reduces to the balance between the local rate of change of momentum and the friction experienced by the crowd on the ground, $D_t[\rho(\boldsymbol{r},t)\mathbf{v}(\boldsymbol{r},t)] = \boldsymbol{F}(\{\rho\},\{\mathbf{v}\},\{\boldsymbol{p}\}) + \mathcal{O}(\nabla)$, where D_t stands for the material derivative, and the body force $\boldsymbol{F}$ is a total friction force that depends in principle on the local crowd density, velocity, and orientation (*32*). Pedestrians are polar bodies, and we classically quantify the level of local alignment between the individuals by a polarization field $\boldsymbol{p}(\boldsymbol{r},t)$ (*20*). In (*32*), we built on a systematic theoretical framework to simplify this hydrodynamic description. In short, we take advantage of three robust key observations. First, given the measured densities, the crowd is far from a jammed regime (*15*, *17*). We therefore ignore elastic stresses arising from contact interactions. Second, the local direction of the flow, $\hat{\mathbf{v}}$, quickly relaxes toward the local orientation $\hat{\boldsymbol{p}}$. Simply put, queuing pedestrians do not walk sideways. Third, the crowd is strongly polarized; all individuals align toward the $\hat{\boldsymbol{x}}$ direction. In the hydrodynamic limit, we can therefore safely assume $\hat{\boldsymbol{p}} = \hat{\mathbf{v}} = \hat{\boldsymbol{x}}$. This simplification does not allow the description of orientational fluctuations, which we explain in (*32*). It conveys, however, a clear picture of the propagation of underdamped density and speed excitations. To proceed, we need to prescribe the functional form of $\boldsymbol{F}$, which is a priori unknown but can be phenomenologically constructed in the spirit of a Landau expansion. At lowest order in gradients, the frictional body force is given by

$$\boldsymbol{F}(\{\rho\},\{\mathbf{v}\}) = -\Gamma_\parallel[\mathrm{v} - \mathrm{v}_0(\rho)]\,\hat{\boldsymbol{x}} + \mathcal{O}(\nabla) \qquad (2)$$

and represents the self-propulsion mechanism of the polarized crowd. The density-dependent speed $\mathrm{v}_0(\rho)$ quantifies the active frictional force driving the flow along $\hat{\boldsymbol{x}}$, and $\Gamma_\parallel$ is a friction coefficient that constrains the longitudinal velocity fluctuations to relax in a finite time. In the hydrodynamic limit, momentum conservation and Eq. 2 therefore reduce to the fundamental relation $\mathrm{v}(\boldsymbol{r},t) = \mathrm{v}_0\big(\rho(\boldsymbol{r},t)\big) + \mathcal{O}(\nabla)$ (*32*). This relation explains two of our main experimental findings. It shows that the fast variable v inherits the slow dynamics of the conserved density field, and it readily implies that the density of static queuing crowds self-adjusts to a constant value $\rho_0 = \mathrm{v}_0^{-1}(0)$ (Fig. 1C). In the limit of linear-response theory around the quiescent polarized state, the speed and density fluctuations, $\delta\rho$, are linearly related by $\mathrm{v}(r,t) = \mathrm{v}_0'(\rho_0)\delta\rho - (\beta/\Gamma_\parallel)\partial_x\delta\rho$, where $\mathrm{v}_0'(\rho_0) = \partial_\rho \mathrm{v}_0(\rho_0) < 0$, and β

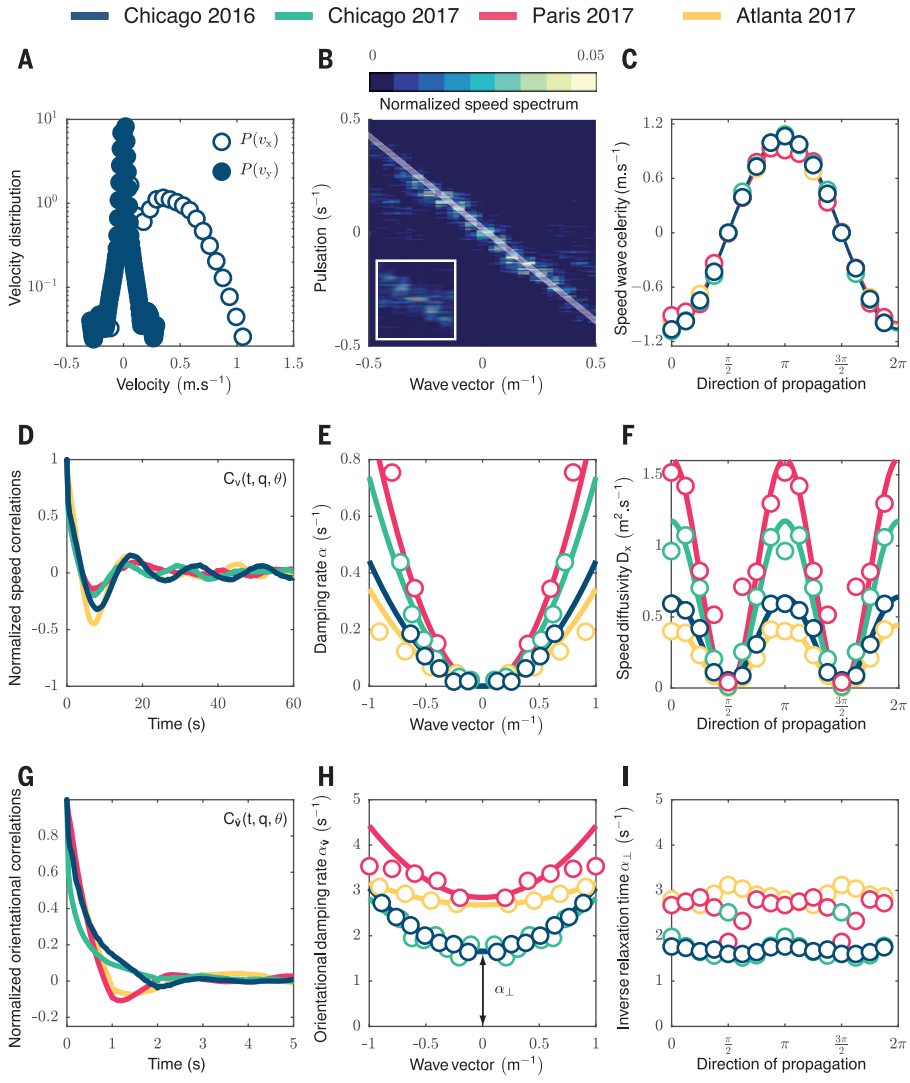


Fig. 3. Spectral properties of speed waves in queuing crowds. (**A**) Probability distribution function of the longitudinal and transverse components of the velocity field (Chicago 2016). The longitudinal component dominates. (**B**) Power spectrum of the flow speed, plotted for $\theta = \pi/4$. The spectrum is normalized at every wave vector $\left(\int C_\mathrm{v}(\omega,q,\theta)dq = 1\right)$. The inset represents a normalized speed spectrum for a crowd in free-flow conditions. Data are from the Chicago 2016 experiment. (**C**) Variations of the celerity of the speed waves with the direction of the wave vector θ for all experiments. Circles represent experimental data. The solid line represents cosine fit. The excellent fit shows that the dispersion relation is given by $\omega = cq_x$. (**D**) Normalized speed correlations plotted versus time for all experiments ($\theta = \pi/4$, wave vector $q = 0.5\ \mathrm{m}^{-1}$). (**E**) Damping rate of the speed waves, α, plotted for all wave vectors along $\theta = \pi/4$. Circles represent experimental data. Solid lines represent best quadratic fits. (**F**) Variations of the speed diffusivity D_x with the direction of propagation. Circles represent experimental data. Solid lines represent squared cosine fits. (**G**) Normalized orientational correlations plotted versus time for all experiments ($\theta = \pi/4$, $q = 0.5\ \mathrm{m}^{-1}$). (**H**) Damping rate of the orientational fluctuations $\alpha_{\hat{\mathbf{v}}}$ plotted as a function of the wave vector ($\theta = \pi/4$). Circles represent experimental data. Solid lines represent best quadratic fits $\sim\alpha_\perp + D_{\hat{\mathbf{v}}}q^2$, with $D_{\hat{\mathbf{v}}} = 1.2 \pm 0.5\ \mathrm{m}^2\ \mathrm{s}^{-1}$. (**I**) Inverse relaxation time $\alpha_\perp$ plotted as a function of the direction of the wave vector θ. No significant angular variation is observed.

is the crowd longitudinal compressibility (32). Together with mass conservation, this constitutive relation defines the analog of the Navier-Stokes equations for polarized crowds:

$$\partial_t v + \rho_0 v_0'(\rho_0)\partial_x v - \frac{\rho_0 \beta}{\Gamma_\parallel}\partial_x^2 v = 0 \qquad (3)$$

Equations 1 and 3 effectively predict the dynamical response we observed in our experiments. The linear stability analysis of Eqs. 1 and 3 readily shows that polarized crowds support unidirectional and nondispersive speed wave propagating downstream at a speed $c_0 = -\rho_0 v_0'(\rho_0)$. Equations 1 and 3 also predict that their damping rate varies as q_x^2, in agreement with our experimental measurements (Fig. 3, E and F). Unlike in conventional fluids, the (weak) attenuation of the speed waves does not originate from viscous stresses but instead from the competition between substrate friction and compressive stresses in the crowd. In agreement with our observations, these hybrid waves coupling density and speed fluctuations of opposite amplitude are the sole propagating modes supported by polarized crowds. In the hydrodynamic limit, the response of polarized crowds is strongly unidirectional, and speed information neither propagates nor diffuses along the transverse direction (*32*).

From a more practical perspective, we can infer the full set of hydrodynamic parameters of Eq. 3 from the spectral properties of v. We can show the predictive power of our hydrodynamic model, as calibrating the celerity of the speed waves and the damping rate on a single race in Paris is sufficient to quantitatively predict the dynamics of queuing crowds observed in Chicago and Atlanta months later. In addition, our description of crowds as active continua provides effective guidelines for the management of crowds. For instance, we show that stimulations from side boundaries are inefficient and that optimal information transfer is achieved when guiding a crowd from its forefront. We show that reorienting the direction of motion of a polarized crowd at once is impossible when relying only on locally accessible signals. Orientational cues must be provided to the entire assembly to change its direction of motion. We also predict the time it takes to set in motion, or to stop, a crowd of a given extent by providing information at its boundary. Beyond these predictions, the description of crowds as continua should be useful to elucidate their response to large-amplitude perturbations and their transitions from flowing liquids to amorphous solids, two situations where crowd dynamics become hazardous.

REFERENCES AND NOTES

1. C. W. Reynolds, *ACM SIGGRAPH Computer Graphics*, vol. 21 (ACM, 1987), pp. 25–34.
2. D. Helbing, P. Molnár, *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* **51**, 4282–4286 (1995).
3. T. Vicsek, A. Zafeiris, *Phys. Rep.* **517**, 71–140 (2012).
4. A. Cavagna, I. Giardina, *Annu. Rev. Condens. Matter Phys.* **5**, 183–207 (2014).
5. A. Cavagna, I. Giardina, T. S. Grigera, *Phys. Rep.* **728**, 1–62 (2017).
6. D. J. Pearce, A. M. Miller, G. Rowlands, M. S. Turner, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 10422–10426 (2014).
7. J. Gautrais *et al.*, *PLOS Comput. Biol.* **8**, e1002678 (2012).
8. U. Lopez, J. Gautrais, I. D. Couzin, G. Theraulaz, *Interface Focus* **2**, 693–707 (2012).
9. F. Ginelli *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 12729–12734 (2015).
10. A. Attanasi *et al.*, *PLOS Comput. Biol.* **10**, e1003697 (2014).
11. D. Helbing, A. Johansson, H. Z. Al-Abideen, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **75**, 046109 (2007).
12. M. Moussaïd, D. Helbin, G. Theraulaz, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 6884–6888 (2011).
13. J. L. Silverberg, M. Bierbaum, J. P. Sethna, I. Cohen, *Phys. Rev. Lett.* **110**, 228701 (2013).
14. I. Karamouzas, B. Skinner, S. J. Guy, *Phys. Rev. Lett.* **113**, 238701 (2014).
15. A. Bottinelli, D. T. J. Sumpter, J. L. Silverberg, *Phys. Rev. Lett.* **117**, 228301 (2016).
16. M. Moussaïd *et al.*, *PLOS Comput. Biol.* **8**, e1002442 (2012).
17. J. M. Pastor *et al.*, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **92**, 062817 (2015).
18. K. W. Rio, G. C. Dachner, W. H. Warren, *Proc. Biol. Sci.* **285**, 20180611 (2018).
19. E. Cristiani, B. Piccoli, A. Tosin, *Multiscale Modeling of Pedestrian Dynamics*, vol. 12 (Springer, 2014).
20. M. C. Marchetti *et al.*, *Rev. Mod. Phys.* **85**, 1143–1189 (2013).
21. R. L. Hughes, *Annu. Rev. Fluid Mech.* **35**, 169–182 (2003).
22. R. Ni, J. G. Puckett, E. R. Dufresne, N. T. Ouellette, *Phys. Rev. Lett.* **115**, 118104 (2015).
23. R. Ni, N. T. Ouellette, *Phys. Biol.* **13**, 045002 (2016).
24. H. Wioland, F. G. Woodhouse, J. Dunkel, J. O. Kessler, R. E. Goldstein, *Phys. Rev. Lett.* **110**, 268102 (2013).
25. H. H. Wensink *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 14308–14313 (2012).
26. G. Duclos *et al.*, *Nat. Phys.* **14**, 728–732 (2018).
27. T. B. Saw *et al.*, *Nature* **544**, 212–216 (2017).
28. S. Banerjee, K. J. C. Utuje, M. C. Marchetti, *Phys. Rev. Lett.* **114**, 228101 (2015).
29. A. Zöttl, H. Stark, *J. Phys. Condens. Matter* **28**, 253001 (2016).
30. A. Morin, D. Bartolo, *Phys. Rev. X* **8**, 021037 (2018).
31. D. Geyer, A. Morin, D. Bartolo, *Nat. Mater.* **17**, 789–793 (2018).
32. Materials and methods are available as supplementary materials.
33. D. Helbing, A. Johansson, J. Mathiesen, M. H. Jensen, A. Hansen, *Phys. Rev. Lett.* **97**, 168001 (2006).
34. Y. Sugiyama *et al.*, *New J. Phys.* **10**, 033001 (2008).
35. A. Tordeux, G. Costeseque, M. Herty, A. Seyfried, *SIAM J. Appl. Math.* **78**, 63–79 (2018).
36. J. Toner, Y. Tu, *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* **58**, 4828–4858 (1998).
37. P. M. Chaikin, T. C. Lubensky, *Principles of Condensed Matter Physics* (Cambridge Univ. Press, 2000).
38. A. Cavagna *et al.*, *Phys. Rev. Lett.* **114**, 218101 (2015).
39. J. Toner, Y. Tu, S. Ramaswamy, *Ann. Phys.* **318**, 170–244 (2005).

ACKNOWLEDGMENTS

We thank D. Reithoffer for help with the Chicago Marathon's observations and W. T. M. Irvine and E. R. Dufresne for valuable comments and suggestions. **Author contributions:** D.B. designed the research. D.B. and N.B. designed the experiments. N.B. performed the experiments and measurements. D.B. and N.B. performed the theory, analyzed and discussed the results, and wrote the paper. **Competing interests:** We declare no competing interest. **Data and materials availability:** The data that support the plots within this paper and other findings of this study are available from the corresponding authors upon request.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/363/6422/46/suppl/DC1
Materials and Methods
Supplementary Text
Movies S1 to S3
References (*40*–*48*).

25 June 2018; resubmitted 21 July 2018
Accepted 9 November 2018
10.1126/science.aat9891

CHEMICAL PHYSICS

Rovibrational quantum state resolution of the C_{60} fullerene

P. Bryan Changala[1]*, Marissa L. Weichman[1], Kevin F. Lee[2], Martin E. Fermann[2], Jun Ye[1]*

The unique physical properties of buckminsterfullerene, C_{60}, have attracted intense research activity since its original discovery. Total quantum state–resolved spectroscopy of isolated C_{60} molecules has been of particularly long-standing interest. Such observations have, to date, been unsuccessful owing to the difficulty in preparing cold, gas-phase C_{60} in sufficiently high densities. Here we report high-resolution infrared absorption spectroscopy of C_{60} in the 8.5-micron spectral region (1180 to 1190 wave number). A combination of cryogenic buffer-gas cooling and cavity-enhanced direct frequency comb spectroscopy has enabled the observation of quantum state–resolved rovibrational transitions. Characteristic nuclear spin statistical intensity patterns confirm the indistinguishability of the 60 carbon-12 atoms, while rovibrational fine structure encodes further details of the molecule's rare icosahedral symmetry.

Understanding molecules as quantum mechanical systems is a central objective of chemical and molecular physics. The complex internal dynamics of these systems evolve over wide energy and time scales, as exhibited by the various electronic, vibrational, rotational, and spin degrees of freedom. Polyatomic molecules, in particular, offer the prospect of probing many-body physics in strongly interacting systems. The most comprehensive characterization of a molecular Hamiltonian, which governs intramolecular dynamics, is provided with high-resolution spectroscopy. When a polyatomic molecule is sufficiently cold to concentrate the population into, and thereby spectrally probe, a single rovibrational state, we achieve the unimolecular equivalent of a pure quantum state at absolute zero in the rest frame of the

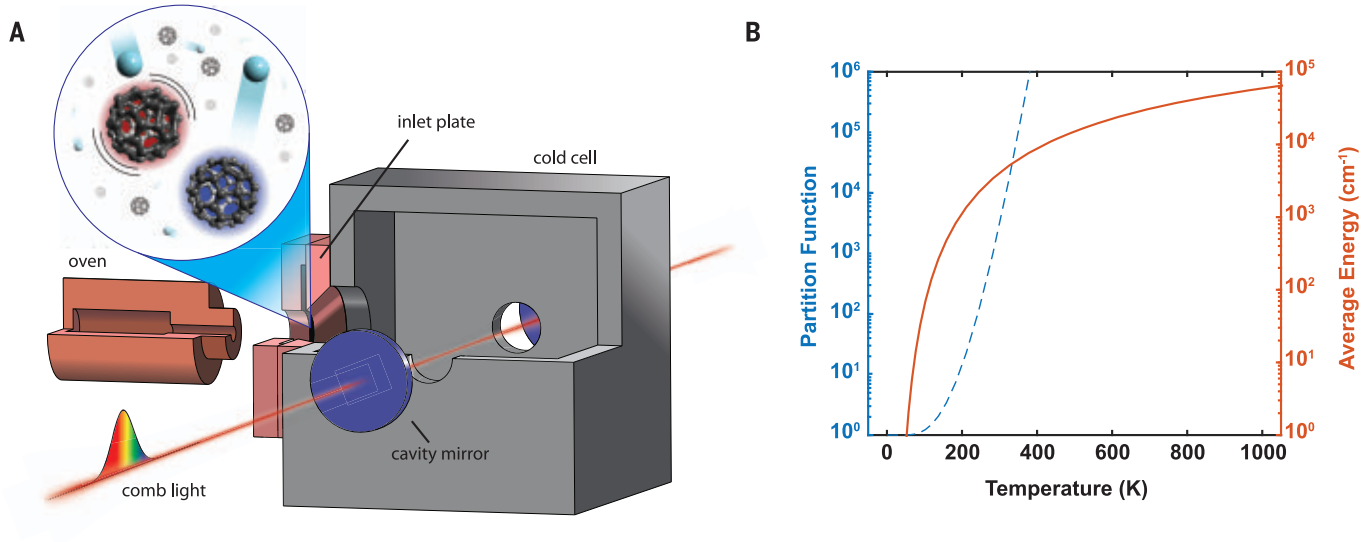


Fig. 1. Cooling and comb spectroscopy of gas-phase C_{60}. (**A**) Sublimated C_{60} vapor exits the oven source and enters a cryogenic cell, where it thermalizes via collisions with cold buffer gas introduced through an annular slit inlet plate surrounding the entrance aperture (see enlarged area). Mid-IR frequency comb light is coupled to an optical enhancement cavity surrounding the cell. The optical absorption spectrum is measured with a scanning arm Fourier transform spectrometer (not pictured). (**B**) The vibrational partition function (blue dashed line) and average vibrational energy (red solid line) increase strongly as a function of temperature. About 6 to 8 eV of vibrational energy must be removed per molecule to cool C_{60} from the initial oven temperature to below 150 K, at which point the vibrational partition function is approximately equal to unity.

molecule. The precise measurement of transition energies between individual molecular eigenstates yields detailed information about strong, multibody interactions between atoms in a unimolecular polyatomic lattice, thus providing profound insights into complex molecular structure and ensuing interaction dynamics.

Here we report a rotationally resolved spectrum of the 8.5-μm vibrational band of buckminsterfullerene, C_{60}. Following the discovery of C_{60} by Kroto *et al.* in 1985 (*1*), infrared (IR) and ^{13}C nuclear magnetic resonance spectroscopy confirmed its caged, icosahedral structure (*2*–*7*). Subsequent spectroscopic and analytical techniques, including x-ray and electron diffraction (*8*, *9*), optical Raman and neutron scattering (*10*–*15*), matrix isolation IR spectroscopy [see (*16*–*18*) and references therein], and photoelectron spectroscopy (*19*, *20*), have greatly advanced our understanding of this unique molecule. Spectroscopy has also played a central role in the astronomical detection of C_{60} and its derivatives (*21*, *22*). High-resolution IR absorption measurements may help resolve current uncertainties regarding the physical state of astronomical C_{60} (*18*). However, to date, there have been no reports of rovibrational quantum state–resolved measurements of C_{60} molecules. The experiments reported here thus establish C_{60} as the largest molecule, and the only example of rare icosahedral symmetry, for which a complete internal quantum state–resolved spectrum has been observed.

[1]JILA, National Institute of Standards and Technology and University of Colorado, Department of Physics, University of Colorado, Boulder, CO 80309, USA. [2]IMRA America, Inc., Ann Arbor, MI 48105, USA.
*Corresponding author. Email: bryan.changala@colorado.edu (P.B.C.); ye@jila.colorado.edu (J.Y.)

Although quantum state–resolved rovibrational spectroscopy is routine for small, light molecules, systems as large and heavy as C_{60} are much less amenable to high-resolution characterization owing to several intrinsic and technical challenges. The increase in both the number of vibrational modes and the magnitude of the moment of inertia for every additional atom results in considerably more rotation-vibration states populated at a given internal temperature. Rovibrational states excited by an IR photon may be strongly coupled to a highly congested manifold of background dark states, the density of which grows rapidly with increasing internal energy, leading to intramolecular vibrational redistribution (IVR) (*23*). The Doppler broadening of optical transitions due to finite translational temperature serves only to exacerbate this spectral congestion. Furthermore, the low gas-phase densities achievable for heavy, nonvolatile species require high detection sensitivity.

These various experimental challenges are addressed by cooling the translational and internal temperatures of gas-phase molecular samples and probing them at lower internal energy with longer wavelength light. The method of cryogenic buffer-gas cooling is particularly effective for large, heavy molecules (*24*, *25*). We have recently demonstrated the integration of a buffer-gas cooling source with cavity-enhanced direct frequency comb spectroscopy (CE-DFCS) in the mid-IR (*26*, *27*), which enables sensitive, broadband, high-resolution absorption measurements (*28*, *29*). We have since made substantial changes to the buffer-gas cooling conditions to permit the preparation and detection of cold, gas-phase samples of even heavier molecules and have extended the spectral window of this apparatus to the long-wave IR (LWIR) region (*30*). We have targeted the 8.5-μm vibrational band because it is the lowest-energy IR active mode that falls within our accessible wavelength region.

Table 1. Fitted spectroscopic parameters of Eq. 6 for the R branch. The residuals (Fig. 4B) have a small root-mean-square error of 7.4×10^{-5} cm^{-1}, slightly larger than the 1σ line-center measurement uncertainty of 2.5×10^{-5} cm^{-1}.

Parameter	Value (cm^{-1})
$\nu_0 + (2\bar{B} + \Delta B)(1 - 2\zeta)$	1184.86196(3)
$2\bar{B}(1 - \zeta) + \Delta B(2 - \zeta)$	0.0078300(3)
ΔB	$-2.876(6) \times 10^{-7}$

Figure 1A depicts a simplified view of the apparatus used for C_{60} cooling and spectroscopy. A 950 K copper oven sublimates solid C_{60} samples, generating gas-phase molecules with an average internal energy of 6 to 8 eV per molecule populating 10^{26} to 10^{30} vibrational quantum states, as shown in Fig. 1B. These hot molecules flow into a cell anchored to a cryogenic cold finger, where they are thermalized close to the cell-wall temperature via collisions with cold buffer-gas atoms introduced through an annular slit inlet plate surrounding the cell entrance aperture. We interrogate the cold gas-phase molecules with CE-DFCS by coupling a frequency comb into a high-finesse optical cavity surrounding the cold cell, which enhances the absorption signal by a factor on the order of the cavity finesse (F = 6000). The LWIR frequency comb light centered near 8.5 μm is

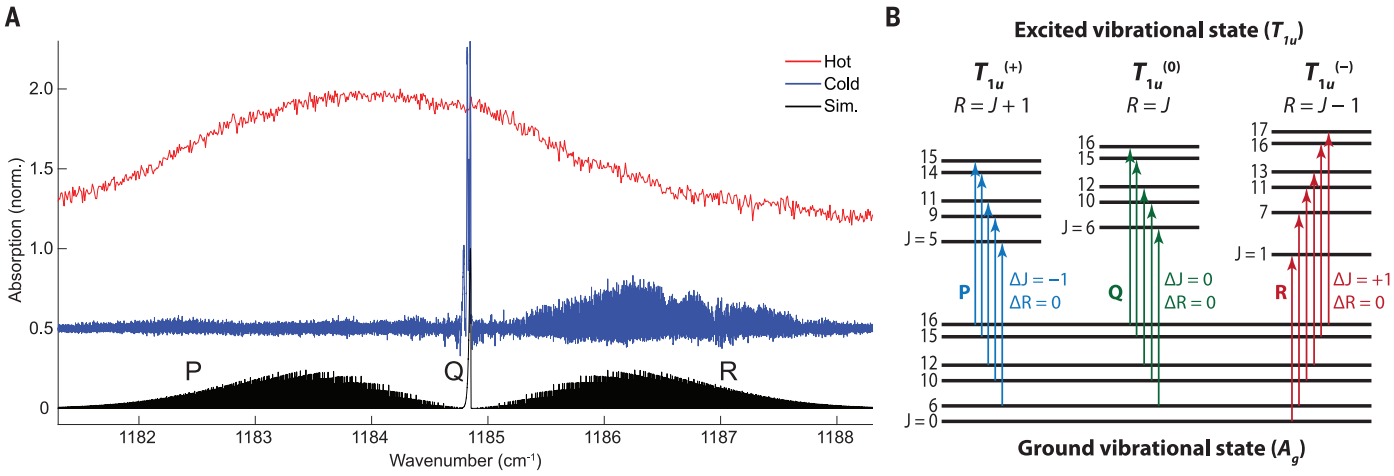


Fig. 2. Spectroscopic patterns of the IR active vibrational band of $^{12}C_{60}$ near 8.5 μm. (**A**) A simulated (sim.) spectrum (black trace) is compared to a measured spectrum of cold (blue trace) and hot (red trace) C_{60}. The measured hot spectrum shows broad, unresolved absorption owing to many thermally occupied vibrational states. The cold spectrum exhibits sharp, well-resolved rotational structure from transitions out of the ground vibrational state. norm., normalized to peak absorption. (**B**) Rovibrational transitions between the ground vibrational state and the excited state follow zeroth-order selection rules of $\Delta J = 0, \pm 1$ and $\Delta R = 0$. These lead to independent P ($\Delta J = -1$), Q ($\Delta J = 0$), and R ($\Delta J = +1$) branches that access the upper-state manifolds labeled $T_{1u}^{(+)}$ (for $R = J + 1$), $T_{1u}^{(0)}$ (for $R = J$), and $T_{1u}^{(-)}$ (for $R = J - 1$), respectively.

produced by difference frequency generation (DFG) with two near-IR frequency combs originating from a single mode-locked fiber laser (*31*). The comb contains narrow teeth at optical frequencies $\nu_m = m \times f_{rep} + f_0$, where m is the integer mode index, f_{rep} is the repetition rate, and the offset frequency f_0 can be introduced via an external acousto-optic modulator before the difference frequency step. The intensity of each comb tooth transmitted through the cavity is read out using a broadband scanning-arm Fourier transform interferometer (*32*, *33*). Additional experimental details are provided in the materials and methods (*34*).

Our first attempts at observing cold gas-phase C_{60} with low-pressure helium buffer gas conditions similar to our previous work (*26*, *27*) yielded no detectable absorption. However, when the vacuum chamber was flooded with a high pressure of helium buffer gas, a single broad, unresolved absorption feature appeared, as shown by the red trace in Fig. 2A. We attribute this spectrum to partially cooled C_{60} molecules that remain warm enough to occupy many vibrational quantum states. This is not surprising: As can be seen in Fig. 1B, even at room temperature, the vibrational partition function is greater than 10^3. This finding suggested that both a higher number of collisions and more-efficient energy transfer per collision would be required to thermalize C_{60} to its ground vibrational state (*35*). Indeed, we ultimately produced a sufficiently dense, cold C_{60} sample by (i) increasing the buffer-gas mass by switching from helium to argon and (ii) carefully optimizing the buffer-gas flow and oven positioning relative to the inlet slit. The spectrum acquired at these conditions is shown by the blue trace in Fig. 2A and exhibits well-resolved rovibrational fine structure, with narrow linewidths on the order of 20 MHz (fig. S1). The peak absorption, near the band origin, is 10% of the cavity-transmitted comb mode intensity. From the magnitude of the integrated absorption cross section (*17*), we estimate the number density of cold C_{60} to be 4×10^{11} cm^{-3}. Observing the appearance and evolution between the broad and narrow signals was greatly facilitated by the wide spectral bandwidth of the frequency comb, which covers the entire breadth of the observed vibrational band. The inferred rotational temperature is about 150 K (*34*), nearly equal to the cell-wall temperature of 135 K, which is kept well above argon's condensation point of 87 K.

The observed fine structure in the infrared spectrum encodes fundamental details of the quantum mechanical structure of C_{60}. To the zeroth order, the rotations of C_{60} can be considered as those of a spherical top with total angular momentum operator $\mathbf{J}$ (*36*). The associated rotational quantum states are $|J, k, m\rangle$, where $J = 0, 1, 2, \ldots$ is the total angular momentum quantum number and $k, m = -J, \ldots, +J$ are the projection quantum numbers of the body-fixed component (J_z) and lab-fixed component (J_Z) of $\mathbf{J}$, respectively. The triply degenerate vibrational mode of T_{1u} symmetry that gives rise to the infrared band can be modeled as a three-dimensional isotropic harmonic oscillator with vibrational angular momentum operator $\boldsymbol{\ell}$. Its quantum states are $|n, \ell, k_\ell\rangle$, where n is the total number of vibrational quanta; $\ell = n, n-2, n-4, \ldots$ is the vibrational angular momentum quantum number; and $k_\ell = -\ell, \ldots, +\ell$ is the projection quantum number of the body-frame projection (ℓ_z) of $\boldsymbol{\ell}$.

The uncoupled rovibrational product wavefunctions $|J, k, m\rangle|n, \ell, k_\ell\rangle$ are simultaneously eigenfunctions of $\mathbf{J}^2, \boldsymbol{\ell}^2, J_z, \ell_z$, and J_Z. It is useful to define the "pure rotational" angular momentum $\mathbf{R} = \mathbf{J} - \boldsymbol{\ell}$, the eigenfunctions of which can be constructed by transforming the uncoupled product wavefunctions using standard angular momentum coupling relations (*36*). This yields total coupled rovibrational wavefunctions of the form $|R, k_R, J, \ell, n, m\rangle$, where R is the angular momentum quantum number of $\mathbf{R}$ and $k_R = -R, \ldots, +R$ is the body-fixed projection. As usual, the values of R satisfy the triangle inequality $R = |J - \ell|, \ldots, J + \ell$. In this work, we are concerned only with the ground vibrational state with $n = \ell = 0$ and the excited T_{1u} vibrational state, populated by the IR photon, with $n = \ell = 1$. Therefore, in the ground vibrational state, $R = J$; similarly, in the excited state where $\ell = 1$, R is restricted to $J, |J \pm 1|$.

The energies of the states we observe are determined by effective rotational Hamiltonians for each vibrational state. The simplest possible effective Hamiltonian for the ground vibrational state is that of a rigid spherical top

$$H_{gr} = B''\mathbf{J}^2 \qquad (1)$$

where B'' is the ground state rotational constant, which is inversely proportional to the moment of inertia. The ground state wavefunctions $|R = J, k_R, J, \ell = 0, n = 0, m\rangle$ are eigenstates of H_{gr} with energies

$$E_{gr} = B''J(J+1) \qquad (2)$$

This energy is independent of k_R and m, leading to the usual $(2R+1)(2J+1) = (2J+1)^2$ spherical-top ground-state degeneracy factor.

The excited vibrational state is described to lowest order by a slightly more sophisticated effective Hamiltonian,

$$H_{ex} = \nu_0 + B'\mathbf{J}^2 - 2B'\zeta(\mathbf{J}\cdot\boldsymbol{\ell}) \qquad (3)$$

where ν_0 is the vibrational band origin, and B' is the excited state rotational constant, which

differs slightly from B'' owing to changes of the moment of inertia upon vibrational excitation. The new rightmost term arises from Coriolis forces that couple the total angular momentum **J** and the vibrational angular momentum $\boldsymbol{\ell}$, with $\ell = 1$. The ζ constant encodes the strength of this coupling, which is determined by the geometric details of the vibrational normal mode. The excited state wavefunctions $|R, k_R, J, \ell = 1, n = 1, m\rangle$ are eigenstates of H_{ex} with energy levels at

$$E_{ex} = \nu_0 + B'J(J+1) - B'\zeta[J(J+1) + \ell(\ell+1) - R(R+1)] \quad (4)$$

again with a degeneracy of $(2R + 1)(2J + 1)$. As $R = J, |J \pm 1|$, the excited state energies sort into three distinct manifolds (*37*)

$$\begin{aligned} E_{ex}^{(+)} &= E_J + 2B'\zeta J, & R &= J+1 \\ E_{ex}^{(0)} &= E_J - 2B'\zeta, & R &= J \\ E_{ex}^{(-)} &= E_J - 2B'\zeta(J+1), & R &= J-1 \end{aligned} \quad (5)$$

where $E_J = \nu_0 + B'J(J + 1)$ is the pure vibrational and rigid rotor contribution to the energy. Physically, these manifolds correspond to states where **J** and $\boldsymbol{\ell}$ are mutually antiparallel, perpendicular, and parallel, respectively.

Rovibrational transitions between the ground and excited T_{1u} vibrational states of spherical tops such as C_{60} are governed by the usual strict $\Delta J = 0, \pm 1$ rule and an additional $\Delta R = \Delta k_R = 0$ rule (*36*). These allowed transitions are illustrated in the level diagram of Fig. 2B. Unlike less symmetric molecules, these selection rules dictate that the usual P ($\Delta J = -1$), Q ($\Delta J = 0$), and R ($\Delta J = +1$) transitions reach mutually exclusive sets of upper state quantum levels. These three manifolds are labeled $T_{1u}^{(+)}$, $T_{1u}^{(0)}$, and $T_{1u}^{(-)}$, according to the energy expressions in Eq. 5.

Inspection of the level diagram in Fig. 2B shows that states with certain values of R are missing. This is, in fact, an exceptional example of nuclear spin statistics at work. The carbon nuclei in pure $^{12}C_{60}$ are each identical spin-0 bosons, so any permutation of nuclei must leave the total molecular wavefunction unchanged. This imposes the strict condition that only states with a total rovibronic symmetry of A_g (+ parity) or A_u (– parity) in the I_h point group may exist. Group theoretical analysis (*38*) of the rovibrational wavefunctions shows that this condition is met only with certain linear combinations of k_R states for a given value of R. In fact, only a single such linear combination is possible for R = 0, 6, 10, 12, 15, 16, 18, 20 to 22, and 24 to 28, with other values of $R < 30$ having no allowed states. (For levels with $R \geq 30$, the number of allowed states is equal to 1 plus the number of states for R minus 30.) The unusual patterns of allowed angular momentum quantum numbers are intimately related to the two-, three- and fivefold symmetry axes of an icosahedron. In the high-R limit, only 1 in 60 states exist, owing to the drastic effects of these ^{12}C nuclear spin statistics.

Taking the zeroth-order energies, selection rules, and spin statistics all together leaves the predicted spectrum plotted in black in Fig. 2A. It consists of a sharp Q branch surrounded by P and R branches containing lines evenly spaced by approximately $(B'' + B')(1 - \zeta) \approx 0.0078$ cm^{-1}. The qualitative appearance of the measured R and Q branch regions is consistent with the simulation, whereas there is substantial disagreement in the P branch. The portions of the spectrum shown in Fig. 3 provide a closer view of this behavior.

The R branch exhibits a regularly spaced progression of transitions R(J) that we have assigned from $J \approx 60$ to 360. Transitions outside this range are below our detection sensitivity. Such high values of the total angular momentum quantum number have been rarely observed, if ever, by rotationally resolved frequency domain spectroscopy. Portions of the measured and simulated R branch from J = 160 to 200 are shown in Fig. 3A. Despite the noise in the measured absorption, these transitions clearly show the expected discrete intensity variations in the correct integer ratios. Such patterns are a basic consequence of the quantum mechanical indistinguishability and the perfect icosahedral arrangement of the carbon nuclei that make up $^{12}C_{60}$.

Quantitative analysis of the R branch transition frequencies permits extraction of spectroscopic constants. The energy expressions in Eqs. 2 and 5 yield expected transition frequencies of

$$\nu[\mathrm{R}(J)] = \nu_0 + (2\bar{B} + \Delta B)(1 - 2\zeta) + J[2\bar{B}(1 - \zeta) + \Delta B(2 - \zeta)] + J^2\Delta B \quad (6)$$

where $\bar{B} \equiv (B' + B'')/2$ is the mean value of the lower- and upper-state rotational constants and $\Delta B \equiv B' - B'' \ll \bar{B}$ is their difference. Figure 4A shows the measured positions (*34*) as a function

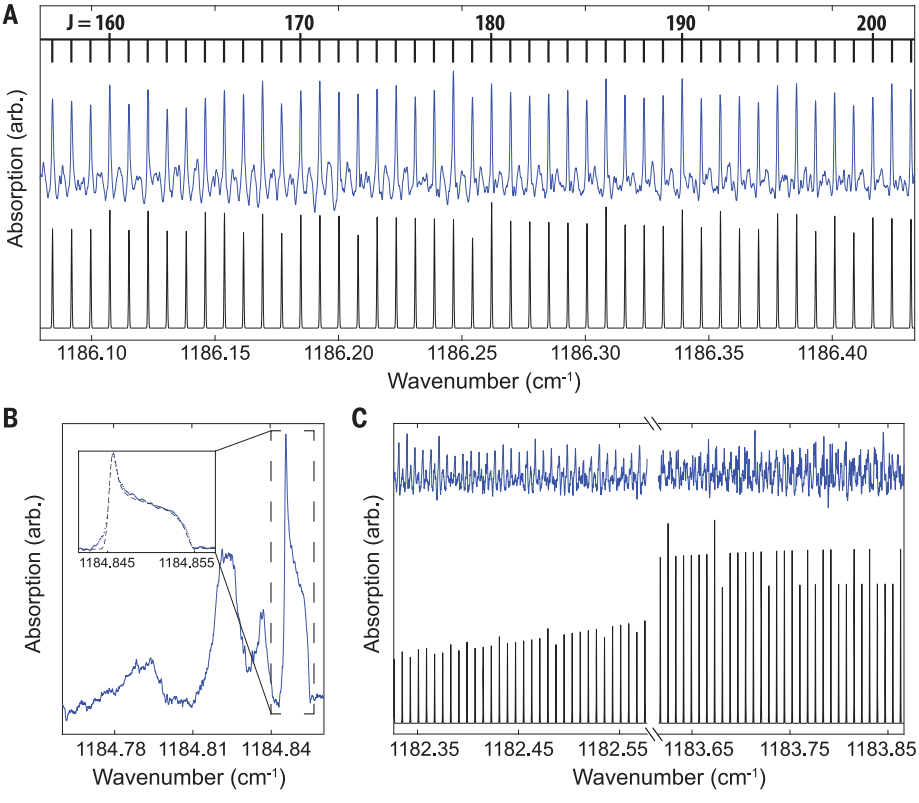


Fig. 3. Detailed views of portions of the measured IR band. (**A**) The R branch shows agreement between the expected intensity patterns from the simulation (black trace) and the measured spectrum (blue trace). The tie line above the spectrum indicates the lower state J value of each observed R(J) transition. (**B**) The Q branch region of the spectrum contains several features. The highest wavenumber feature is assigned as the Q branch of the $^{12}C_{60}$ isotopologue. In the inset, the dashed line represents a fit to a simple quartic centrifugal distortion contour. The additional features at lower frequencies are likely due to the singly substituted $^{13}C^{12}C_{59}$ isotopologue. (**C**) These two portions of the P branch (blue trace) are representative of the disagreement with the zeroth-order simulation determined from parameters fitted to the R branch (black trace). The structure not captured by the simulation is evidence of nonscalar centrifugal distortion effects. arb., arbitrary units.

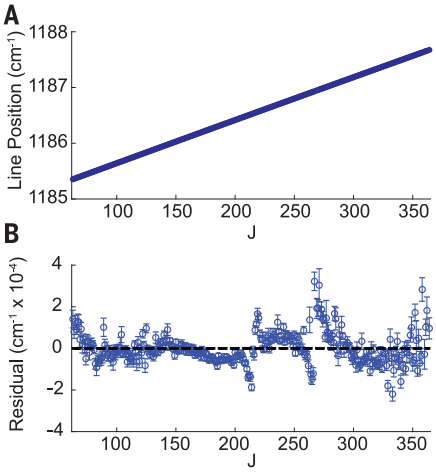


Fig. 4. Fit results for the R branch. (**A**) The R(*J*) line positions plotted versus lower-state *J* display a very linear trend over *J* = 60 to 360. The individual line positions are listed in (*39*). (**B**) The residuals from the fit of Eq. 6 to these line positions, summarized in Table 1, exhibit apparent avoided crossings near *J* = 215 and 275, which are possible signatures of local dark-state perturbers in the upper state. The error bars are 1σ line-center uncertainties determined from lineshape fit residuals (*34*).

of lower state *J*, which follow the expected nearly linear dependence. Figure 4B shows the residuals from a fit of Eq. 6 to the measured line positions, displaying two avoided crossings arising from perturbations in the excited state. The fitted spectroscopic parameters are summarized in Table 1. The R branch transition frequencies are well reproduced despite the simplicity of the zeroth-order Hamiltonian, which ignores centrifugal distortion effects, and the very high range of *J*. A complete listing of the ~300 transition frequencies used in this fit is given in data S1 (*39*).

The quantum state–resolved spectrum of C_{60} provides structural information of isolated gas-phase molecules through the rotational fine structure. Although the transitions included in our initial analysis do not yet allow an independent determination of B'' and ζ, if we assume a range of ζ = −0.30 to −0.45 based on theoretical calculations (*37*), we can estimate $B'' = \frac{1}{hc}\frac{\hbar^2}{2I} \approx 0.0027$ to 0.0030 cm^{-1}, where *I* is the effective moment of inertia of the ground vibrational state, $\hbar$ is Planck's constant *h* divided by 2π, and *c* is the speed of light. Given $I = \frac{2}{3}mr^2$ for a spherical shell of mass *m* and radius *r*, the corresponding range of radii is 3.4 to 3.6 Å. This is consistent with a previous gas-phase electron diffraction measurement of 3.557(5) Å, which includes thermal averaging effects that lengthen the measured radius relative to that of the vibrational ground state (*8*). Further analysis of the rotational fine structure of $^{12}C_{60}$ (and ultimately $^{12}C_{59}{}^{13}C$) will be necessary to constrain B'' and ζ independently and completely determine the gas-phase structural parameters. Our measured value of Δ*B* implies that the effective C_{60} radius increases by only 0.005% upon excitation of the observed vibrational mode, which is primarily of a surface-tangent C–C bond stretching character. The narrow IR transition linewidths (about 20 MHz) to the excited vibrational state provide a lower bound for its IVR lifetime of at least 8 ns, despite being embedded in a dense manifold of dark vibrational states. This is consistent with our expectation that the high degree of icosahedral symmetry substantially restricts rovibrational coupling.

The Q branch region is shown in Fig. 3B. There are several unresolved features here, though each is still quite narrow on an absolute scale of 0.01 to 0.03 cm^{-1}. The highest frequency feature is assigned as the Q branch of the $^{12}C_{60}$ isotopologue in its ground vibrational state. Centrifugal distortion effects create a band head observed near *J* = 250 (inset of Fig. 3B). The remaining features in the Q branch region are not definitively assigned. Although they are possibly hot-band transitions of the $^{12}C_{60}$ isotopologue, we believe they most likely derive from the singly substituted $^{12}C_{59}{}^{13}C$ isotopologue. Despite a ^{13}C natural abundance of only 1.1%, the 60 equivalent substitution sites lead to a notably high ($^{12}C_{59}{}^{13}C$):$^{12}C_{60}$ ratio of about 2:3. The substitution breaks the icosahedral symmetry of C_{60}, splitting the threefold degeneracy of the vibrational level and nullifying the nuclear spin statistics. Many more rotational levels and transitions are expected, which will be further split by the nonspherical moments of inertia (*40*).

Finally, two representative portions of the P branch are shown in Fig. 3C. Here, the zeroth-order simulation fails to capture either the position or number of observed transitions. This complicated fine structure is likely due to high-order centrifugal distortion terms not included in the simulated spectrum (*41*). The zeroth-order Hamiltonians, Eqs. 1 and 3, contain only scalar terms that preserve the spherical degeneracy of the (2*R* + 1) body-fixed projections of **R**. Whereas most of these substates are eliminated by the ^{12}C nuclear spin statistics, the degeneracy of the remaining substates can be broken by nonscalar centrifugal distortion terms. These so-called "icosahedral splitting" terms (*41*) lead to subsequent splittings of the observed transitions. In the ground state, the lowest-order nonscalar centrifugal distortion term scales as J^6, whereas such terms can appear in the excited state that scale only as J^4. Owing to the large *J* values observed here, it is not surprising that such effects become important. However, to date, there have been no theoretical predictions of the magnitude of these icosahedral splitting terms. A full analysis of this portion of the spectrum is most effectively treated within the irreducible spherical tensor formalism (*42*).

The present experiments point toward an exciting future direction of fullerene research. The general applicability of buffer-gas cooling establishes the possibility of similar studies, using vibrational, electronic, or other spectroscopies, on larger fullerenes such as C_{70}; endofullerenes, wherein an atom or small molecule is encapsulated in a closed fullerene cage; or even pure $^{13}C_{60}$, which represents a pristine example of a spin-½ network on a spherical lattice. Ultimately, precision spectroscopy of such targets is the first step toward single quantum state preparation and control of large molecular systems.

REFERENCES AND NOTES

1. H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl, R. E. Smalley, *Nature* **318**, 162–163 (1985).
2. H. Ajie *et al.*, *J. Phys. Chem.* **94**, 8630–8633 (1990).
3. R. Taylor, J. P. Hare, A. K. Abdulsada, H. W. Kroto, *J. Chem. Soc. Chem. Commun.* **1990**, 1423–1424 (1990).
4. W. Krätschmer, K. Fostiropoulos, D. R. Huffman, *Chem. Phys. Lett.* **170**, 167–170 (1990).
5. W. Krätschmer, L. D. Lamb, K. Fostiropoulos, D. R. Huffman, *Nature* **347**, 354–358 (1990).
6. C. S. Yannoni, P. P. Bernier, D. S. Bethune, G. Meijer, J. R. Salem, *J. Am. Chem. Soc.* **113**, 3190–3192 (1991).
7. C. S. Yannoni, R. D. Johnson, G. Meijer, D. S. Bethune, J. R. Salem, *J. Phys. Chem.* **95**, 9–10 (1991).
8. K. Hedberg *et al.*, *Science* **254**, 410–412 (1991).
9. S. Liu, Y.-J. Lu, M. M. Kappes, J. A. Ibers, *Science* **254**, 408–410 (1991).
10. S. F. Parker *et al.*, *Phys. Chem. Chem. Phys.* **13**, 7789–7804 (2011).
11. L. Pintschovius, *Rep. Prog. Phys.* **59**, 473–510 (1996).
12. D. S. Bethune, G. Meijer, W. C. Tang, H. J. Rosen, *Chem. Phys. Lett.* **174**, 219–222 (1990).
13. D. S. Bethune *et al.*, *Chem. Phys. Lett.* **179**, 181–186 (1991).
14. K. Prassides *et al.*, *Chem. Phys. Lett.* **187**, 455–458 (1991).
15. R. L. Cappelletti *et al.*, *Phys. Rev. Lett.* **66**, 3261–3264 (1991).
16. N. Sogoshi *et al.*, *J. Phys. Chem. A* **104**, 3733–3742 (2000).
17. S. Iglesias-Groth, F. Cataldo, A. Manchado, *Mon. Not. R. Astron. Soc.* **413**, 213–222 (2011).
18. A. C. Brieva, R. Gredel, C. Jäger, F. Huisken, T. Henning, *Astrophys. J.* **826**, 122 (2016).
19. X.-B. Wang, C.-F. Ding, L.-S. Wang, *J. Chem. Phys.* **110**, 8217–8220 (1999).
20. D.-L. Huang, P. D. Dau, H.-T. Liu, L.-S. Wang, *J. Chem. Phys.* **140**, 224315 (2014).
21. J. Cami, J. Bernard-Salas, E. Peeters, S. E. Malek, *Science* **329**, 1180–1182 (2010).
22. E. K. Campbell, M. Holz, D. Gerlich, J. P. Maier, *Nature* **523**, 322–323 (2015).
23. D. J. Nesbitt, R. W. Field, *J. Phys. Chem.* **100**, 12735–12756 (1996).
24. D. Patterson, E. Tsikata, J. M. Doyle, *Phys. Chem. Chem. Phys.* **12**, 9736–9741 (2010).
25. J. Piskorski, D. Patterson, S. Eibenberger, J. M. Doyle, *ChemPhysChem* **15**, 3800–3804 (2014).
26. P. B. Changala, B. Spaun, D. Patterson, J. M. Doyle, J. Ye, *Appl. Phys. B* **122**, 292 (2016).
27. B. Spaun *et al.*, *Nature* **533**, 517–520 (2016).
28. F. Adler, M. J. Thorpe, K. C. Cossel, J. Ye, *Annu. Rev. Anal. Chem. (Palo Alto, Calif.)* **3**, 175–205 (2010).
29. M. J. Thorpe, J. Ye, *Appl. Phys. B* **91**, 397–414 (2008).
30. K. Iwakuni *et al.*, *Appl. Phys. B* **124**, 128 (2018).
31. K. F. Lee, C. J. Hensley, P. G. Schunemann, M. E. Fermann, *Opt. Express* **25**, 17411–17416 (2017).
32. J. Mandon, G. Guelachvili, N. Picqué, *Nat. Photonics* **3**, 99–102 (2009).
33. F. Adler *et al.*, *Opt. Express* **18**, 21861–21872 (2010).
34. See supplementary materials.
35. J. T. Stewart, B. E. Brumfield, B. M. Gibson, B. J. McCall, *ISRN Phys. Chem.* **2013**, 675138 (2013).
36. C. di Lauro, in *Rotational Structure in Molecular Infrared Spectra* (Elsevier, 2013), chap. 10, pp. 225–245.

37. D. E. Weeks, W. G. Harter, *Chem. Phys. Lett.* **176**, 209–216 (1991).
38. P. R. Bunker, P. Jensen, *Mol. Phys.* **97**, 255–264 (1999).
39. P. B. Changala, M. L. Weichman, K. F. Lee, M. E. Fermann, J. Ye, External data for "Rovibrational quantum state resolution of the C_{60} fullerene," Version 1, Harvard Dataverse (2018); https://doi.org/10.7910/DVN/RVVAN6.
40. T. C. Reimer, W. G. Harter, *J. Chem. Phys.* **106**, 1326–1335 (1997).
41. W. G. Harter, D. E. Weeks, *J. Chem. Phys.* **90**, 4727–4743 (1989).
42. V. Boudon *et al.*, *J. Mol. Spectrosc.* **228**, 620–634 (2004).

ACKNOWLEDGMENTS
The authors thank H. Green for technical advice during the design of the oven source and J. Doyle for insightful discussions. **Funding:** This work was supported by AFOSR grant no. FA9550-15-1-0111, the Gordon and Betty Moore Foundation, the DARPA SCOUT Program, NIST, and NSF PHYS-1734006. M.L.W. is supported through an NRC Postdoctoral Fellowship. **Author contributions:** P.B.C., M.L.W., and J.Y. performed the experiment. K.F.L. and M.E.F. built the DFG-based comb. All authors contributed to the writing of the paper. **Competing interests:** K.F.L. and M.E.F. have submitted a patent (WO2017209989A1) on portions of the DFG frequency comb instrument. **Data and materials availability:** All data are available in the main text, in the supplementary materials, or through Harvard Dataverse (*39*).

SUPPLEMENTARY MATERIALS
www.sciencemag.org/content/363/6422/49/suppl/DC1
Materials and Methods
Fig. S1
References (*43–45*)

29 August 2018; accepted 8 November 2018
10.1126/science.aav2616

MESOSCOPIC PHYSICS

Counter-propagating charge transport in the quantum Hall effect regime

Fabien Lafont[1,2]*†, Amir Rosenblatt[1]*, Moty Heiblum[1], Vladimir Umansky[1]

The quantum Hall effect, observed in a two-dimensional (2D) electron gas subjected to a perpendicular magnetic field, imposes a 1D-like chiral, downstream, transport of charge carriers along the sample edges. Although this picture remains valid for electrons and Laughlin's fractional quasiparticles, it no longer holds for quasiparticles in the so-called hole-conjugate states. These states are expected, when disorder and interactions are weak, to harbor upstream charge modes. However, so far, charge currents were observed to flow exclusively downstream in the quantum Hall regime. Studying the canonical spin-polarized and spin-unpolarized $\nu = 2/3$ hole-like states in GaAs-AlGaAs heterostructures, we observed a significant upstream charge current at short propagation distances in the spin unpolarized state.

Elementary charge excitations in the quantum Hall effect (QHE) flow downstream along the edge of a two-dimensional electron gas (2DEG), with the downstream chirality imposed by the magnetic field (*1*). In the fractional regime (*2*), this statement remains valid only for particle-like (Laughlin's) states (*3–5*); by contrast, hole-like states (filling factors ν so that $1/2 + n < \nu < 1 + n$ with $n = 0, 1, 2...$) are expected to harbor counter-propagating (downstream and upstream) charge excitations (*6*). In a noninteracting and scattering-free model, a downstream $\nu = 1$ charge mode was predicted to be accompanied by an upstream $\nu = 1/3$ mode, leading to a two-terminal conductance of $4e^2/3h$, where e and h are the electron charge and the Planck constant, respectively. However, experimentally, only downstream charge modes (*7*, *8*) with a two-terminal conductance of $2e^2/3h$ accompanied by upstream neutral modes (*9–15*) have been found. A recent experiment (*16*) measured conductance of an unequilibrated downstream channels at narrow regions (4 μm wide) of the polarized $\nu = 2/3$ state; the results were consistent with the model from (*6*), but no direct measurement of the upstream current was made. Although the majority of the studies were concentrated on the spin-polarized $\nu = 2/3$ state, there has been recent interest in its spin-unpolarized counterpart (*17–24*) as a potential host for para-fermions when coupled to superconducting contacts (*25–27*). In the composite fermion (CF) picture, one can construct two kinds of states in the $\nu = 2/3$: an unpolarized state, emerging at lower magnetic fields, with two quantum levels that have the same orbital quantum number but opposite spin configurations: (0, ↑) and (0, ↓) (Fig. 1B) (*28*), and a polarized state, emerging at high magnetic fields, with two quantum levels having the same spin but different orbitals (0, ↑) and (1, ↑) (*29*). The majority of previous experiments in the unpolarized state focused on characterizing the spin domains structure in the bulk (*23*, *24*, *30*) or the nuclear spin polarization occurring at high currents (*18*, *19*, *21*, *22*, *30–35*). Still, the configuration of edge channels for this state remains elusive: On the one hand, no upstream channel is expected in the CF picture; on the other, because the effective K-Matrix in the CF basis is the same for both $\nu = 2/3$ states, an upstream mode should occur also in the unpolarized case (*36*). We studied the two flavors of the $\nu = 2/3$ state along a short distance (a few micrometers) and found a substantial upstream charge current only in the spin-unpolarized state. Consequently, the two-terminal

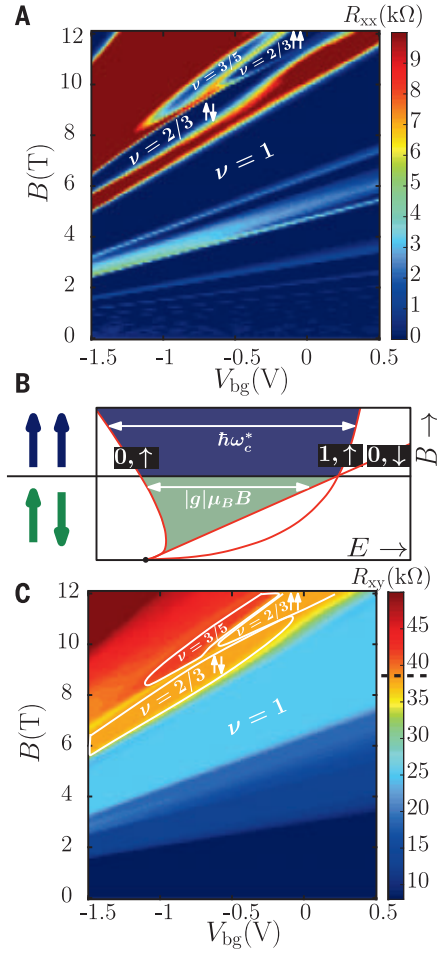


Fig. 1. Longitudinal and transverse magnetoresistances measured in a 40-μm-wide Hall bar sample. (**A**) Longitudinal four-terminal magnetoresistance versus backgate voltage measured by using $I = 1$ nA at $T = 40$ mK. A clear nondissipated state, $R_{xx} \approx 0$, is visible for the $\nu = 2/3$ polarized and unpolarized states. (**B**) Sketch of the evolution of the relevant energy scales. At low field, a gap exists between the (0, ↑) and (0, ↓) states,corresponding to the spin unpolarized state, whereas at higher fields, thanks to the different B dependency of the Coulomb ($\propto l_B^{-1} \sim \sqrt{B}$, where l_B is the magnetic length) and Zeeman ($\propto B$) energies, the gap exists between the (0, ↑) and the (1, ↑) λ levels corresponding to the polarized state. (**C**) Four-terminal transverse magnetoresistance as function of the backgate voltage. The $\nu = 2/3$ polarized and unpolarized quantum Hall plateaus exhibit a resistance $R_{xy} \approx (2e^2/3h)^{-1} \approx 38.7$ kilohms (dashed line on the color bar).

[1]Braun Center for Submicron Research, Department of Condensed Matter Physics, Weizmann Institute of Science, Rehovot 76100, Israel. [2]College de France, 11 place Marcelin Berthelot, 75231 Paris Cedex 05, France.
*These authors contributed equally to this work.
†Corresponding author. Email: lafont.fabien@gmail.com

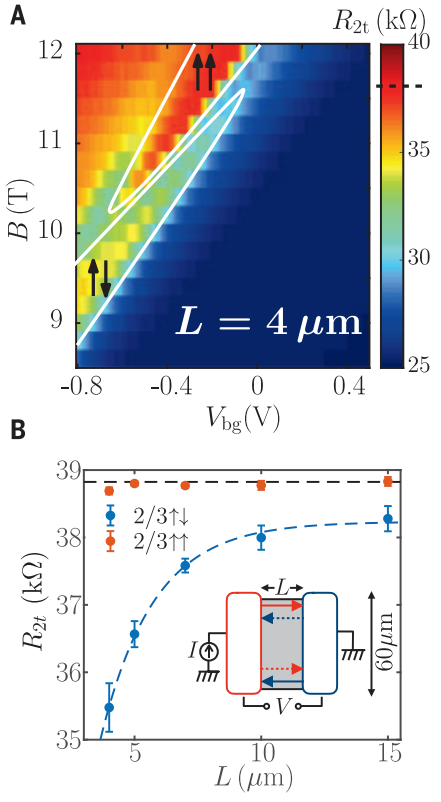


Fig. 2. Deviation from the quantized Hall resistance value owing to the upstream current. (**A**) Two-terminal magnetoresistance versus backgate voltage for a L = 4-μm-long and 60-μm-wide sample measured at T ~ 35 mK and I = 1 nA. A clear difference appears between the spin-polarized state, which remains quantized, and the spin-unpolarized state, which deviates substantially from the quantized value. (**B**) Evolution of the two-terminal resistance averaged over an area of (B, V_{bg}) corresponding to the polarized and unpolarized ν = 2/3 states, as a function of the length L [(*38*), section 2]. The dashed black line is the quantized value $(2e^2/3h)^{-1}$, and the dashed blue line is an exponential fit $R(x) = [R(0) - R(\infty)]e^{-x/l_0} + R(\infty)$, where $R(0)$ = 20 ± 13 kilohms, $R(\infty)$ = 38.2 ± 0.3 kilohms, and l_0 = 2.1 ± 0.8 μm.

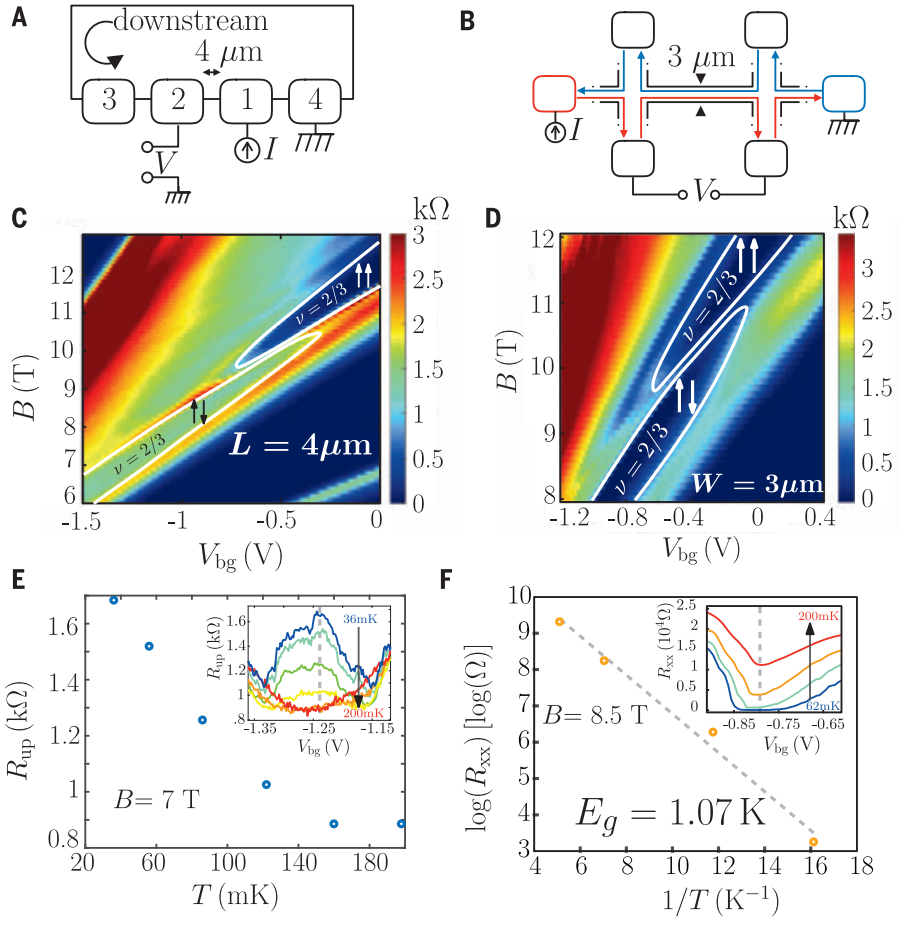


Fig. 3. Different sample geometries validating the presence of a counter-propagating charge flow. (**A**) Sketch of the three-terminal measurement. Four contacts are aligned on a single edge of the sample, 4 μm apart. Current I = 1 nA was sourced at contact 1, and voltage was measured between the contact 2 placed upstream and the ground; contact 4 was grounded, and contact 3 was floating. (**B**) Sketch of the four-terminal R_{xx} measurement on a narrow, 3-μm-wide and 25-μm-long Hall bar. The red lines represent the biased-edge channels, whereas the blue ones represent the grounded-edge channels. (**C**) Three-terminal magnetoresistance versus the backgate voltage for the measurement scheme presented in (A). A clear finite resistance appears in the unpolarized region. (**D**) Four-terminal R_{xx} versus backgate voltage for the measurement scheme presented in (B). The R_{xx} values are low in both polarized and unpolarized regions, in contrast to (C). (**E**) Evolution of the resistance measured in the unpolarized regime at B = 7 T and V_{bg} = –1.24 V as function of temperature. (Inset) Resistance in the unpolarized regime as function of the backgate voltage for different temperatures (36, 56, 86, 122, 160, and 198 mK). (**F**) Evolution of the log of the longitudinal resistance versus the inverse of the temperature in the Hall bar geometry. (Inset) Evolution of the resistance versus the backgate voltage in the unpolarized regime for several temperatures (62, 85, 142, and 196 mK). The extracted activation gap is E_g = 1.07 K.

resistance deviates from the quantized one at ν = 2/3. The GaAs-AlGaAs heterostructure used to study the two ν = 2/3 states had to be carefully designed (with the 2DEG confined in a narrow, 12-nm-wide quantum-well) because we aimed to have the transition between the two states at a sufficiently high carrier density (and magnetic field), corresponding to having high mobility throughout the transition region in the phase space between the two states. A conductive n^+ GaAs layer was grown ~1 μm below the 2DEG and served as a backgate, capable of tuning the density from 1×10^{11} to 2.5×10^{11} cm^{-2}, with a corresponding low-temperature dark mobility of 1.5×10^6 to 3.5×10^6 cm^2 V^{-1} s^{-1}. Lock-in measurements were performed at ~80 Hz with an input current I = 1 nA and an electron temperature of ~35 mK [additional fabrication information is provided in (*37*), section 1].

The evolution of the four-terminal longitudinal (R_{xx}) and transverse resistance (R_{xy}), measured in a 40-μm-wide Hall-bar geometry, is plotted on Fig. 1, A and C. As reported previously (*17*, *21*, *22*, *30*, *31*), a clear transition between the two-spin varieties of the ν = 2/3 states is visible in R_{xx} (around V_{bg} = –0.5 V and B = 10 T) (Fig. 1A). The finite R_{xx} region corresponds to the point at which the system undergoes a first-order quantum phase transition between the spin-unpolarized and the spin-polarized ν = 2/3 state. The transverse resistance $R_{xy} \simeq (2e^2/3h)^{-1} \simeq$ 38.7 kilohms, however, remains constant on both sides of the transition. As predicted in (*6*), the presence of an upstream current leads to the deviation of the two-terminal resistance from the canonical value $R_{2t} \simeq$ 38.7 kilohms. We therefore have conducted two-terminal resistance measurements of several samples, consisting of two 60-μm-wide ohmic contacts separated by a distance L ranging from 4 to 15 μm. The large aspect ratio (of width to length) minimizes backscattering between the propagating edge modes on

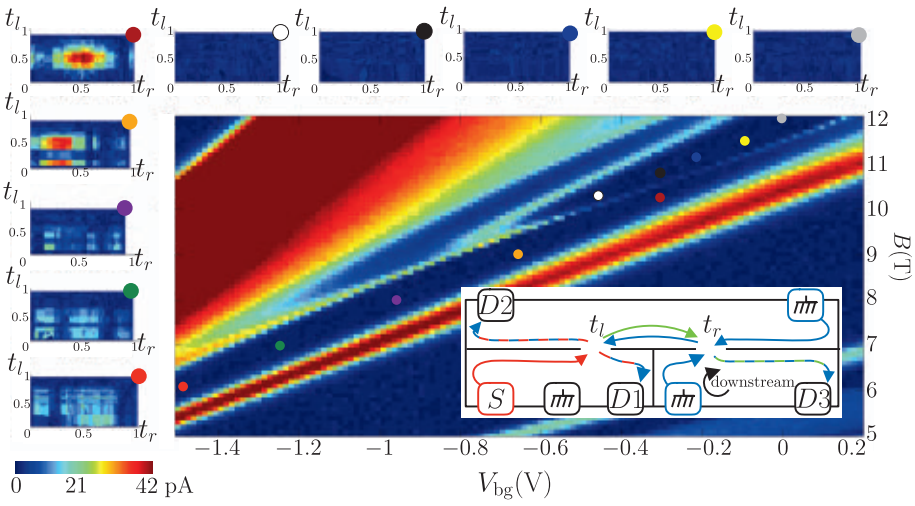


Fig. 4. Generation of upstream charge current by a quantum point contact. Evolution of the current measured in D3 as function of t_l and t_r for different points in the (V_{bg}, B) phase space (points shown in the large image). A significant current is measurable in the unpolarized region (red, green, purple, orange, and brown points), whereas no signal was measurable in the polarized region (white, black, blue, yellow, and gray points). (**Inset**) Two successive quantum point contacts setup. Current is sourced at S (red), flowing downstream to the left QPC; there, it is split to downstream (red/blue) and upstream (green) charge currents. The unequilibrated upstream current reaches the second QPC and turns back to downstream (green/blue), where we measured its voltage at D3. The sketch is not to scale; the distance between the two QPC is 700 nm, and the distance between the QPC and the nearest ohmic contact is 30 μm.

opposite sides of the mesa. As visible on Fig. 2A for L = 4 μm, in the (B, V_{bg}) phase space corresponding to the polarized state, we found a two-terminal resistance $R_{2t}^{\uparrow\uparrow}(L=4) \simeq 38.6 \pm 0.1$ kilohms. However, for the unpolarized state, the resistance plateau was found to deviate from the quantized value, showing $R_{2t}^{\uparrow\downarrow}(L=4) \simeq 35.5$ kilohms.

Measuring the evolution of $R_{2t}^{\uparrow\uparrow}$ and $R_{2t}^{\uparrow\downarrow}$ with length, we found $R_{2t}^{\uparrow\uparrow}$ independent of contact separation (Fig. 2B, orange circles), whereas $R_{2t}^{\uparrow\downarrow}$ increases with L, approaching the quantized value for L = 15 μm (Fig. 2B, blue circles). Exponential fit of the two-terminal resistance is presented in Fig. 2B (dashed blue line), $[R(0) - R(\infty)]e^{-x/l_0} + R(\infty)$, where $R(0)$ = 20 ± 13 kilohms is the resistance at zero distance, $R(\infty)$ = 38.2 ± 0.3 kilohms is the resistance at infinite distance, and l_0 = 2.1 ± 0.8 μm is the characteristic equilibration length [additional details on Fig. 2B are provided in (*37*), section 2]. Moreover, the resistance $R(0)$ is in agreement with the two-terminal resistance predicted for unequilibrated channels proposed in (*6*), $R_{2t} = (4e^2/3h)^{-1} \approx 19.4$ kilohms. These observations might have been possible at short distance because of the reduction of scattering events and the screening of the Coulomb interraction by the back gate placed 1 μm away from the 2DEG.

Bearing in mind that a finite R_{xx} caused by dissipation processes at short contact separation can lead to similar observations, a few additional configurations were tested. One of them was a configuration that uses a complementary Hall bar structure, with a narrow Hall channel width of 3 μm (Fig. 3B); this was necessary in order to ensure the lack of backscattering along distances under consideration. The measured R_{xx} (Fig. 3D) for the two ν = 2/3 states by using this geometry was negligibly small, ensuring that edge states located on opposite sides of the Hall bar (Fig. 3B, red and blue lines) do not exchange particles. This observation is in agreement with the relatively large gap (~1 K for the unpolarized state) extracted from the temperature evolution of R_{xx}^{4p} presented on Fig. 3F and is in agreement with previous measurements (*38*, *39*). Furthermore, testing a Corbino geometry sample ensured a negligible bulk conductance of both ν = 2/3 states [(*37*), section 3]. Last, a three-terminal configuration, with contacts aligned on a single edge of the mesa (each separated by 4 μm), allowed separation of the upstream current from the downstream one (Fig. 3A), providing a direct measurement for the upstream conductance. Current I was sourced via contact 1 and drained via contact 4 to the ground. A finite voltage V_{up} was measured at contact 2 for the unpolarized state only as visible on Fig. 3C. The resistance, defined as $R_{up} = V_{up}/I$, continuously dropped with increasing temperature up to 200 mK (Fig. 3E). This dependence has an opposite trend to that of usual dissipative processes such as variable range hopping or activation mechanisms, ruling them out as alternative explanations. A complementary measurement of the downstream resistance, $R_d = V_d/I$, was done by sourcing current via contact 2 and measuring the voltage at contact 1. The upstream and downstream conductances, calculated by using the Landauer-Buttiker formalism (*40*, *41*) [(*38*), section 4], leads to $G_d \approx 0.687\ e^2/h$ and $G_{up} \approx 0.026\ e^2/h$ or equivalent to a two-terminal resistance $(G_{up} + G_d)^{-1} \approx 1.40\ h/e^2 \approx 36.2$ kilohms, which is in agreement with the two-terminal configuration at 4 μm presented above (Fig. 2). The mobility of the 2DEG in proximity to an alloyed ohmic contact is degraded, and its density is increased, which limited us on the minimal distance between ohmic contacts to 4 μm.

In order to probe the edge modes at shorter distances, we used a configuration consisting of two, gate-defined quantum point contacts (QPCs) separated by 700 nm (Fig. 4, inset), with all ohmic contacts placed far away (above 30 μm). A current I = 1 nA was sourced via contact S; currents were monitored at the drains while scanning the transmissions of the left and right QPCs t_l and t_r. This was done at different points in the (V_{bg}, B) phase space for both spin polarization of the ν = 2/3 states, indicated by the colored circles in Fig. 4. In the polarized state, all of the current flowed to drains D1 and D2 independent of t_r, which is consistent with downstream channels, and zero current was measured at D3 (Fig. 4, white, black, blue, yellow, and gray points). However, in the unpolarized state (Fig. 4, red, green, purple, orange, and brown points), substantial signal was found in D3, simultaneously decreasing the current measured in D1 and D2 result in overall current conservation [(*37*), section 5]. This "upstream effect" can be explained by the appearance of an upstream current between the two QPCs (Fig. 4, inset, green arrow), which emerges from the left QPC, flows a short distance to the right QPC, and scatters back to the downstream channel, finally arriving at D3. A maximum current at D3 was measured when $t_l = t_r = 0.5$ [a toy model for this effect is presented in (*37*), section 6].

The present set of experiments revealed counterpropagation of charged particles in the fractional quantum Hall effect regime. This present experiment may induce future theoretical works of the less understood unpolarized ν = 2/3 state.

REFERENCES AND NOTES

1. K. V. Klitzing, G. Dorda, M. Pepper, *Phys. Rev. Lett.* **45**, 494–497 (1980).
2. D. C. Tsui, H. L. Stormer, A. C. Gossard, *Phys. Rev. Lett.* **48**, 1559–1562 (1982).
3. R. B. Laughlin, *Phys. Rev. Lett.* **50**, 1395–1398 (1983).
4. C. W. J. Beenakker, *Phys. Rev. Lett.* **64**, 216–219 (1990).
5. X. Wen, *Mod. Phys. Lett. B* **05**, 39–46 (1991).
6. A. H. MacDonald, *Phys. Rev. Lett.* **64**, 220–223 (1990).
7. R. C. Ashoori, H. L. Stormer, L. N. Pfeiffer, K. W. Baldwin, K. West, *Phys. Rev. B Condens. Matter* **45**, 3894–3897 (1992).
8. R. Sabo *et al.*, *Nat. Phys.* 10.1038/nphys4010 (2017).
9. C. L. Kane, M. P. A. Fisher, J. Polchinski, *Phys. Rev. Lett.* **72**, 4129–4132 (1994).
10. Y. Meir, *Phys. Rev. Lett.* **72**, 2624–2627 (1994).
11. J. Wang, Y. Meir, Y. Gefen, *Phys. Rev. Lett.* **111**, 246803 (2013).
12. A. Bid *et al.*, *Nature* **466**, 585–590 (2010).
13. I. Gurman, R. Sabo, M. Heiblum, V. Umansky, D. Mahalu, *Nat. Commun.* **3**, 1289 (2012).
14. H. Inoue *et al.*, *Nat. Commun.* **5**, 4067 (2014).
15. A. Rosenblatt *et al.*, *Nat. Commun.* **8**, 2251 (2017).
16. A. Grivnin *et al.*, *Phys. Rev. Lett.* **113**, 266803 (2014).
17. J. P. Eisenstein, H. L. Stormer, L. N. Pfeiffer, K. W. West, *Phys. Rev. B Condens. Matter* **41**, 7910–7913 (1990).
18. S. Kronmüller *et al.*, *Phys. Rev. Lett.* **81**, 2526–2529 (1998).
19. S. Kronmüller *et al.*, *Phys. Rev. Lett.* **82**, 4070–4073 (1999).

20. T. Chakraborty, *Adv. Phys.* **49**, 959–1014 (2000).
21. J. H. Smet, R. A. Deutschmann, W. Wegscheider, G. Abstreiter, K. von Klitzing, *Phys. Rev. Lett.* **86**, 2412–2415 (2001).
22. S. Kraus *et al.*, *Phys. Rev. Lett.* **89**, 266801 (2002).
23. J. Hayakawa, K. Muraki, G. Yusa, *Nat. Nanotechnol.* **8**, 31–35 (2013).
24. J. N. Moore, J. Hayakawa, T. Mano, T. Noda, G. Yusa, *Phys. Rev. Lett.* **118**, 076802 (2017).
25. R. S. K. Mong *et al.*, *Phys. Rev. X* **4**, 011036 (2014).
26. D. J. Clarke, J. Alicea, K. Shtengel, *Nat. Phys.* **10**, 877–882 (2014).
27. T. Wu *et al.*, *Phys. Rev. B* **97**, 245304 (2018).
28. J. K. Jain, *Phys. Rev. Lett.* **63**, 199–202 (1989).
29. I. V. Kukushkin, K. V. Klitzing, K. Eberl, *Phys. Rev. Lett.* **82**, 3665–3668 (1999).
30. B. Verdene *et al.*, *Nat. Phys.* **3**, 392–396 (2007).
31. J. Huels, J. Weis, J. Smet, K. Klitzing, Z. Wasilewski, *Phys. Rev. B* **69**, 085319 (2004).
32. S. Hennel *et al.*, *Phys. Rev. Lett.* **116**, 136804 (2016).
33. H. Cho *et al.*, *Phys. Rev. Lett.* **81**, 2522–2525 (1998).
34. Y. Q. Li, V. Umansky, K. von Klitzing, J. H. Smet, *Phys. Rev. B* **86**, 115421 (2012).
35. J. H. Smet *et al.*, *Nature* **415**, 281–286 (2002).
36. Y.-H. Wu, G. J. Sreejith, J. K. Jain, *Phys. Rev. B* **86**, 115127 (2012).
37. Supplementary text is available as supplementary materials.
38. L. W. Engel, S. W. Hwang, T. Sajoto, D. C. Tsui, M. Shayegan, *Phys. Rev. B Condens. Matter* **45**, 3418–3425 (1992).
39. G. S. Boebinger, A. M. Chang, H. L. Stormer, D. C. Tsui, *Phys. Rev. Lett.* **55**, 1606–1609 (1985).
40. M. Büttiker, *Phys. Rev. Lett.* **57**, 1761–1764 (1986).
41. S. Datta, *Electronic Transport in Mesoscopic Systems* (Cambridge Univ. Press, 1995), vol. 3.
42. A. Rosenblatt, F. Lafont, Replication Data for: Counter-propagating charge transport in the quantum Hall effect regime. Harvard Dataverse (2018);

ACKNOWLEDGMENTS

We thank A. Stern and Y. Meir for fruitful discussions. We thank D. Mahalu for her precious help in the ebeam lithography process. **Funding:** We acknowledge the European Research Council under the European Community's Seventh Framework Program, grant agreement 339070; the partial support of the Minerva Foundation, grant 711752; and, together with V.U., the German-Israeli Foundation (GIF), grant I-1241-303.10/2014, and the Israeli Science Foundation (ISF). **Author contributions:** F.L. and A.R. contributed equally to this work in sample design, device fabrication, measurement set-up, data acquisition, data analysis and interpretation, and writing of the paper. M.H. guided the experimental work and contributed in data interpretation and writing of the paper. V.U. contributed in molecular beam epitaxy growth. **Competing interests:** The authors declare that they have no competing financial interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials and deposited at (*42*).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/363/6422/54/suppl/DC1
Supplementary Text
Figs. S1 to S5

1 November 2017; accepted 8 November 2018
10.1126/science.aar3766

NANOMATERIALS

Fluorine-programmed nanozipping to tailored nanographenes on rutile TiO_2 surfaces

M. Kolmer[1,2]*, R. Zuzak[1], A. K. Steiner[3], L. Zajac[1], M. Engelund[4], S. Godlewski[1], M. Szymonski[1], K. Amsharov[3]*

The rational synthesis of nanographenes and carbon nanoribbons directly on nonmetallic surfaces has been an elusive goal for a long time. We report that activation of the carbon (C)–fluorine (F) bond is a reliable and versatile tool enabling intramolecular aryl-aryl coupling directly on metal oxide surfaces. A challenging multistep transformation enabled by C–F bond activation led to a dominolike coupling that yielded tailored nanographenes directly on the rutile titania surface. Because of efficient regioselective zipping, we obtained the target nanographenes from flexible precursors. Fluorine positions in the precursor structure unambiguously dictated the running of the "zipping program," resulting in the rolling up of oligophenylene chains. The high efficiency of the hydrogen fluoride zipping makes our approach attractive for the rational synthesis of nanographenes and nanoribbons directly on insulating and semiconducting surfaces.

Carbon-based nanostructures synthesized through rational surface-assisted C–C coupling on single-crystal metal surfaces (*1*, *2*) include individual isomers of fullerenes (*3*, *4*) and fullerene fragments (*5*, *6*), the chirality-pure carbon nanotubes (*7*), atomically precise nanographenes (NGs) (*8*–*10*), and graphene nanoribbons (GNRs) with a well-defined periphery (*6*, *11*–*15*). The consecutive, thermally triggered cyclodehydrogenation of the polycyclic aromatic hydrocarbon (PAH) precursor bearing required C–C connectivity to the target nanostructure represents the key transformation in the on-surface synthesis strategy. The synthesis of hexabenzo[*bc,ef,hi,kl,no,qr*]coronene (HBC) by Weiss *et al.* showed that this step can be realized efficiently under ultrahigh vacuum (UHV) conditions on atomically clean metal surfaces (*16*) (Fig. 1A). The catalytic activity of the metals substantially reduced the activation barrier of the cyclization (*8*). In 2010, Cai *et al.* applied a similar strategy to form atomically precise GNRs on the Au(111) surface (*11*) (Fig. 1B). This discovery paved the way toward the fabrication of complex molecular nanoarchitectures on selected noble metal surfaces (*17*). However, for most practical applications, a carbon-based nanostructure must be transferred onto insulating or semiconducting surfaces (*18*, *19*).

An attractive yet challenging way to tackle this problem is the controlled synthesis of carbon nanostructures directly on technologically relevant nonmetallic substrates, such as metal oxide surfaces (*20*–*22*). However, all reported attempts to perform the cyclization on metal oxides have been unsuccessful because of the lack of catalytic activity in the cyclodehydrogenation process (*20*) (Fig. 1C). The cyclization of PAH precursors on such surfaces requires high temperatures, which leads to a loss of selectivity. Thus, the rational synthesis of tailored carbon-based nanoarchitectures on metal oxide surfaces requires the development of an alternative cyclization technique. Previously, we have found that intramolecular aryl-aryl coupling can be effectively realized through C–F bond activation on γ-Al_2O_3 under relatively mild conditions (*23*–*27*). Further exploration revealed that metal oxides of III and IV groups also displayed activity in cyclodehydrofluorination at elevated temperatures (*28*). Among them, bulk powders of titanium dioxide activated C–F bonds at 570 K, which made it an attractive candidate for on-surface investigations performed under UHV conditions.

We present the rational on-surface synthesis of NGs on a semiconducting rutile TiO_2(011) surface through dominolike HF zipping of programmed fluoroarene precursors (Fig. 1D). The high potential of the approach was demonstrated by a challenging transformation consisting of the formal rolling up of the linear oligophenylene chain around a phenyl moiety, yielding NG HBC (Fig. 1D). In contrast to the commonly used rigid design of precursors, our approach allows the regioselectivity of the cyclization to be unambiguously programmed by F atom positions, providing sufficient flexibility in the design of precursor molecules.

To investigate the HF-zipping process on a rutile surface, two model NGs—namely, DBPP (dibenzo[*ij,rst*]phenanthro[9,10,1,2-*defg*:]pentaphene) and HBC—were chosen as target compounds (Fig. 2, A and B). DBPP, which can be considered

[1]Centre for Nanometer-Scale Science and Advanced Materials, NANOSAM, Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, Łojasiewicza 11, 30-348 Kraków, Poland. [2]Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. [3]Department of Organic Chemistry, Friedrich Alexander University Erlangen-Nuremberg, 91058 Erlangen, Germany. [4]Espeem S.A.R.L., L-4365 Esch-sur-Alzette, Luxembourg.
***Corresponding author. Email: konstantin.amsharov@fau.de (K.A.); kolmerma@ornl.gov (M.K.)**

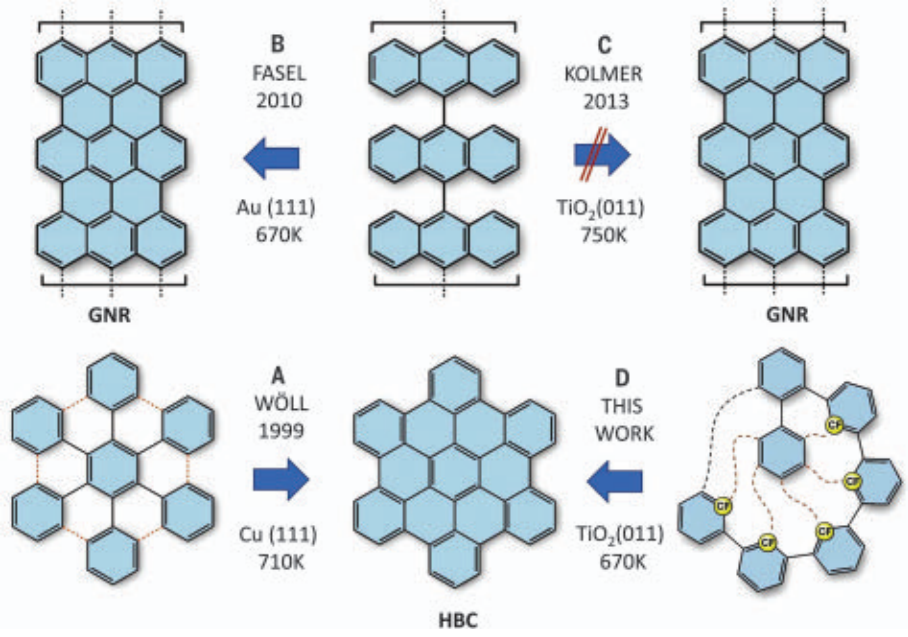


Fig. 1. Selected examples of on-surface syntheses of NGs and nanoribbons by a bottom-up approach. (**A**) First on-surface synthesis of NG HBC (*16*); (**B**) rational synthesis of GNRs on Au(111) (*11*); (**C**) attempts to perform cyclodehydrogenation on a metal oxide surface (*20*); (**D**) first rational on-surface synthesis of NGs on a nonmetallic surface (this work).

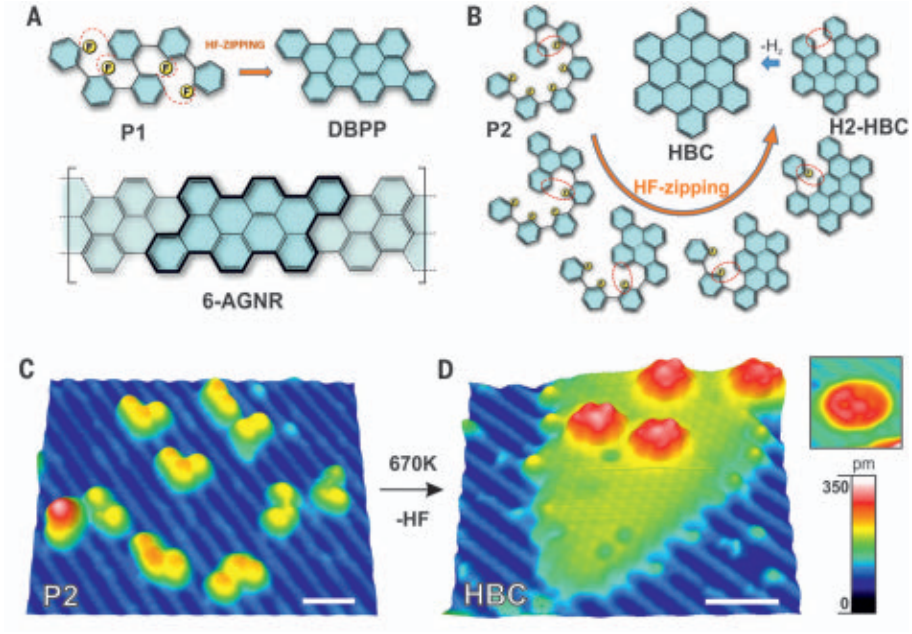


Fig. 2. HF elimination–based zipping of precursors. Schemes of (**A**) DBPP and (**B**) HBC formation concepts. (**C**) Three-dimensional (3D) visualization of an STM image (+2 V; 10 pA) of P2 deposited on rutile (011)-(2×1) at RT and annealed at 570 K for 10 min. Note different configurations of molecules. (**D**) 3D STM image (+2.5 V; 50 pA) obtained after sample annealing at 670 K. Single HBC molecules are adsorbed on newly formed ad-islands with different surface reconstruction (yellow-green). (Inset) STM image (+2 V, 10 pA, 2.5 nm by 2.5 nm) of a single HBC molecule. Scale bars in (C) and (D) are 2 nm.

an ultrashort armchair GNR (6-AGNR), represents the model for the on-surface synthesis of GNRs, and HBC is one of the smallest and best-characterized NGs (*29*–*33*). Both NGs possess easy-to-recognize geometries in scanning tunneling microscopy (STM) images. The required specially "programmed" fluorinated oligophenylenes P1 and P2 were obtained by multistep organic synthesis (for details, see the supplementary materials). The key feature of the cyclodehydrofluorination is the "switchable" activity of the C–F bond. Only C–F bonds with close proximity to a C–H bond displayed activity, whereas peripheral C–F remained completely intact. This reactivity enabled tandem cyclization via HF elimination in a truly dominolike fashion, because each subsequent cyclization step led to the "activation" of one new C–F bond (*26*). The active H–F pairs for the HF-zipping concept are shown schematically in Fig. 2, A and B.

All on-surface experiments, including low-temperature STM, x-ray photoemission spectroscopy (XPS), and mass spectrometry (MS) studies, were performed in situ under UHV conditions. We deposited precursor molecules by using standard Knudsen cells on the (2×1) reconstructed (011) face of the rutile TiO_2 (for details, see the supplementary materials). To thermally induce the transformation of P1, we started with submonolayer deposition of precursor molecules on the substrate kept at room temperature (RT) and then heated the substrate to ~570 K (bulk activation temperature) for 10 min. Under these mild conditions, most P1 molecules desorbed from the surface, leaving almost bare surface terraces with no clear evidence of successful HF elimination (see the supplementary materials). At a higher annealing temperature of ~670 K, particularly flat molecules with the specific rhomboid shape expected for DBPP were observed in our STM experiment. However, because of the appreciable thermal desorption of P1, DBPP molecules were adsorbed only on chemically active sites (step edges or domain boundaries), which complicated the accurate interpretation of their nonuniform STM contrast (see the supplementary materials).

With the larger precursor P2, after deposition onto an RT substrate, P2 molecules were found mostly in globular form on reconstructed terraces of rutile (011) with STM. However, after annealing at ~570 K, the molecules remained on the surface (Fig. 2C). Moreover, we observed a general change in P2 appearance and observed elongated geometries with lengths up to ~2.5 nm. These geometries correspond to different possible configurations of P2 on the surface (see detailed analysis in the supplementary materials), consistent with the expected flexibility of the precursors. This observation points out that the globular geometry, the favorable gas-phase configuration of P2 preserved after deposition, was only metastable on the surface.

To induce the transformation of P2, we heated the rutile (011) substrate to ~670 K for 10 min.

Although most of the molecules were found on step edges, some molecules appeared directly on newly formed ad-islands that allowed their high-quality STM visualization (Fig. 2D). The round shape and the size (~1.6-nm full width at half maximum, as shown by STM image cross sections in Fig. 3, A and B) of the molecules are in agreement with the expected HBC model (Fig. 3C), our unoccupied-states STM image simulations (see the supplementary materials), and previous reports (*29*, *30*, *32*). The presence of HBC molecules demonstrated the feasibility of the HF-zipping strategy on the rutile surface. The programmed P2 precursor was transformed via five consecutive cyclodehydrofluorination steps followed by the final cyclodehydrogenation into the target NG molecule.

Among molecules found on ad-islands, some possessed an additional protrusion located beside the center of the molecule, as marked by arrows in Fig. 3A. We attributed these images to molecules of helicenelike H2-HBC, presented in Fig. 3D, in which the final cyclodehydrogenation step was not accomplished. The presence of two H atoms in the fjord region caused deviation from planar geometry, with two corresponding benzene rings oriented out of plane. The high-resolution STM image of an HBC and H2-HBC pair in Fig. 3E confirms the expected position of the protrusion on the molecules. The uncopied-states STM contrast of H2-HBC is also reproduced in our STM image simulation (see the supplementary materials), which is in good agreement with the experimental image. The formation of the H2-HBC intermediate after only the first five cyclization reactions activated via F was rather unexpected, because the last cyclization step should occur spontaneously given the high strain in the fjord region. However, the observation of the H2-HBC intermediate indicates that, on the TiO_2 surface, the cyclodehydrofluorination process had a markedly lower activation barrier than cyclodehydrogenation. A similar strategy was in fact completely inefficient on metal surfaces. As we found for the Au(111) surface, catalytic activation of the competitive cyclodehydrogenation processes resulted in undefined molecular structures formed from P2 after their annealing (see results and discussion in the supplementary materials). Thus, metal oxide surface assistance in the cyclodehydrofluorination is the key aspect of the successful HF zipping (*28*). Although the cyclization mechanism is not fully understood, the most probable scenario includes the C–F bond polarization by the active Ti center, allowing synchronous Friedel-Crafts–like arylation (*34*, *35*).

The STM data show that the HF-zipping strategy was efficient, as many single-target molecules were found locally on the surface. Other than H2-HBC and HBC, no other molecular species were observed on the ad-islands where high-quality STM imaging was possible. To shed more light on the global outcome of the thermally induced HF elimination reaction, we combined the STM data with the XPS chemical analysis. The C1s core-level photoemission spectrum of P2 molecules deposited on the rutile (011) surface at RT (Fig. 4A) showed an asymmetric signal composed of two peaks separated by ~2 eV in binding energy (E_B), which correspond to C–C (red line, E_B = 284.4 eV) and C–F (green line, E_B = 286.3 eV) contributions (*36*, *37*). Figure 4B presents the F1s core-level region. For RT deposition, the single peak observed at E_B = 686.8 eV is related to an organofluorine component (*13*, *36*, *37*). The absence of a peak at ~684.5 eV suggests the lack of a F–Ti component (*38*). Thus, the XPS analysis confirmed that the C–F bonds in the precursor molecules were intact after deposition on the surface at RT.

After annealing at 670 K for 10 min, the XPS signal from the C1s region consisted of only a single peak at E_B = 284.5 eV, corresponding to a C–C component (Fig. 4A), confirming global scission of C–F bonds in P2 molecules caused by efficient HF elimination. The corresponding STM images (Fig. 3, A and E) show that the flat-lying HBC molecules were found on the newly formed reconstructed areas of the surface, forming ad-islands attached to TiO_2(011)-(2×1) step edges or domain boundaries. These structures were not observed in a control experiment, where we directly deposited HBC molecules on the rutile (011) and annealed the sample to 670 K (see the supplementary materials). Moreover, other halogenoarenes with bromine (*20*–*22*) or iodine (*21*) substituents annealed under similar conditions at rutile titania surfaces did not produce the observed ad-islands.

An obvious hypothesis is that F atoms are building blocks in these new structures. However, Fig. 4B shows that after annealing, the XPS signal from the F1s region was strongly reduced, consisted of only organic fluorine (<10% of the initial intensity), and provided no indication of a F–Ti component at ~684.5 eV (arrow). Lack of F at the surface region motivated us to look for fluorine-containing molecules desorbing from the rutile (011) surface after the HF-zipping reaction by monitoring corresponding MS signals during controlled sample heating from RT up to 770 K. We observed signals only for molecules of $TiOF_2$ (Fig. 4C) and TiOF (see the supplementary materials). Despite the sharp peak located at ~500 K, the main fraction of $TiOF_2$ desorbed when the temperature exceeded 670 K, consistent with our previous experimental observations.

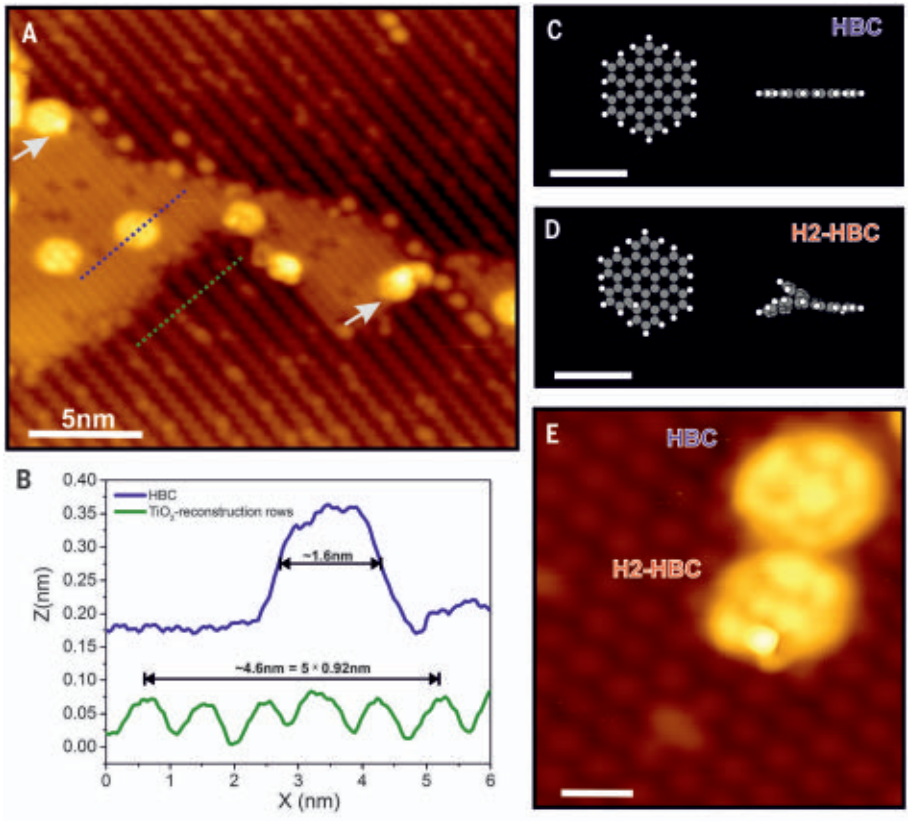


Fig. 3. HF versus H_2 elimination. (**A**) STM image (+2 V, 50 pA) presenting molecules of HBC and H2-HBC (arrows) formed from P2 after annealing at 670 K. (**B**) Cross sections along purple and green lines in (A). (**C** and **D**) Optimized geometries of (C) HBC and (D) H2-HBC molecules. Gray and white circles correspond to carbon and hydrogen atoms, respectively. (**E**) High-resolution STM image (+2.7 V, 200 pA) of HBC and H2-HBC on newly formed reconstruction of rutile (011). Scale bars in (C) to (E) are 1 nm.

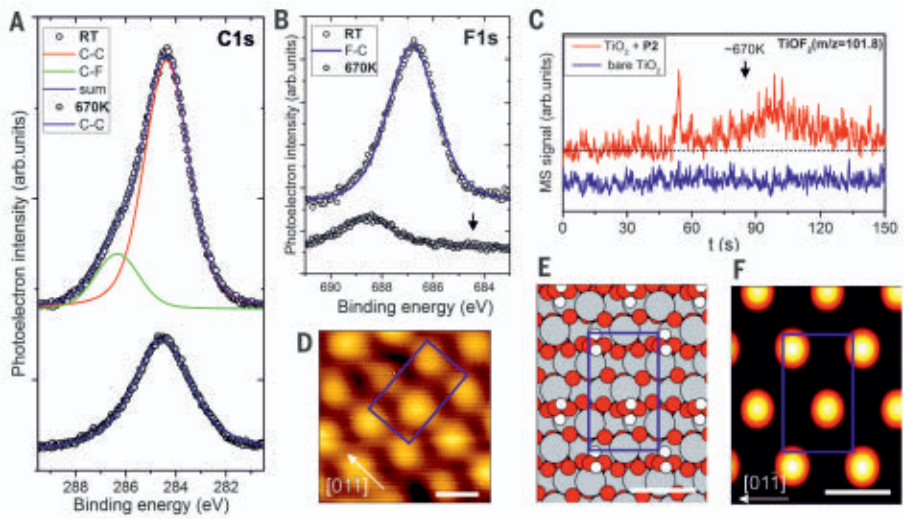


Fig. 4. Chemical analysis. (**A**) XPS C1s core-level spectra measured for P2 deposited on rutile (011) at RT (top) and after annealing to 670 K (bottom). arb., arbitrary. (**B**) XPS F1s core-level spectra measured for P2 deposited on rutile (011) at RT and after annealing to 670 K. The arrow marks the energy of the eventual F–Ti component. (**C**) Mass spectrometer signals of a $TiOF_2$ molecule (m/z = 101.8) registered during controlled heating of bare and P2-covered rutile TiO_2(011) from RT to around 770 K. The main fraction of thermally desorbed $TiOF_2$ molecules from the P2-covered surface is registered for temperatures exceeding 670 K. (**D**) High-resolution STM image (+2.7 V, 200 pA) presenting new reconstruction of the ad-island with apparent $c(2\times1)$ symmetry. (**E**) Calculated structure of the hydroxylated rutile TiO_2(011)-(1×1) surface forming (2×1) reconstruction. Gray, red, and white circles correspond to titanium, oxygen, and hydrogen atoms, respectively. (**F**) Unoccupied-states STM image simulation of the structure presented in (E). Scale bars in (D) to (F) are 0.5 nm. Purple rectangles mark corresponding unit cells.

Desorption of $TiOF_2$ points to strong Ti–F covalent bonding of F to the surface during the HF zipping, which may indirectly explain the observed surface reconstruction under newly formed HBC molecules. Adsorbates strongly interacting with the surface are known to induce changes in rutile titania reconstructions, as previously reported for (110) (*39*) and (011) (*40*, *41*) faces. Recent work by Balajka *et al.* (*41*) reported a water-induced reconstruction of a rutile (011) surface to form a bulk-terminated (1×1) face covered with a hydroxyl group overlayer forming (2×1) reconstruction. In the case of a HF elimination reaction, the desorption of $TiOF_2$ species also formally leads to the formation of water molecules, $TiO_2 + 2HF \rightarrow TiOF_2 + H_2O$, which then could dissociate and locally create the hydroxyl-rich reconstruction. Close inspection of the ad-island structure performed by high-resolution STM characterization (Fig. 4D) shows bright protrusions forming rows along the [01-1] direction, with neighboring ones shifted by half of the lattice, giving apparent pseudo-hexagonal $c(2\times1)$ symmetry. Results of our density functional theory calculations shown in Fig. 4E present the exact structure of the ad-islands with four hydroxyl groups (two dissociated water molecules) per (2×1) unit cell. Corresponding unoccupied-states STM image simulation (Fig. 4F) confirmed that for high positive bias voltages, this (2×1) surface reconstruction gave the apparent $c(2\times1)$ symmetry observed in the experimental data. These results support the rutile titania surface participation in intermediate states of the cyclodehydrofluorination and reflect its crucial role in this process. Our results provide a pathway toward custom-designed sp^2 carbon–based nanostructure formation by direct on-surface synthesis methods on technologically relevant semiconducting or insulating surfaces.

REFERENCES AND NOTES

1. J. Méndez, M. F. López, J. A. Martín-Gago, *Chem. Soc. Rev.* **40**, 4578–4590 (2011).
2. P. A. Held, H. Fuchs, A. Studer, *Chemistry* **23**, 5874–5892 (2017).
3. G. Otero *et al.*, *Nature* **454**, 865–868 (2008).
4. K. Amsharov *et al.*, *Angew. Chem. Int. Ed. Engl.* **49**, 9392–9396 (2010).
5. K. T. Rim *et al.*, *Angew. Chem. Int. Ed. Engl.* **46**, 7891–7895 (2007).
6. N. Abdurakhmanova *et al.*, *Carbon* **84**, 444–447 (2015).
7. J. R. Sanchez-Valencia *et al.*, *Nature* **512**, 61–64 (2014).
8. M. Treier *et al.*, *Nat. Chem.* **3**, 61–67 (2011).
9. C. Rogers *et al.*, *Angew. Chem. Int. Ed. Engl.* **54**, 15143–15146 (2015).
10. R. Zuzak *et al.*, *Chem. Commun.* **54**, 10256–10259 (2018).
11. J. Cai *et al.*, *Nature* **466**, 470–473 (2010).
12. P. Ruffieux *et al.*, *Nature* **531**, 489–492 (2016).
13. H. Hayashi *et al.*, *ACS Nano* **11**, 6204–6210 (2017).
14. C. Ma *et al.*, *Nat. Commun.* **8**, 14815 (2017).
15. D. J. Rizzo *et al.*, *Nature* **560**, 204–208 (2018).
16. K. Weiss *et al.*, *Angew. Chem. Int. Ed. Engl.* **38**, 3748–3752 (1999).
17. L. Talirz, P. Ruffieux, R. Fasel, *Adv. Mater.* **28**, 6222–6231 (2016).
18. P. B. Bennett *et al.*, *Appl. Phys. Lett.* **103**, 253114 (2013).
19. J. P. Llinas *et al.*, *Nat. Commun.* **8**, 633 (2017).
20. M. Kolmer *et al.*, *Angew. Chem. Int. Ed. Engl.* **52**, 10300–10303 (2013).
21. M. Kolmer *et al.*, *Chem. Commun.* **51**, 11276–11279 (2015).
22. G. Vasseur *et al.*, *J. Am. Chem. Soc.* **138**, 5685–5692 (2016).
23. K. Y. Amsharov, M. A. Kabdulov, M. Jansen, *Angew. Chem. Int. Ed. Engl.* **51**, 4594–4597 (2012).
24. N. Suzuki, T. Fujita, K. Y. Amsharov, J. Ichikawa, *Chem. Commun.* **52**, 12948–12951 (2016).
25. O. Papaianina, K. Y. Amsharov, *Chem. Commun.* **52**, 1505–1508 (2016).
26. A. K. Steiner, K. Y. Amsharov, *Angew. Chem. Int. Ed. Engl.* **56**, 14732–14736 (2017).
27. O. Papaianina *et al.*, *Angew. Chem. Int. Ed. Engl.* **56**, 4834–4838 (2017).
28. K. Amsharov, *Phys. Status Solidi B* **253**, 2473–2477 (2016).
29. P. Samorí, N. Severin, C. D. Simpson, K. Müllen, J. P. Rabe, *J. Am. Chem. Soc.* **124**, 9454–9457 (2002).
30. L. Gross *et al.*, *Phys. Rev. B* **71**, 165428 (2005).
31. W. H. Soe *et al.*, *ACS Nano* **6**, 3230–3235 (2012).
32. L. Gross *et al.*, *Science* **337**, 1326–1329 (2012).
33. T. Ardhuin, O. Guillermet, A. Gourdon, S. Gauthier, *J. Phys. Chem. C* **122**, 11905–11910 (2018).
34. J. Ichikawa, M. Yokota, T. Kudo, S. Umezaki, *Angew. Chem. Int. Ed. Engl.* **47**, 4870–4873 (2008).
35. D. Sharapa, A.-K. Steiner, K. Amsharov, *Phys. Status Solidi B* 10.1002/pssb.201800189 (2018).
36. F. Blobner *et al.*, *J. Phys. Chem. C* **119**, 15455–15468 (2015).
37. R. Han *et al.*, *Chem. Commun.* **52**, 9805–9808 (2016).
38. J.-Y. Ruzicka *et al.*, *RSC Adv.* **4**, 20649–20658 (2014).
39. D. Silber *et al.*, *Nat. Commun.* **7**, 12888 (2016).
40. Q. Cuan, J. Tao, X. Q. Gong, M. Batzill, *Phys. Rev. Lett.* **108**, 106105 (2012).
41. J. Balajka *et al.*, *J. Phys. Chem. C* **121**, 26424–26431 (2017).

ACKNOWLEDGMENTS

Funding: The research was supported by the Polish Ministry of Science and Higher Education, contract no. 0341/IP3/2016/74. R.Z. acknowledges support received from the National Science Center, Poland (2017/24/T/ST5/00262). A.K.S. and K.A. thank the Deutsche Forschungsgemeinschaft (DFG-SFB 953 "Synthetic Carbon Allotropes" project A6, and AM407) for financial support. Work was partially conducted at the Center for Nanophase Materials Sciences (CNMS), which is a DOE Office of Science User Facility. The research was partially supported from basic and statutory funds of the Jagiellonian University in Krakow provided by the Polish Ministry of Science and Higher Education. **Author contributions:** K.A. and M.K. conceived the project. K.A. and A.K.S. carried out the precursor synthesis and analysis. M.K. and R.Z. conducted the on-surface synthesis and low-temperature STM, XPS, and MS analyses with support from L.Z., S.G., and M.S. M.E. conducted the computations. M.K. and K.A. prepared the manuscript with feedback from all other authors. **Competing interests:** None declared. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/363/6422/57/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S34
Table S1
References (*42*–*45*)

21 September 2018; accepted 16 November 2018
10.1126/science.aav4954

ATOMIC PHYSICS

Laser cooling of ions in a neutral plasma

Thomas K. Langin, Grant M. Gorman, Thomas C. Killian*

Laser cooling of a neutral plasma is a challenging task because of the high temperatures typically associated with the plasma state. By using an ultracold neutral plasma created by photoionization of an ultracold atomic gas, we avoid this obstacle and demonstrate laser cooling of ions in a neutral plasma. After 135 microseconds of cooling, we observed a reduction in ion temperature by up to a factor of four, with the temperature reaching as low as 50(4) millikelvin. This pushes laboratory studies of neutral plasmas deeper into the strongly coupled regime, beyond the limits of validity of current kinetic theories for calculating transport properties. The same optical forces also retard the plasma expansion, opening avenues for neutral-plasma confinement and manipulation.

The application of laser cooling to trapped ions (*1*–*3*) and neutral atoms (*4*) has driven groundbreaking advances in physics, including quantum computing (*5*), Wigner crystallization (*6*–*8*), and quantum degeneracy (*9*). More recently, laser-cooling techniques have been applied to molecules (*10*), solids (*11*, *12*), and mesoscopic quantum objects (*13*). In this study, using an ultracold neutral plasma (UNP) created by photoionizing Sr atoms in an ultracold gas (*14*, *15*), we demonstrate laser cooling of ions in a neutral plasma.

A major motivation for cooling ions in a neutral plasma is to study transport, thermalization, and collective modes when the ratio between the Coulomb potential energy and thermal kinetic energy of ions is large [$\Gamma_i = e^2/4\pi\varepsilon_0 a k_B T_i \gtrsim 1$, where T_i is the ion temperature, $a = (3/4\pi n_i)^{1/3}$ is the Wigner-Seitz radius for density n_i, e is the elementary charge, ε_0 is the vacuum permittivity, and k_B is the Boltzmann constant]. Standard kinetic theories (*16*) fail to describe important plasma properties under such conditions of strong Coulomb coupling because they neglect effects of spatial correlations. Such conditions are found in white dwarf stars (*6*, *17*), cores of Jovian planets (*6*), dusty plasmas (*18*), trapped nonneutral plasmas (*3*, *7*, *8*), laser-produced plasmas important for studies of warm dense matter and inertial confinement fusion (*17*, *19*), and UNPs. State-of-the-art predictions of transport and thermalization rates for high-density strongly coupled plasmas (*20*) are currently obtained from direct molecular dynamics simulation (*21*–*23*) of the Yukawa one-component-plasma (OCP) model (*6*, *17*, *24*). UNPs are highly faithful realizations of the Yukawa OCP model (*15*). With precise temperature and density diagnostics (*14*) and pump-probe techniques for studying kinetic processes (*25*), UNPs offer many advantages for studying the effects of strong coupling on collisional processes (*15*, *24*, *25*) and validating numerical methods.

With the use of standard methods for UNP creation, coupling in these systems is limited to $\Gamma_i \lesssim 3$ by disorder-induced heating (DIH), which occurs immediately after plasma creation from the disordered atomic or molecular gas. DIH decreases the ion Coulomb energy and increases the temperature to $T_{DIH} \approx e^2/12\pi\varepsilon_0 a k_B$ as ions develop spatial correlations (*14*, *15*, *26*). For typical UNP densities, $T_{DIH} \sim 1$ K. Many schemes have been proposed to overcome this limit, such as precorrelating the system before ionization by using Rydberg blockade (*27*, *28*), Fermi repulsion (*26*), an optical lattice (*28*), and molecular Rydberg plasmas in a supersonic beam (*29*). Sequential excitation to higher ionization states (*30*) has been shown to increase Γ by 40%. Experiments using Rydberg blockade (*31*) and molecular plasmas (*29*) have yielded promising results, but no measured values of Γ have been reported. In this work, we realized laser-cooling proposals (*32*, *33*) and achieved Γ_i as high as 11(1).

Laser cooling works through velocity-dependent scattering and exchange of momentum between near-resonant photons and ions, molecules, or atoms. For simple Doppler cooling (*4*) using a transition with natural linewidth γ (expressed in hertz), excitation wavelength λ, and laser detuning $\Delta \sim -\gamma$, particles with velocity substantially outside the capture range ($v_c = \lambda\gamma$) are Doppler-shifted too far out of resonance for appreciable light scattering. Thus, cooling is most effective in systems with $v_T \lesssim v_c$ for characteristic thermal velocity $v_T = (k_B T/m)^{1/2}$, where m is the particle mass. For typical optical transitions, $v_c \approx 10$ m/s, requiring $T \lesssim 1$ K. This is one reason that laser cooling has not been successfully applied to ions in non-ultracold neutral plasmas, which are invariably much hotter than UNPs. UNPs provide the required low initial ion temperature, but high collision rates (*25*) and hydrodynamic expansion of the plasma into surrounding vacuum (*14*, *15*) create an environment that differs markedly from other systems that have been laser cooled.

To create UNPs, we initially cooled 5×10^8 ^{88}Sr atoms to $T = 1$ mK and magnetically trapped them in the metastable $5s5p\ ^3P_2$ state by using standard laser-cooling techniques (*34*). Trapping fields and cooling lasers were then extinguished, and the atom cloud expanded for 6 ms before a pulse (10 mJ for 10 ns) of $\lambda_{pulse} = 322$ nm photons from a doubled, pulsed-dye laser ionized 10% of the atoms. The plasma density distribution was well approximated by a slightly asymmetric Gaussian distribution with root mean square (RMS) radii $\sigma_x(0) = 2.4(1)$ mm and $\sigma_{y/z}(0) = 3.1(1)$ mm and peak density $n_i(0) = 1.3(3) \times 10^8$ cm^{-3}. This yielded a peak $T_{DIH} = 0.41(0.03)$ K. λ_{pulse} was tuned such that $\Delta E = hc/\lambda_{pulse} - E_{PI}$, the excess photon energy above the photoionization threshold energy E_{PI}, set the electron temperature to $T_e(0) = 2\Delta E/3k_B = 15.5(3)$ K. (Here, h is Planck's constant, and c is the speed of light.) Electrons

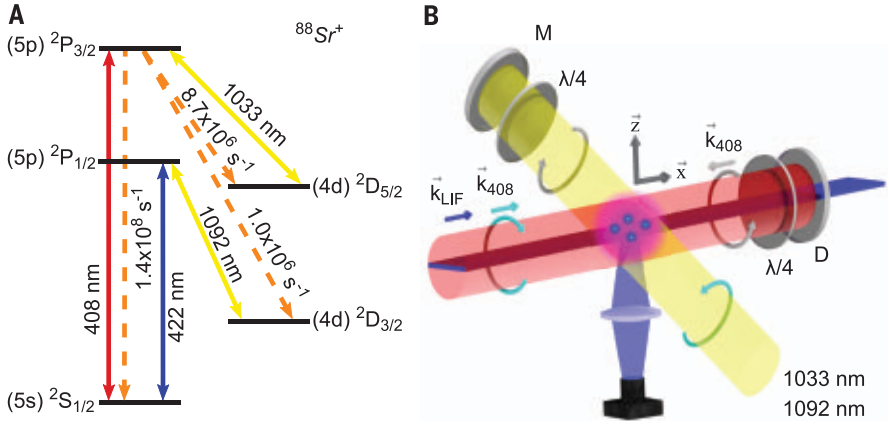


Fig. 1. Principles of laser cooling of a neutral plasma. (**A**) Sr^+-level diagram indicating the wavelengths and decay rates for transitions relevant to cooling and imaging. (**B**) Experimental schematic. Cooling (408-nm) and repumping (1092- and 1033-nm) lasers were applied in counter-propagating configurations with the indicated polarizations. Light at 422 nm for LIF was shaped by a slit to illuminate a central slice of the plasma. It propagated perpendicular to the imaging axis ($\hat{z}$) and, unless otherwise specified, along the laser-cooling axis ($\hat{x}$). Propagation directions for LIF and cooling light are indicated. LIF images were recorded on an intensified charge-coupled device camera. A high-pass dichroic *D* reflected the cooling laser and transmitted the LIF light. M, mirror; λ/4, quarter-wave plate).

Department of Physics and Astronomy, Rice University, 6100 Main Street, Houston, TX 77005, USA.
*Corresponding author. Email: killian@rice.edu

were trapped by Coulomb attraction to the ions and formed a neutralizing background with overall nonneutrality on the order of a few percent (*14*, *15*).

Immediately after plasma formation, counterpropagating σ^+ and σ^- polarized beams near resonance with the 5s $^2S_{1/2}$–5p $^2P_{3/2}$ transition at 408 nm (γ = 22.4 MHz) illuminated the plasma, forming a one-dimensional optical molasses (*4*) for laser cooling along the x direction (Fig. 1). The single-beam peak intensity was I = 100 mW/cm^2 (1/e^2-intensity radius w = 9 mm, and saturation parameter s_0 = 2.3). Lasers at 1092 and 1033 nm repumped ions from long-lived $^2D_{3/2}$ and $^2D_{5/2}$ states, returning them to the cooling cycle. Reductions in cooling efficiency caused by coherent coupling of the $^2S_{1/2}$ and $^2D_{5/2}$ states and resulting electromagnetically induced transparencies (*35*) were minimized by rapid velocity-changing collisions in the plasma (*25*). After laser cooling, spatially resolved measurements of ion temperature and density were performed by using laser-induced fluorescence (LIF) on the $^2S_{1/2}$–$^2P_{1/2}$ transition at 422 nm (*36*). The LIF laser illuminated a central slice of the plasma ($z \approx 0$), providing transverse spatial resolution of 13 μm and resolution along the imaging axis equal to the tight dimension of the LIF light (w_z = 2 mm). Unless otherwise specified, temperature was measured along the laser-cooling axis (Fig. 1).

Figure 2, A to C, shows temperature measurements, spatially resolved along the laser-cooling axis, for three different laser-cooling parameters: no cooling light (yellow), detuning from resonance of Δ = −20 MHz [cooling (red)], and Δ = +20 MHz [heating (blue)]. By 5 μs after plasma formation, cooling and heating beams had little effect. Temperature variation across the sample reflected variation in density and the resulting DIH temperature ($T_{DIH} \propto n_i^{1/3}$). By 60 μs, the ion temperature was substantially altered by the lasers. In the center of the plasma ($x \approx 0$), T_i doubled in the heating configuration and was reduced by half for cooling compared with no cooling light. The trend continued for 135 μs of heating or cooling, with the lowest temperature observed, T_i = 50(4) mK, providing clear evidence of laser cooling.

Cooling was effective only in the central region ($|x| \lesssim 1$ mm), which illustrates the role of plasma expansion. Immediately after plasma formation, the density gradient and electron thermal energy created a radially directed force that drives expansion (*14*). Although electron-ion thermalization has noticeable effects (*37*) and the plasma used in this work was slightly asymmetric, an analytic expansion model for a collisionless, spherically symmetric, Gaussian plasma is a satisfactory approximation for developing an intuitive explanation of laser-cooling results. In this model (*14*, *15*), expansion creates a hydrodynamic plasma-velocity field $\vec{v}_{exp}(r,t) = [t/(t^2 + \tau_{exp}^2)]\vec{r}$ (where r is the displacement from the plasma center, t is time, and $\tau_{exp} = [m_i\sigma(0)^2/k_B T_e(0)]^{1/2}$ is a characteristic time for the expansion). The velocity increases with time and distance from the plasma center, saturating at an RMS value for the entire plasma of $\sqrt{\langle v_{exp}^2 \rangle} = \sigma(0)/\tau_{exp}$ on a time scale of τ_{exp}. Cooling is effective only in regions for which the expansion velocity along the cooling-laser axis remains less than or comparable to the velocity capture range [$|\hat{\mathbf{x}} \cdot \vec{v}_{exp}(\vec{r},t)| \lesssim v_c$] for an appreciable time.

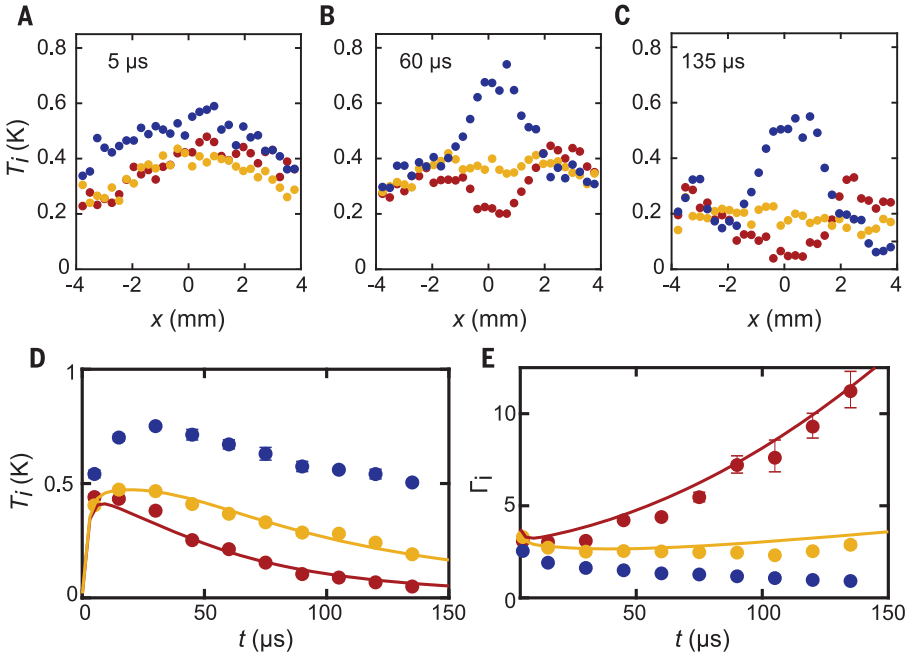


Fig. 2. Ion temperature and coupling variation. Shown are the data for laser detuning Δ = −20 MHz [red (cooling)], Δ = 20 MHz [blue (heating)], and no 408-nm laser (yellow). (**A** to **C**) Temperature at three different times after plasma creation (as indicated in each panel). Each measurement corresponds to a region with Δx = 260 μm and Δy = 4.5 mm centered at y = 0. Cooling and heating were ineffective at large displacement from the plasma center because the plasma expansion Doppler-shifted these ions out of resonance with the lasers. (**D**) Temperature versus time in the central region ($|x| \leq 0.5$ mm). (**E**) Coulomb coupling parameter Γ_i versus time in the central region. Solid lines in (D) and (E) were calculated by using coupled ion and electron kinetic equations (*36*). We do not display results from the model for the heating data because the density perturbation was too severe for the model to be applicable (Fig. 3A). Error bars indicate SD.

For these experiments, $\sigma(0)/\tau_{exp} \approx$ 40 m/s exceeded the velocity capture range, and $\tau_{exp} \approx$ 75 μs was on the order of the minimum time required to scatter enough photons to substantially cool the ions (*4*). Thus, cooling was most effective for $|x| \lesssim 1$ mm, where the expansion velocity stayed relatively small. Analogous statements can be made for the heating configuration. To observe substantial laser cooling, it was essential to create very large plasmas compared with those in previous UNP experiments (*14*, *15*), which increased τ_{exp} and gave more time for laser cooling.

Figure 2D shows the evolution of the ion temperature for the center of the plasma, where cooling and heating were most effective. Fits from an approximate kinetic model (*14*, *37*) accounting for DIH, electron-ion heating, adiabatic cooling, and laser cooling describe the data reasonably well (*36*). The natural dynamics, seen in the data for no laser cooling, was an increase to $T_i \sim$ 0.4 K within the first few microseconds, owing to DIH, followed by cooling caused by adiabatic expansion. In the presence of the cooling lasers, however, the ion temperature dropped much farther than that for no 408-nm light. Even more notable is the increase of Γ_i to 11(1) after 135 μs (Fig. 2E), which is comparable to conditions of interest in white dwarf stars (*6*, *17*), for example. It is also much deeper into the strong coupling regime than has previously been reported with these systems. A value of Γ_i = 4 was obtained (*38*) in a UNP with relatively high density and low electron temperature, which induces substantial screening of ion-ion interactions [screening parameter $\kappa = a/\lambda_D$ = 1 for Debye screening length $\lambda_D = (k_B T_e \varepsilon_0/n_e e^2)^{1/2}$] and reduces the effects of strong coupling. For our conditions, $\kappa \leq$ 0.7, and the screening was weaker.

Although laser-cooling was applied only along one axis, all three degrees of freedom were effectively cooled because of the high collision rate in the plasma. Molecular dynamics simulations (*39*) show that local thermal equilibrium is established on the time scale of a few times the inverse of the ion plasma oscillation frequency, $\omega_{pi}^{-1} = \sqrt{\varepsilon_0 m_i/n_i e^2} \lesssim 2$ μs. We confirmed cross-thermalization by measuring the temperature transverse to the cooling axis (*36*). Lasers along a single axis also produce three-dimensional cooling in trapped ions (*1*, *2*), but this is not true for neutral atoms (*4*) and molecules (*10*), which have much lower collision rates. Collisions did not damp the hydrodynamic expansion velocity transverse to the laser-cooling axis because of conservation of momentum during each collision.

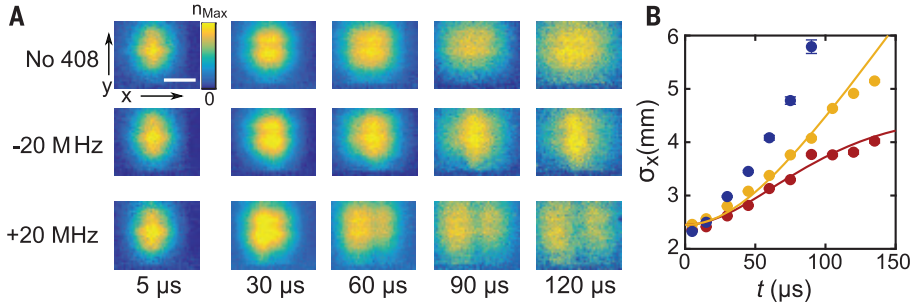


Fig. 3. Influence of laser forces on plasma expansion. (**A**) Evolution of plasma density distribution. Red-detuned optical molasses along $\hat{x}$ retarded expansion, whereas blue-detuned light accelerated it, eventually leading to bifurcation. The scale bar is 5 mm. The color bar is rescaled for each time to n_{max} = (13, 8.5, 4.2, 2.2, 1.3) × 10^7 cm^{-3} for t = (5, 30, 60, 90, 120) μs. (**B**) RMS radius $\sigma_x(t)$ from Gaussian fit to experimental data (same symbols as in Fig. 2). Error bars indicate SD.

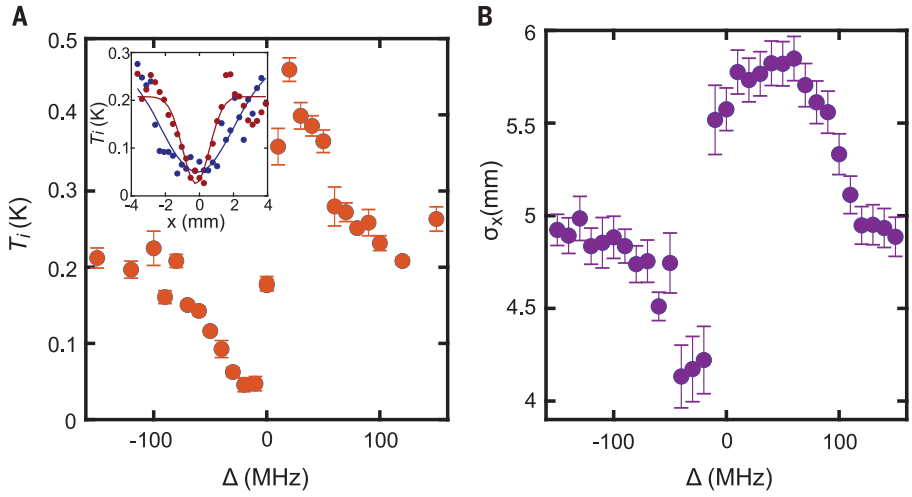


Fig. 4. Effects of detuning. (**A**) Temperature at x = 0 versus detuning for 135 μs of evolution in the optical field. The inset shows T_i versus x for Δ = −10 MHz (red) and Δ = −30 MHz (blue), along with Gaussian fits. (**B**) Plasma size versus detuning at 135 μs of evolution. Here, size is the RMS width calculated numerically from the image. Error bars indicate SD.

The light-scattering force from the cooling lasers also notably retarded the expansion of the plasma, as shown in Fig. 3 for the same initial plasma conditions as in Fig. 2. The kinetic model discussed above provides a reasonable description of the evolution of the RMS size along the laser axis (Fig. 3B). For larger initial plasma size and detuning of the cooling laser, retardation of the expansion was even more effective (*36*), raising the intriguing possibility of confining a neutral plasma with optical forces, perhaps with the addition of spatially varying magnetic fields in a hybrid magneto-optical trap (*4*) and magnetic cusp (*40*) configuration. For blue detuning, the plasma bifurcated because ions were accelerated away from the center until the force diminished when the velocity exceeded the capture range. The marked change in expansion dynamics for blue detuning highlights the potential of laser forces for manipulating a plasma.

Figure 4 illustrates the variation of laser-cooling and heating effects with detuning Δ. After 135 μs of laser application, Γ_i in the central region was minimized for Δ ~ −20 MHz. However, as the lasers were further red detuned, more of the cloud experienced the light-scattering forces for a longer time, creating a larger region of cooled ions at the expense of slightly decreased cooling efficiency (Fig. 4A, inset). Similarly, the acceleration of plasma expansion for the heating configuration was most effective for a relatively large blue detuning, Δ ~ 50 MHz, as shown in the plot of RMS size along the laser axis (Fig. 4B).

With Γ_i = 11, laser-cooled UNPs are already in an interesting regime for applying established techniques for measuring collision rates, transport, and dispersion relations (*15*, *25*). Laser forces also open entirely new possibilities. For example, by patterning the lasers spatially, it should be possible to create reasonably sharp velocity gradients and measure shear viscosity. Heating and cooling separate regions of the plasmas should initiate heat flow and allow measurement of thermal conductivity. There are also straightforward opportunities to improve the laser-cooling process, such as investigating cooling along three dimensions, which might be effective for plasma confinement. Increasing the number of ionized atoms would allow for larger σ_0 and τ_{exp} for a given density. Laser cooling could then be applied for a longer time, leading to lower temperature and higher Γ_i. The lowest temperature theoretically achievable with Doppler cooling on this transition is $T_{Doppler} = \hbar\gamma/2k_B$ = 0.5 mK (*4*), and even with present conditions it is likely that cooling can continue longer than reported here and below 50 mK. However, the natural linewidth of the LIF transition is already large compared with the Doppler broadening at this temperature, limiting further application of this temperature diagnostic. Improving the temperature resolution by, for example, using a narrow two-photon transition to a metastable D state should enable measurement of lower temperature. Then it will be possible to study the laser-cooling and laser-confinement limits imposed by electron-ion heating and three-body recombination (*14*, *15*, *32*).

REFERENCES AND NOTES

1. D. Wineland, R. Drullinger, F. Walls, *Phys. Rev. Lett.* **40**, 1639–1642 (1978).
2. W. Neuhauser, M. Hohenstatt, P. Toschek, H. Dehmelt, *Phys. Rev. Lett.* **41**, 233–236 (1978).
3. D. H. E. Dubin, T. M. O'Neil, *Rev. Mod. Phys.* **71**, 87–172 (1999).
4. H. Metcalf, P. van der Straten, *Laser Cooling and Trapping* (Springer-Verlag, 1999).
5. H. Haffner, C. F. Roos, R. Blatt, *Phys. Rep.* **469**, 155–203 (2008).
6. S. Ichimaru, *Rev. Mod. Phys.* **54**, 1017–1059 (1982).
7. W. M. Itano *et al.*, *Science* **279**, 686–689 (1998).
8. A. Mortensen, E. Nielsen, T. Matthey, M. Drewsen, *Phys. Rev. Lett.* **96**, 103001 (2006).
9. C. J. Pethick, H. Smith, *Bose–Einstein Condensation in Dilute Gases* (Cambridge Univ. Press, ed. 2, 2008).
10. E. S. Shuman, J. F. Barry, D. Demille, *Nature* **467**, 820–823 (2010).
11. D. V. Seletskiy *et al.*, *Nat. Photonics* **4**, 161–164 (2010).
12. R. I. Epstein, M. I. Buchwald, B. C. Edwards, T. R. Gosnell, C. E. Mungan, *Nature* **377**, 500–503 (1995).
13. M. Aspelmeyer, T. J. Kippenberg, F. Marquardt, *Rev. Mod. Phys.* **86**, 1391–1452 (2014).
14. T. C. Killian, T. Pattard, T. Pohl, J. M. Rost, *Phys. Rep.* **449**, 77–130 (2007).
15. M. Lyon, S. L. Rolston, *Rep. Prog. Phys.* **80**, 017001 (2017).
16. L. G. Stanton, M. S. Murillo, *Phys. Rev. E* **93**, 043203 (2016).
17. M. S. Murillo, *Phys. Plasmas* **11**, 2964–2971 (2004).
18. G. E. Morfill, A. V. Ivlev, *Rev. Mod. Phys.* **81**, 1353–1404 (2009).
19. B. A. Remington, R. P. Drake, D. D. Ryutov, *Rev. Mod. Phys.* **78**, 755–807 (2006).
20. J. Clérouin, P. Arnault, C. Ticknor, J. D. Kress, L. A. Collins, *Phys. Rev. Lett.* **116**, 115003 (2016).
21. J. Daligault, *Phys. Rev. Lett.* **108**, 225004 (2012).
22. J. Daligault, K. O. Rasmussen, S. D. Baalrud, *Phys. Rev. E* **90**, 033105 (2014).
23. Z. Donkó, P. Hartmann, *Phys. Rev. E* **69**, 016405 (2004).
24. M. S. Murillo, *Phys. Rev. Lett.* **96**, 165001 (2006).
25. T. S. Strickler, T. K. Langin, P. McQuillen, J. Daligault, T. C. Killian, *Phys. Rev. X* **6**, 021021 (2016).
26. M. S. Murillo, *Phys. Rev. Lett.* **87**, 115003 (2001).
27. G. Bannasch, T. C. Killian, T. Pohl, *Phys. Rev. Lett.* **110**, 253003 (2013).
28. D. O. Gericke, M. S. Murillo, *Contrib. Plasma Phys.* **43**, 298–301 (2003).
29. H. Sadeghi *et al.*, *Phys. Rev. Lett.* **112**, 075001 (2014).
30. M. Lyon, S. D. Bergeson, A. Diaw, M. S. Murillo, *Phys. Rev. E* **91**, 033101 (2015).
31. M. Robert-de-Saint-Vincent *et al.*, *Phys. Rev. Lett.* **110**, 045004 (2013).
32. S. G. Kuzmin, T. M. O'Neil, *Phys. Plasmas* **9**, 3743–3751 (2002).
33. T. Pohl, T. Pattard, J. M. Rost, *Phys. Rev. Lett.* **92**, 155003 (2004).
34. S. B. Nagel *et al.*, *Phys. Rev. A* **67**, 011401 (2003).
35. M. Fleischhauer, A. Imamoglu, J. P. Marangos, *Rev. Mod. Phys.* **77**, 633–673 (2005).

36. Further experimental and simulation details are available as supplementary materials.
37. P. McQuillen, T. Strickler, T. K. Langin, T. C. Killian, *Phys. Plasmas* **22**, 033513 (2015).
38. M. Lyon, S. D. Bergeson, M. S. Murillo, *Phys. Rev. E* **87**, 033101 (2013).
39. S. D. Baalrud, J. Daligault, *Contrib. Plasma Phys.* **57**, 238–251 (2017).
40. H. Grad, *Containment in Cusped Plasmas Systems* (Courant Institute of Mathematical Sciences, New York Univ., 1961).
41. T. K. Langin, G. M. Gorman, T. C. Killian, Replication data for: Laser cooling of ions in a neutral plasma, Version 1, Harvard Dataverse (2018); https://doi.org/10.7910/DVN/8ZOVUB.

ACKNOWLEDGMENTS

We thank I. Plompen for experimental assistance. **Funding:** This work was supported by the Air Force Office of Scientific Research through grant FA9550-17-1-0391 and by the NSF/DOE Partnership in Basic Plasma Science and Engineering through the DOE/SC Office of Fusion Energy Sciences grant DE-SC0014455. **Author contributions:** T.K.L. constructed the apparatus, performed experiments, analyzed data, and performed numerical simulations. G.M.G. performed experiments and analyzed data. T.C.K. conceived of and designed the experiment. All authors discussed results and contributed to the preparation of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data shown in this work can be found in the Harvard Dataverse (*41*).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/363/6422/61/suppl/DC1
Supplementary Text
Figs. S1 to S4
References (*42*–*49*)

14 February 2018; accepted 13 November 2018
10.1126/science.aat3158

NEUROSCIENCE

A collicular visual cortex: Neocortical space for an ancient midbrain visual structure

Riccardo Beltramo[1,2,3]* and Massimo Scanziani[1,2,3]*

Visual responses in the cerebral cortex are believed to rely on the geniculate input to the primary visual cortex (V1). Indeed, V1 lesions substantially reduce visual responses throughout the cortex. Visual information enters the cortex also through the superior colliculus (SC), but the function of this input on visual responses in the cortex is less clear. SC lesions affect cortical visual responses less than V1 lesions, and no visual cortical area appears to entirely rely on SC inputs. We show that visual responses in a mouse lateral visual cortical area called the postrhinal cortex are independent of V1 and are abolished upon silencing of the SC. This area outperforms V1 in discriminating moving objects. We thus identify a collicular primary visual cortex that is independent of the geniculo-cortical pathway and is capable of motion discrimination.

The mammalian cerebral cortex receives sensory information from several modalities. Even within the same modality, sensory information reaches the cortex via anatomically distinct parallel pathways. The relative roles of these distinct pathways in driving cortical responses to a sensory stimulus and the extent to which their sensory representations are spatially segregated in the cortex are still matters of debate (*1*).

In the visual system, dorsolateral geniculate nucleus (dLGN) innervation of the primary visual cortex (V1) is considered the primary entry point of retinal input to the cortex (*2*). V1 lesions abolish or strongly reduce visually evoked activity in several higher cortical visual areas (*3*–*6*). The colliculo-cortical pathway provides visual input to the cortex from the superior colliculus (SC) via the pulvinar nucleus of the thalamus (*7*–*10*). Its inactivation has either no or only a slight and feature-selective effect on cortical visual responses (*9*, *11*–*13*). Thus, despite a clear anatomical link from SC to visual cortex, no cortical area has been identified yet whose visual responses rely entirely on visual activity originating from the SC.

Visual responses in the mouse cortical area postrhinal cortex (POR) are well documented (*14*–*16*). Although generally assumed to rely on V1, their dependence on V1 has not been directly assessed. We determined the impact of V1 on POR's visual responses in awake, head-fixed mice. We optogenetically silenced V1 while performing simultaneous electrophysiological recordings from V1 and POR, ensuring that the receptive fields of the respective recording sites overlapped (Fig. 1 and fig. S1A). Visual areas in mice are anatomically defined by their retinotopic afferent input originating from V1 (*14*, *16*). We thus injected the anterograde viral tracer AAV1.CAG.TdTomato in the posterior portion of V1 and identified POR via transcranial epifluorescence illumination of the labeled V1 axons projecting to the visual areas surrounding V1 (Fig. 1A and fig. S2). Drifting gratings displayed on a monitor at the center of the receptive field triggered responses in 40% (81 of 209) of the units isolated from POR [average firing rate ± SEM of visually evoked responses for regular-spiking (RS), putative excitatory neurons: 1.61 ± 0.25 Hz, n = 41; for fast-spiking (FS), putative inhibitory interneurons: 3.14 ± 0.43 Hz, n = 40; 5 mice]. We silenced V1 by photoactivating cortical inhibitory interneurons expressing channelrhodopsin-2 (ChR2). This approach abolished visual responses in RS neurons across the entire cortical depth (fig. S3) and over large areas of V1 (fig. S4) (*17*). Despite this extensive silencing of V1, however, most of the visual response in POR persisted (Fig. 1, C and D) [21.65 ± 6.51% average decrease ± SEM in visually evoked firing rate of RS cells (P = 0.022, n = 41); 34.5 ± 5.03% of FS cells (P < 0.0001, n = 40, 5 mice; Wilcoxon signed-rank test)]. Whereas the response latencies in V1 and POR were quite similar (fig. S5), the time courses of the peristimulus time histogram (PSTH) in V1 and POR were markedly different (Fig. 1, C and D, and fig. S6).

Which structure other than V1 could relay visual input to POR? The dLGN is also a source of afferent input to other visual areas (*16*). To determine whether POR directly receives input from the dLGN, we injected cholera toxin subunit B (CTB) in POR (Fig. 1E). The dLGN was almost devoid of retrogradely labeled neurons. The vast majority (>99%) of retrogradely labeled cell bodies in the visual thalamus were found in the pulvinar (*18*) [also called the latero-posterior nucleus in rodents (*10*)].

The pulvinar receives a massive afferent input from V1 (*1*, *9*), and its response to visual stimuli depends on V1 (*1*, *19*). Because silencing of V1 has a minor effect on POR activity, the pulvinar might seem a poor candidate for relaying visual activity to POR. However, the pulvinar is also a key node of the colliculo-cortical pathway because it receives direct input from SC (*10*, *20*). If there were pulvinar neurons whose visually evoked activity was driven by SC and unaffected by V1 silencing, then such neurons could mediate responses in POR that are independent of V1. We first determined whether there are neurons in the pulvinar that are visually driven by SC (Fig. 2). Injections in V1 and SC with viral anterograde tracers revealed that axons originating from SC preferentially target the caudal pulvinar, whereas axons originating from V1 preferentially target the rostral pulvinar (Fig. 2A). To determine whether V1 and SC inputs are also separated functionally, we recorded responses from the caudal or rostral pulvinar while optogenetically silencing V1 or SC (Fig. 2, B to D). We presented dark moving dots, visual stimuli known to drive robust activity in the SC (*21*). Photoinhibition of SC strongly attenuated its responses to the visual stimuli, particularly in the stratum opticum (Fig. 2B) (79.34 ± 4.31% average decrease ± SEM in visually evoked firing rate; P < 0.0001, Wilcoxon signed-rank test, n = 33, 5 mice) (see also fig. S7). SC silencing similarly attenuated

[1]Center for Neural Circuits and Behavior, Neurobiology Section, and Department of Neuroscience, University of California San Diego, La Jolla, CA, USA. [2]Department of Physiology, University of California San Francisco, San Francisco, CA, USA. [3]Howard Hughes Medical Institute, University of California San Francisco, San Francisco, CA, USA.
***Corresponding author. Email: ric.beltramo@gmail.com (R.B.); massimo@ucsf.edu (M.S.)**

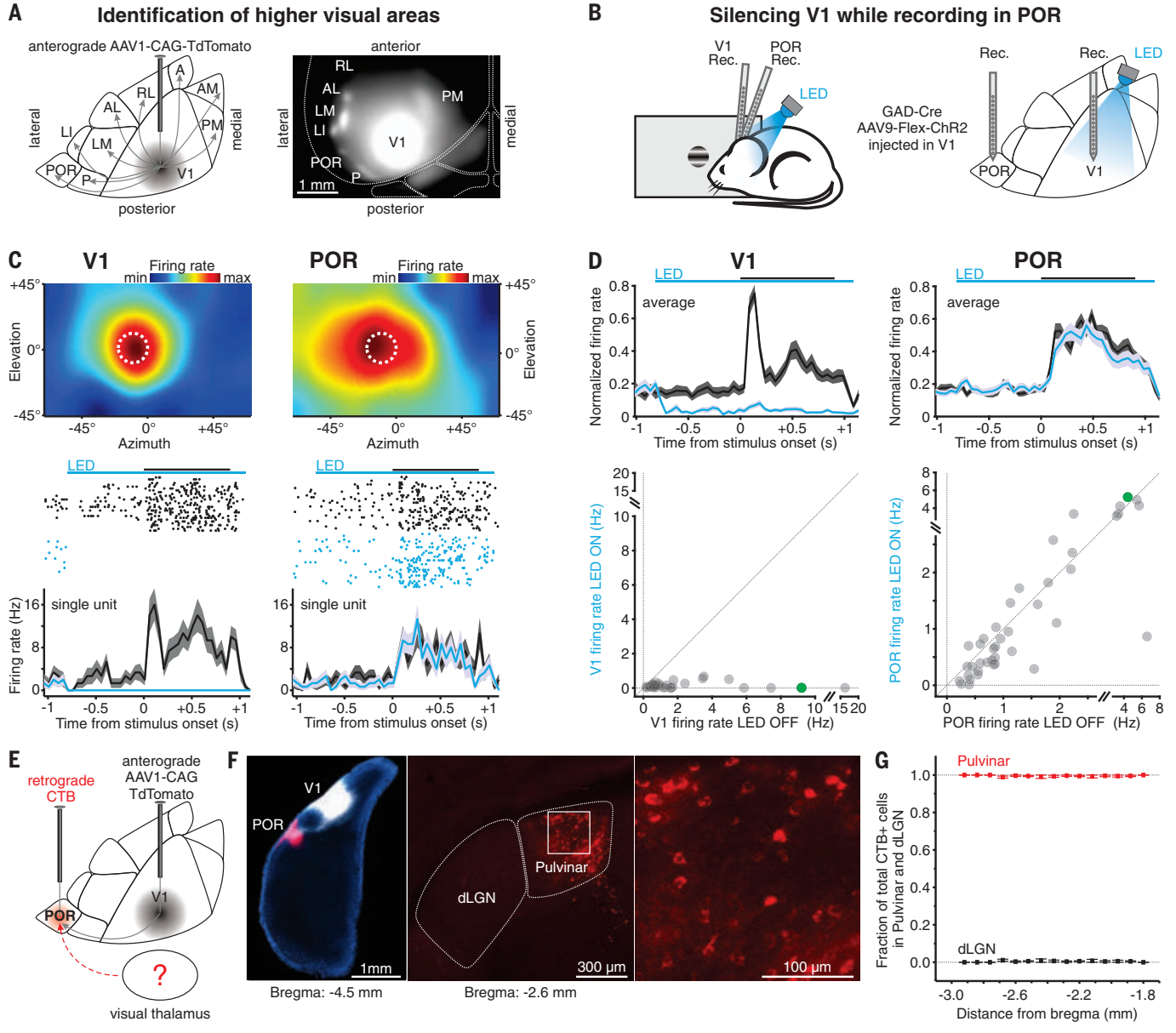


Fig. 1. Visual responses in POR are not driven by the geniculate-V1 pathway. (**A**) Injection of the anterograde tracer AAV1.TdTomato in V1 enabled the visualization of higher visual areas. Delineated cortical areas are as shown in fig. S2. The weaker fluorescence of POR is consistent with its weaker V1 input (*31*). (**B**) Drifting gratings (diameter, 20°) were presented to awake mice conditionally expressing ChR2 in V1 inhibitory neurons. Recordings were simultaneously performed in POR and in V1. A light-emitting diode (LED)–coupled optic fiber was used to silence V1. (**C**) Example experiment. (Left) V1 recordings; (right) POR recordings. (Top) Heatmaps of the receptive fields of multiunit activity. Dotted circle shows position of the visual stimulus. (Bottom) Raster plots and PSTH of RS units under control conditions (black) and during V1 silencing (blue). Black horizontal bar, duration of stimulus presentation; blue horizontal bar, period of V1 silencing. (**D**) Summary of all experiments as described for (C) (33 non-GABAergic RS units in V1 and 41 RS units in POR from five animals). (Top) Summary average PSTHs. The PSTH values for individual units were normalized by their maximum value and then averaged together. (Bottom) Scatter plots reporting the responses of all units measured during the stimulus presentation period (0.9 s). Green data points, example units in (C). (**E**) Injection of the retrograde tracer CTB in POR. (**F**) (Left) TdTomato injection site in V1 with anterograde projection to POR (white) and CTB injection in POR (red). (Middle) CTB+ neurons in dLGN and pulvinar nuclei. (Right) Higher-magnification image of the region delineated by the square in the middle image. (**G**) Distribution of retrogradely labeled dLGN and pulvinar neurons along the thalamic rostro-caudal axis (ratio of CTB+ cells in pulvinar or in dLGN to the total CTB+ cells counted in the two nuclei; 45 coronal sections, three mice).

the visually evoked responses recorded simultaneously in the caudal pulvinar (Fig. 2B) (80.02 ± 5.96% average decrease ± SEM in visually evoked firing rate; $P < 0.0001$, Wilcoxon signed-rank test, $n = 34$, 5 mice). Silencing of V1 had little effect on visually evoked activity in the caudal pulvinar (Fig. 2C) (24.64 ± 5.33% average decrease ± SEM in visually evoked firing rate; $P = 0.0004$, Wilcoxon signed-rank test, $n = 63$, 7 mice) while strongly reducing visual responses in the rostral pulvinar (Fig. 2D) (91.43 ± 4.82% average decrease ± SEM in visually evoked firing rate; $P = 0.0005$, Wilcoxon signed-rank test, $n = 16$, 3 mice).

Could the SC be disynaptically connected to POR through the pulvinar? We used anterograde

trans-synaptic tracing in which transfection of virus harboring Cre recombinase in the presynaptic neuronal population leads to the conditional expression of a reporter gene in the postsynaptic neuronal population (*22*) (Fig. 3, A and B, and fig. S8). We injected this virus in the SC and a virus containing conditional green fluorescent protein (GFP) in the caudal pulvinar. Histological analyses revealed cell bodies expressing GFP in the caudal pulvinar and axonal arborizations densely innervating layers 4 and 1 in POR as well as other cortical visual areas, albeit with the exception of the laterointermediate area (LI), to a much lesser extent than POR (Fig. 3, A and B, and fig. S8), consistent with recent observations (*20*).

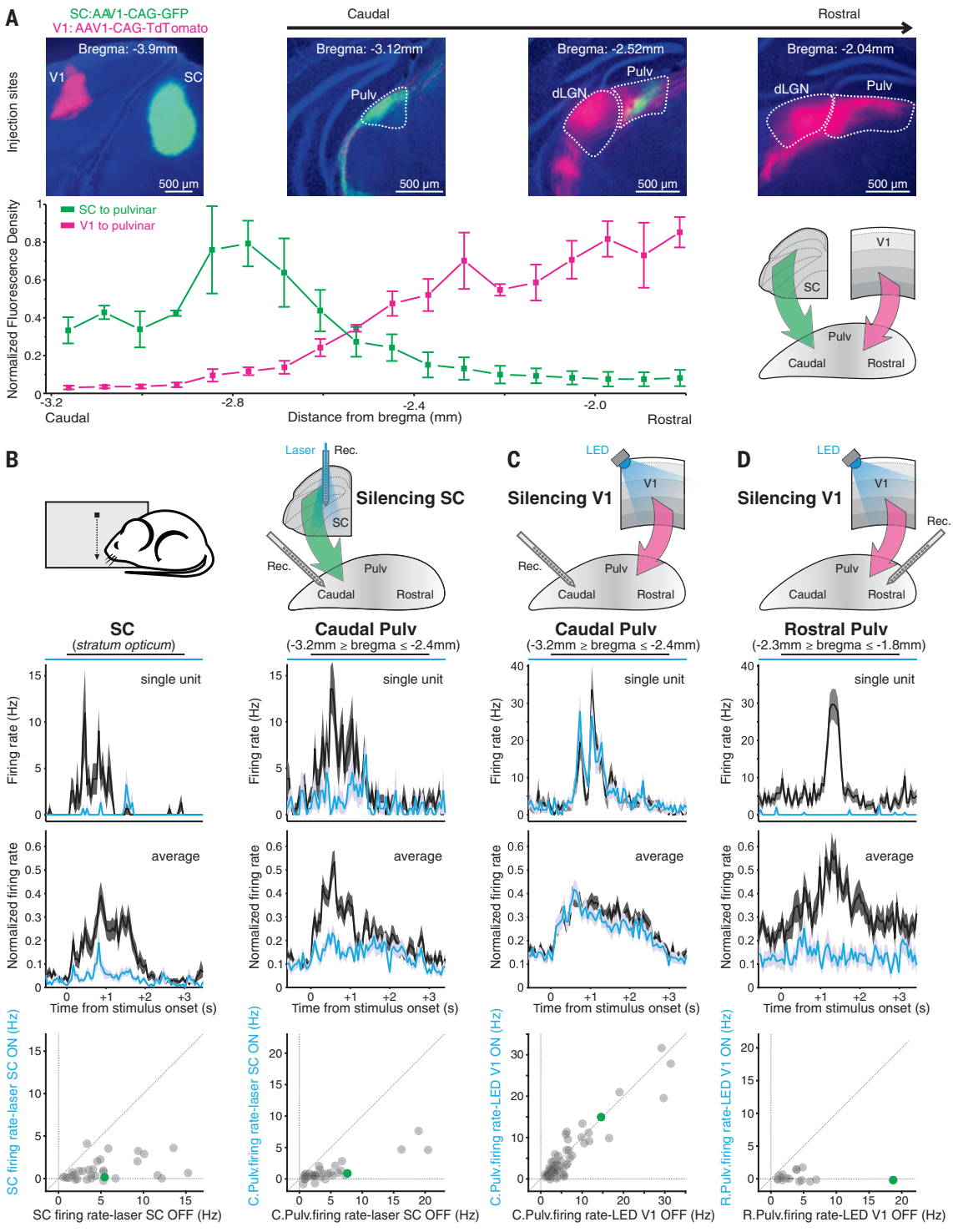


Fig. 2. Anatomical and functional segregation of cortical and collicular inputs in the pulvinar. (**A**) (Top) Coronal sections showing the injection sites of AAV1.CAG.TdTomato in V1 and AAV1.CAG.GFP in SC (left) and of V1 and SC projections to dLGN and pulvinar (Pulv) for three representative sections along the rostro-caudal axis. (Bottom left) Normalized fluorescent density (average ± SEM) along rostro-caudal axis of the pulvinar (green, SC input; magenta, V1 input; 54 coronal sections, three mice). Fluorescence density values were normalized to those for the coronal section displaying maximal tdTomato or GFP expression. (Bottom right) Schematic illustration of the results. (**B**) (Top row) The stimulus was a dark dot moving along a straight trajectory. Recordings were performed simultaneously from stratum opticum of SC (with an optrode for silencing) and caudal pulvinar. (Second row) (Left) PSTH of an SC unit. Black, control; blue, SC silencing; black horizontal bar, period of stimulus presentation; blue horizontal bar, period of SC silencing. (Right) PSTH of a simultaneously recorded unit in the caudal pulvinar. (Third row) Average PSTH for 33 (left) and 34 (right) isolated units in SC and in caudal pulvinar (five animals). (Bottom row) Scatter plots of firing rates averaged over a 450-ms interval (i.e., the response window; see materials and methods for details on the moving dot analysis) during the period of visual stimulation. Green data points, example units shown in the second row. (**C**) (Top row) V1 silencing and simultaneous recordings in caudal pulvinar. (Second row) PSTH of a unit in caudal pulvinar. Black, control; blue, V1 silencing. (Third row) Average PSTH for 63 isolated units (seven animals). (Bottom row) Scatter plot as in (B). (**D**) Similar experiment as shown in (C) but recording from rostral pulvinar (16 units; three animals).

To determine whether visually evoked activity in POR depends on SC, we performed simultaneous recordings from POR and SC while optogenetically silencing SC (Fig. 3, C to F). We presented dark moving dots and ensured that the receptive fields at the recording sites in POR and in the superficial layers of SC overlapped (see materials and methods and figs. S1B and S9). SC silencing almost completely abolished the response of POR to the dots (Fig. 3D) (93.77 ± 1.95% average decrease ± SEM in visually evoked firing rate; $P < 0.0001$, Wilcoxon signed-rank test, $n = 96, 5$ mice), across cortical depths in a homogeneous manner (fig. S10). The reduction of POR responses to dots was particularly pronounced for the portion of

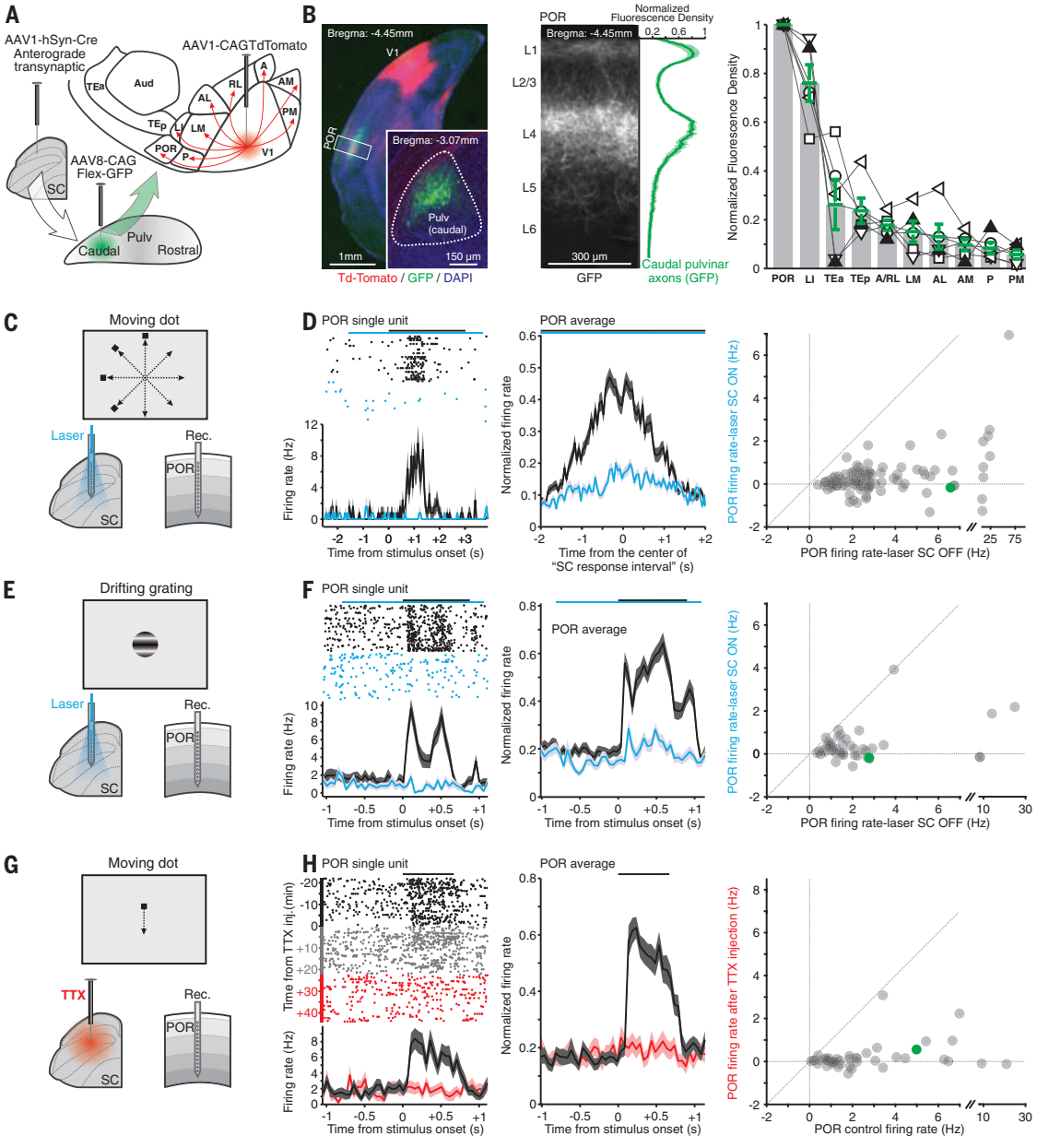


Fig. 3. POR is driven by the colliculo-cortical pathway. (**A**) Anterograde trans-synaptic tracing. AAV1.hSyn.Cre was injected in SC and AAV8.CAG.Flex.GFP was injected in caudal pulvinar. AAV1.CAG.TdTomato injected in V1 was used to identify higher visual areas. Delineated cortical areas are as shown in fig. S8. (**B**) (Left) Double labeling of axons targeting POR. Red, V1 afferents (for POR localization); green, caudal pulvinar afferents. (Inset) GFP expression in the caudal pulvinar. (Middle) Magnification of the region marked by the rectangle in the left panel (GFP channel) and summary distribution of GFP fluorescence (caudal pulvinar afferents) along POR cortical depth (five animals). (Right) Fluorescence density distribution of trans-synaptically labeled caudal pulvinar axons across cortical areas. Green bars, normalized averages and SEM (five animals) across 10 distinct cortical areas. For each animal, the fluorescence density of each area was normalized to that for POR. Each symbol corresponds to a different animal. Black triangle, experiment on the left. (**C**) SC silencing and POR recording. The visual stimulus was a dark dot moving at a speed of 30° per second along the two cardinal and two oblique axes for 0.67 or 3 s (i.e., the time required to cover, at 30° per second, trajectories of 20° or 90° of visual space, respectively). (**D**) (Left) Raster plot and PSTH of a POR unit under control conditions (black) and during SC silencing (blue). (Middle) Aligned average PSTH for 96 isolated units (five animals). The average PSTH was generated by aligning the response of each isolated POR unit to the center of the "SC response interval" for the same stimulus (time 0 s) (see materials and methods for a description of the moving dot analysis; see also fig. S9). (Right) Scatter plot of firing rates of POR units, averaged over a 450-ms window (i.e., the response window; see materials and methods description of the moving dot analysis) within the SC response interval. Green data point, example on the left. (**E** and **F**) Results of experiments similar to those shown in (C) and (D) but with a circular drifting grating patch as the stimulus (diameter: 20°; 44 isolated units; two animals). The average PSTH was aligned to the visual stimulus onset. Scatter plot, neuronal responses measured during stimulation period (0.9 s), as in Fig. 1D. (**G**) SC silencing and POR recording. The visual stimulus was a dark dot moving at 30° per second along a straight trajectory of 20° of visual space (stimulus duration, 0.67s). The superficial layers of SC were silenced via local application of TTX. (**H**) (Left) Raster plot and PSTH of an isolated POR unit under control conditions (black) and ~20 min after the beginning of TTX infusion in SC (red). The gray portion of the raster plot indicates the initial ~20 min of TTX infusion. (Middle) Average PSTH, aligned to the visual stimulus onset, for 44 isolated units (three animals). (Right) Scatter plot as in (D).

the POR receptive field that overlapped with the receptive field silenced in SC. That was evident for dot trajectories covering a large fraction of visual space, where SC silencing created a "scotoma" in the response of POR neurons (fig. S9). To directly compare the effect of SC silencing with that of V1 silencing, we also presented drifting gratings (Fig. 3, E and F) identical to those used to measure V1 silencing (Fig. 1, C and D). Optogenetic silencing of SC suppressed the response of POR to drifting gratings (Fig. 3F), highlighting the much stronger impact of SC than V1 on visually evoked activity in POR (79.83 ± 4.46% average decrease ± SEM in visually evoked firing rate; $P < 0.0001$, Wilcoxon signed-rank test, $n = 44$, 2 mice).

Because SC sends sparse inhibitory GABAergic projections to the dLGN (*21*), our optogenetic activation of GABAergic projection neurons in SC might suppress dLGN neurons. The dLGN projects to other visual areas in addition to V1 (*16*) (although not POR) (Fig. 1E). If visual responses in POR depended on a geniculate input relayed via other visual areas, rather than on a collicular

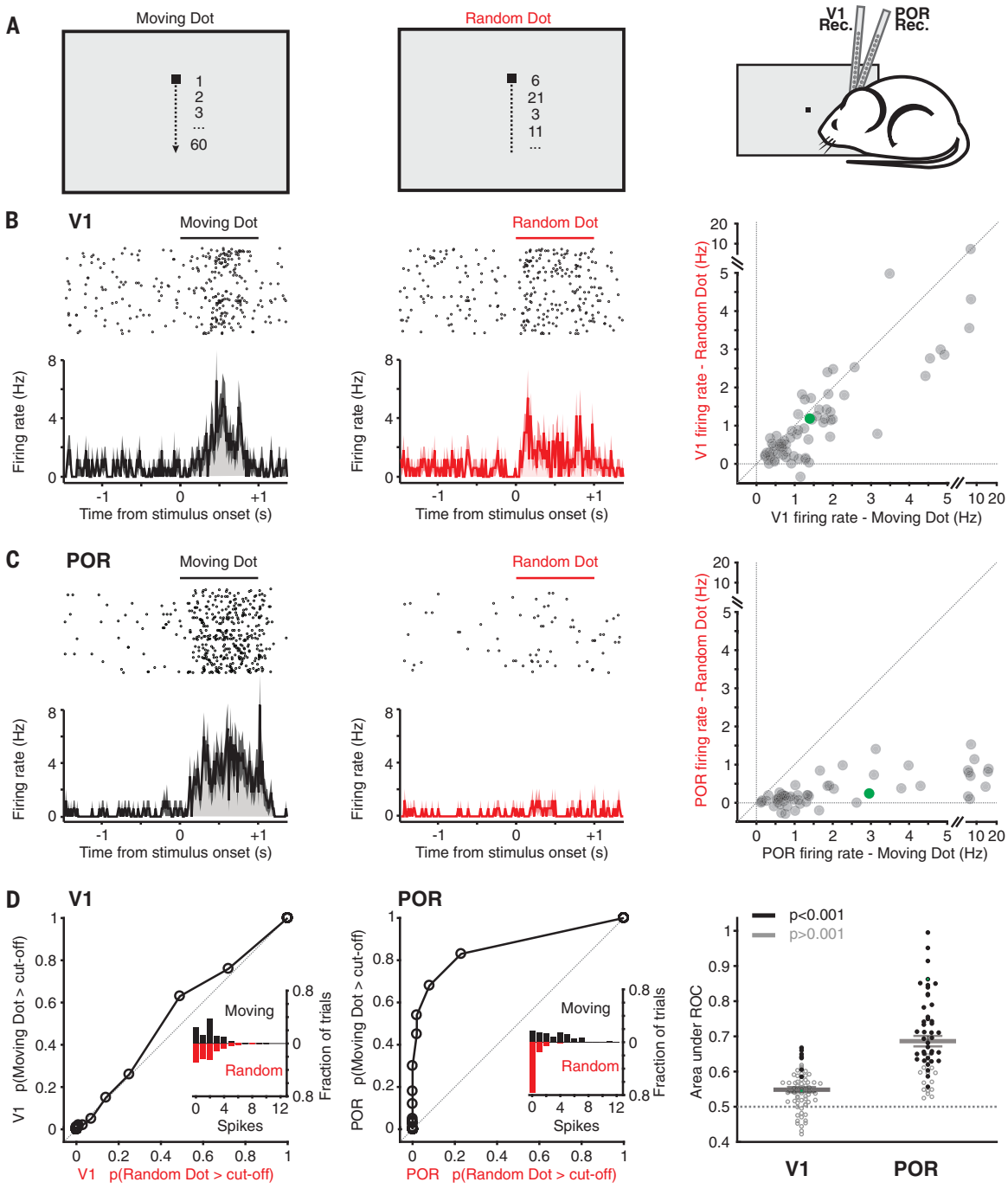


Fig. 4. POR neurons discriminate moving dots from random dots. (**A**) (Left) The stimulus was a dark dot moving for 1 s at a speed of 30° per second along a straight trajectory. (Middle) Random dot: The stimulus was made of the same frames as the moving dot but was presented in random order. Each iteration showed a different random sequence. (Right) Recordings were performed in V1 or POR. (**B**) Raster plot and PSTH of a V1 unit in response to a moving dot (left) or a random dot (middle), and the scatter plot (right) of average firing rates of V1 units during the period of stimulus presentation (1 s) for moving versus random dots (67 units, five animals). Green data point, example unit shown on the left. (**C**) As in (B) but for POR (56 units, five animals). Note that the shown POR unit did not respond to random dots. (**D**) (Left) ROC curve for the V1 isolated unit shown in (B). (Inset) Distribution of spikes for moving dots (black bars) and random dots (red bars). (Middle) ROC curve for the POR isolated unit shown in (C). (Right) Comparison of area under ROC curve (average ± SEM) for all responsive POR and V1 units. Black filled data points denote $P < 0.001$ for difference in spike distributions between responses to moving and random dots (Wilcoxon rank sum test). Gray bars show areas under ROC: for V1, 0.55 ± 0.007, 67 units, five mice; for POR: 0.69 ± 0.015, 56 units, five mice ($P < 0.0001$, Wilcoxon rank sum test). Green data points, example units shown in (B) and (C).

input, optogenetic activation of GABAergic SC neurons could reduce putative geniculate-mediated visual responses in POR. We therefore blocked neuronal activity in SC with tetrodotoxin (TTX) injected in the stratum opticum of SC (Fig. 3, G and H, and fig. S11). Again, we ensured that the receptive field of the TTX injection site matched the receptive field of the recording site in POR. TTX application abolished visually evoked responses in POR to both moving dots (Fig. 3H) (94.46 ± 3.18% average decrease ± SEM in visually evoked firing rate; $P < 0.0001$, $n = 44$, 3 mice) and drifting gratings (fig. S11) (92.46 ± 3.79% average decrease in visually evoked firing rate; $P < 0.0001$, Wilcoxon signed-rank test, $n = 39$, 3 mice). Baseline activity in POR was unaffected, indicating that the suppression of visual responses was not due to a direct action of TTX on POR.

Compared with V1, does POR capture distinct properties of the visual world? The ability to respond to small objects moving in the field of view is a characteristic property of the SC (*21*). We compared the response of V1 and POR to small moving objects. The response of a neuron to a dot moving on a monitor may report different aspects of the stimulus. In one case, the neuronal response may simply report local changes in luminance within the neuron's receptive field that occur as the dot moves along its trajectory. If so, the exact sequence of the changes in luminance along the trajectory may not be relevant for eliciting a response. In the other case, the neuron may selectively respond to the motion of the dot, i.e., to changes in luminance occurring sequentially at adjacent spatial positions. We therefore presented two different stimuli: a "moving dot" and a "random dot" (Fig. 4A). The moving dot stimulus consisted of a small dot shifting position incrementally along a linear trajectory through the receptive field of the neuron (the same type of stimulus as used in experiments described above). For the random dot stimulus, the dot positions along the same trajectory were randomized (Fig. 4A). Except for the temporal order of the dot positions, the two stimuli were identical. V1 neurons responded almost equally well to moving and to random dots (Fig. 4B) [25.19% (67 of 266) of the units isolated from V1 responded to moving and/or random dots; average firing rate ± SEM for moving dot: 1.62 ± 0.19 Hz; for random dot: 1.13 ± 0.16 Hz; $n = 67$; 5 mice]. For all responsive V1 neurons, the moving dot produced only a 1.4-fold-higher average firing response than the random dot. The response of POR neurons to the same stimuli, however, showed a marked preference for the moving dot over the random dot (Fig. 4C). For all responsive POR neurons, the moving dot produced a 9.2-fold-higher average firing response than the random dot [33.73% (56 of 166) of the units isolated from POR responded to either stimulus; average firing rate ± SEM for moving dot: 2.77 ± 0.5 Hz; for random dot: 0.3 ± 0.05 Hz; $P < 0.0001$, Wilcoxon signed-rank test, $n = 56$; 5 mice]. This was not because POR neurons responded much more to moving dots than did V1 neurons, but mainly because POR neurons responded much less to random dots than did V1 neurons. Receiver operator characteristics (ROC) analysis showed that an independent observer is better at discriminating moving from random dots based on the response of POR than the response of V1 neurons (Fig. 4D). Finally, we used moving dots to compare the sizes of receptive fields of POR and V1 neurons (fig. S12). Neurons in POR had significantly larger receptive fields than those in V1 (medians ± interquartile ranges of receptive field area; POR: 735.5 ± 631 square degrees, $n = 62$, 5 mice; V1: 157 ± 71.25 square degrees, $n = 31$, 8 mice; $P < 0.0001$, Wilcoxon rank sum test).

These results show that a cortical area, POR, is driven by visual information conveyed via the colliculo-cortical pathway rather than by the geniculate-V1 pathway. The weak impact of V1 silencing on POR is consistent with the weak projection of the former onto the latter area (*14*). The superior ability of POR to discriminate moving objects compared with that of V1 is in agreement with previous descriptions of enhanced motion selectivity in lateral cortical visual areas (*23*, *24*). However, though this selectivity to moving objects was believed to result from the hierarchical cortical processing of geniculate inputs entering V1 (*23*, *24*), our data indicate that this property of POR depends on collicular input and could be directly inherited from SC. Similarly, the expanded representation of the upper visual hemifield reported in POR (*24*, *25*) could also be directly inherited from the equally biased representation in SC (*26*). Because the colliculo-cortical pathway is believed to be phylogenetically older than the geniculate-V1 pathway (*27*), it is tempting to regard POR as an ancestral primary visual area.

Previous reports of SC lesions in cats (*12*) and rodents (*9*) described only a slight reduction, if any, of visually evoked activity in various visual cortical areas. Furthermore, in rodents, this reduction depends on the features of the stimulus (*9*). However, none of these recordings was performed in POR. In primates, SC silencing and lesions also have little or no impact on visual response of several visual cortical areas unless V1 has been ablated first (*11*). Although primate vision may rely less on SC than in other mammals, again, these recordings were not performed in the parahippocampal cortical areas TF and TH, the primate areas most similar to POR (*28*). On the other hand, the driving rather than modulatory impact of SC on the thalamus reported here is not without precedent. In primates, the SC drives the mediodorsal nucleus of the thalamus (*29*) to relay corollary oculomotor activity to the frontal eye fields. Primates can also partially recover from the inability to detect visual stimuli caused by V1 lesions, a phenomenon called blindsight and that involves SC (*30*). Whether some aspect of blindsight depends on the impact of SC on visual cortex remains to be established.

These results show a spatial and functional segregation of the sensory representation of the two main visual pathways to visual cortex, the geniculate-V1 and colliculo-cortical pathways, and define a specialized cortical area whose responses to visual stimuli are driven by the colliculo-cortical pathway.

REFERENCES AND NOTES

1. S. M. Sherman, R. W. Guillery, *Exploring the Thalamus and Its Role in Cortical Function*. (MIT Press, 2001).
2. D. J. Felleman, D. C. Van Essen, *Cereb. Cortex* **1**, 1–47 (1991).
3. P. H. Schiller, J. G. Malpeli, *Brain Res.* **126**, 366–369 (1977).
4. H. Sherk, *J. Neurophysiol.* **41**, 204–215 (1978).
5. P. Girard, P. A. Salin, J. Bullier, *J. Neurophysiol.* **66**, 1493–1503 (1991).
6. S. Molotchnikoff, F. Hubert, *Brain Res.* **510**, 223–228 (1990).
7. R. A. Berman, R. H. Wurtz, *J. Neurosci.* **30**, 6342–6354 (2010).
8. D. C. Lyon, J. J. Nassi, E. M. Callaway, *Neuron* **65**, 270–279 (2010).
9. M. Tohmi, R. Meguro, H. Tsukano, R. Hishida, K. Shibuki, *Curr. Biol.* **24**, 587–597 (2014).
10. N. A. Zhou, P. S. Maire, S. P. Masterson, M. E. Bickford, *Vis. Neurosci.* **34**, E011 (2017).
11. H. R. Rodman, C. G. Gross, T. D. Albright, *J. Neurosci.* **10**, 1154–1164 (1990).
12. T. Ogino, K. Ohtsuka, *Invest. Ophthalmol. Vis. Sci.* **41**, 955–960 (2000).
13. A. Zénon, R. J. Krauzlis, *Nature* **489**, 434–437 (2012).
14. Q. Wang, A. Burkhalter, *J. Comp. Neurol.* **502**, 339–357 (2007).
15. C. R. Burgess *et al*., *Neuron* **91**, 1154–1169 (2016).
16. L. L. Glickfeld, S. R. Olsen, *Annu. Rev. Vision Sci.* **3**, 251–273 (2017).
17. A. D. Lien, M. Scanziani, *Nat. Neurosci.* **16**, 1315–1323 (2013).
18. N. Zhou, S. P. Masterson, J. K. Damron, W. Guido, M. E. Bickford, *J. Neurosci.* **38**, 347–362 (2018).
19. D. B. Bender, *Brain Res.* **279**, 258–261 (1983).
20. C. Bennett, S. D. Gale, M. E. Garrett, M. L. Newton, E. M. Callaway, G. J. Murphy, S. R. Olsen, bioRxiv 395244 [Preprint]. 7 November 2018. https://doi.org/10.1101/395244.
21. S. D. Gale, G. J. Murphy, *J. Neurosci.* **34**, 13458–13471 (2014).
22. B. Zingg *et al*., *Neuron* **93**, 33–47 (2017).
23. B. Vermaercke *et al*., *J. Neurophysiol.* **112**, 1963–1983 (2014).
24. N. Nishio *et al*., *Sci. Rep.* **8**, 11136 (2018).
25. M. E. Garrett, I. Nauhaus, J. H. Marshel, E. M. Callaway, *J. Neurosci.* **34**, 12587–12600 (2014).
26. U. C. Dräger, D. H. Hubel, *J. Neurophysiol.* **39**, 91–101 (1976).
27. I. T. Diamond, W. C. Hall, *Science* **164**, 251–262 (1969).
28. S. C. Furtak, O. J. Ahmed, R. D. Burwell, *Neuron* **76**, 976–988 (2012).
29. R. H. Wurtz, M. A. Sommer, J. Cavanaugh, *Prog. Brain Res.* **149**, 207–225 (2005).
30. D. A. Leopold, *Annu. Rev. Neurosci.* **35**, 91–109 (2012).

ACKNOWLEDGMENTS

We thank all the members of the Scanziani lab for discussions about the project and comments on the manuscript; H. Karten for advice during the course of the study; L. Frank, R. Nicoll, and E. Feinberg for critical reading of the manuscript; and M. Mukundan, J. Evora, N. Kim, and Y. Li for technical support. **Funding:** This project was supported by the European Molecular Biology Organization Long Term Fellowship, the Human Frontier Science Program Long Term Fellowship, and the Howard Hughes Medical Institute. **Author contributions:** R.B. and M.S. designed the study. R.B. conducted all experiments and analyses. R.B. and M.S. wrote the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data and analyses necessary to understand and assess the conclusions of the manuscript are presented in the main text and in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/363/6422/64/suppl/DC1
Materials and Methods
Figs. S1 to S12
References (*31*–*41*)

3 August 2018; accepted 21 November 2018
10.1126/science.aau7052

CLIMATE CHANGE

The Little Ice Age and 20th-century deep Pacific cooling

G. Gebbie[1]* and P. Huybers[2]

Proxy records show that before the onset of modern anthropogenic warming, globally coherent cooling occurred from the Medieval Warm Period to the Little Ice Age. The long memory of the ocean suggests that these historical surface anomalies are associated with ongoing deep-ocean temperature adjustments. Combining an ocean model with modern and paleoceanographic data leads to a prediction that the deep Pacific is still adjusting to the cooling going into the Little Ice Age, whereas temperature trends in the surface ocean and deep Atlantic reflect modern warming. This prediction is corroborated by temperature changes identified between the HMS Challenger expedition of the 1870s and modern hydrography. The implied heat loss in the deep ocean since 1750 CE offsets one-fourth of the global heat gain in the upper ocean.

Downcore temperature profiles found in boreholes from the Greenland (*1*) and West Antarctic ice sheets (*2*) enable the recovery of past surface temperatures. These borehole inversions indicate a globally coherent pattern of cooling from the Medieval Warm Period to the Little Ice Age that is also documented in recent land (*3*) and ocean (*4*) proxy compilations. The ocean adjusts to surface temperature anomalies over time scales greater than 1000 years in the deep Pacific (*5*, *6*), which suggests that it too hosts signals related to Common Era changes in surface climate (*7*). But whether these signals are predictable or detectable in the face of three-dimensional ocean circulation and mixing processes, let alone invertible for surface characteristics, has been unclear.

To explore how Common Era changes in surface temperature could influence the interior ocean, we first inverted modern-day tracer observations for ocean circulation using a previously described methodology (*8*). In this inversion, the net effects of sub–grid-scale processes on advective and diffusive transport are empirically constrained at a 2° resolution in the horizontal and 33 levels in the vertical. When integrated with prescribed surface values, the estimated circulation gives accurate predictions of interior $\delta^{13}C$ (*9*) and radiocarbon values (*6*). The relative influences of Antarctic Bottom Water and North Atlantic Deep Water are also captured (*8*) and agree with estimates made using related approaches (*10*).

It is also possible to represent the transient oceanic response to changing surface conditions. A 2000-year simulation is performed by initializing our empirical circulation model at equilibrium in 15 CE and prescribing globally coherent surface temperature anomalies (*4*) that propagate into the ocean interior (see supplementary materials). The resulting estimate, referred to as EQ-0015, indicates that disparate modern-day temperature trends are expected at depth (Fig. 1). At depths below 2000 m, the Atlantic warms at an average rate of 0.1°C over the past century, whereas the deep Pacific cools by 0.02°C over the past century.

The pattern of temperature trends can be understood as a basic consequence of an advective-diffusive adjustment to surface conditions. Deep Atlantic waters are directly replenished by their formation in the North Atlantic, but deep Pacific waters must propagate from the Atlantic and Southern oceans. Radiocarbon observations (*11*) indicate that most waters in the deep Atlantic were last at the surface 1 to 4 centuries ago, whereas most deep Pacific waters have longer memory due to isolation from the atmosphere for 8 to 14 centuries (*6*). As a result of differing response times, Atlantic temperature trends reflect warming over recent centuries, including that associated with anthropogenic influences, whereas the Pacific is still cooling as a consequence of ongoing replacement of Medieval Warm Period waters by Little Ice Age waters.

The simulated magnitude of temperature changes also reflects an advective-diffusive response to surface conditions. EQ-0015 indicates deep-Pacific cooling of 0.1°C following the temperature maximum associated with the Medieval Warm Period, whereas the faster-responding deep Atlantic cools by as much as 0.3°C. Ocean circulation can be likened to a filter through which interior water properties inherit a temporally smoothed version of surface signals. Signals in the deep Pacific are more heavily smoothed and have a more attenuated signal than their Atlantic counterparts because they are subject to mixing over a longer journey (*12*). The incomplete response of the subsurface to rapid surface changes also leads to delays seen in EQ-0015 being shorter than those indicated by radiocarbon-age analysis (*13*).

Implicit in the EQ-0015 simulation is that temperature anomalies are transported according to a statistically steady ocean circulation. Estimates of circulation strength over the Common Era, however, suggest variations by as much as ±25% for components of the Atlantic circulation (*14*, *15*). If we instead modify circulation rates to covary with surface temperature anomalies such that advective and diffusive fluxes are changed by ±25% in the Little Ice Age relative to the 1990s, the magnitude of our results is altered (fig. S3), but not the qualitative pattern. In a general circulation model not subject to such simplified assumptions, the centennial-scale subsurface temperature response is also well approximated by the transport of an unchanging circulation (*16*). Of course, it cannot be excluded that changes in deep circulation—for example, in response to altered deep water formation rates or winds (*17*)—counteract the basic pattern of temperature response expected from modern circulation. The results of EQ-0015 are thus considered a prediction that requires further testing.

Differences in the simulated timing and magnitude of temperature trends between the Atlantic and Pacific offer a fingerprint of historical changes in surface temperature. To compare this fingerprint against observations, we turn to the deep-ocean temperature measurements from the HMS Challenger expedition that were obtained near the beginning of the instrumental era, 1872–1876 CE. There were 5010 temperature observations along the cruise track, including 4081 observations below the mixed layer and 760 observations from deeper than 2000 m (Fig. 2). Previous analysis (*18*) showed a 0.4°C warming between the 1870s and 2000s in the upper 500 m of the ocean, tapering off to values indistinguishable from zero at 1800 m depth. Challenger temperature trends were not assessed at deeper levels, however, over concerns regarding depth-dependent biases.

Our focus is to test the model prediction of deep-Pacific cooling. Therefore, we guard against observational biases that would predispose results toward such a trend. In particular, we adjust Challenger temperatures to be 0.04°C cooler per kilometer of depth in keeping with a previously used correction for the effects of compression (*18*, *19*). Another concern is that the rope used for measurements may not have paid out entirely in the vertical, causing depths to be overestimated. But comparing Challenger reports of ocean depth against modern bathymetry (*20*) indicates that, if anything, depths are underestimated, possibly because the hemp rope used aboard the Challenger stretched (fig. S4). We apply no further depth corrections because underestimates would only bias our analysis toward showing greater warming. The exception is in the Southern Ocean, where strong currents are expected to cause greater horizontal deflection of the line (*18*); data south of 45°S are therefore excluded. Finally, the max-min thermometer used on the Challenger would have been biased in regions with vertical temperature inversions. To mitigate the influence of such reversals, we also exclude the 164 data points that are located in temperature inversions in modern climatology (*21*), leaving a total of 3212 observations.

The most recent top-to-bottom global assessment of ocean temperatures comes from the World Ocean Circulation Experiment (WOCE) campaign of the 1990s. Interpolating WOCE

[1]Department of Physical Oceanography, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA. [2]Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA 02139, USA.
*Corresponding author. Email: ggebbie@whoi.edu

observations (*21*) to the Challenger data locations permits for comparison of temperatures across more than a century. The squared cross-correlation between WOCE and Challenger temperatures is 0.97 and remains high at 0.92 after removing a global-mean vertical profile from each individual profile. Comparison of other 20th-century hydrographic data also indicated only minor density perturbations on the basic oceanic structure (*22*). Similarity of the oceanic temperature and density structure over time supports the interpretation of changes in circulation since the Little Ice Age as involving only minor perturbations.

Despite overall consistency, there are systematic differences between WOCE and Challenger temperatures. The upper 1000 m of the ocean hosts pervasive warming (Fig. 3), as found earlier (*18*). Basin-wide warming is also found to 2800 m depth in the Atlantic and is significant at the 95% confidence level. Significance levels are computed accounting for the effects of high-frequency motions incurred by internal waves, mesoscale eddies, and wind variability (see supplementary materials). In the deep Pacific, we find basin-wide cooling ranging from 0.02° to 0.08°C at depths between 1600 and 2800 m (Fig. 3) that is also statistically significant. The basic pattern of Atlantic warming and deep-Pacific cooling diagnosed from the observations is consistent with our model results, although the observations indicate stronger cooling trends in the Pacific. Note that the difference between Atlantic and Pacific trends is particularly diagnostic because it is insensitive to choices regarding depth-dependent bias corrections.

The bulk of the Challenger observations that indicate 20th-century cooling are found in the Pacific between 2000 and 4000 m depth. We estimate the integrated rate of heat loss in this Pacific layer to be 1 TW. Although a warming trend was identified in repeat hydrographic observations available over recent decades for the abyssal ocean below 4000 m (*23*), trend estimates specifically for the deep Pacific between 2000 and 4000 m depth were found to be insignificant at 6 ± 7 TW (5 to 95% confidence interval) over the period 1991–2010 (*24*). Reanalysis products augment the hydrographic data with other observational and numerical model information, but no consensus on the sign of deep-Pacific temperature trends has emerged amongst these estimates (*25*). Some reanalyses do, however, show a pattern of Atlantic warming and deep-Pacific cooling that is congruent with our findings (*26*, *27*) (see supplementary materials). Whereas it was suggested that this deep-Pacific cooling in reanalyses originates from model initialization artifacts and weak data constraints (*25*), our results indicate that such temperature drifts should be expected on physical grounds. We also emphasize that there is a major caveat in all these comparisons, in that rate estimates may be sensitive to decadal variability and the time periods over which trends are computed (*7*).

The EQ-0015 simulation is independent of the Challenger observations, and these two indications of deep-ocean temperature trends can be

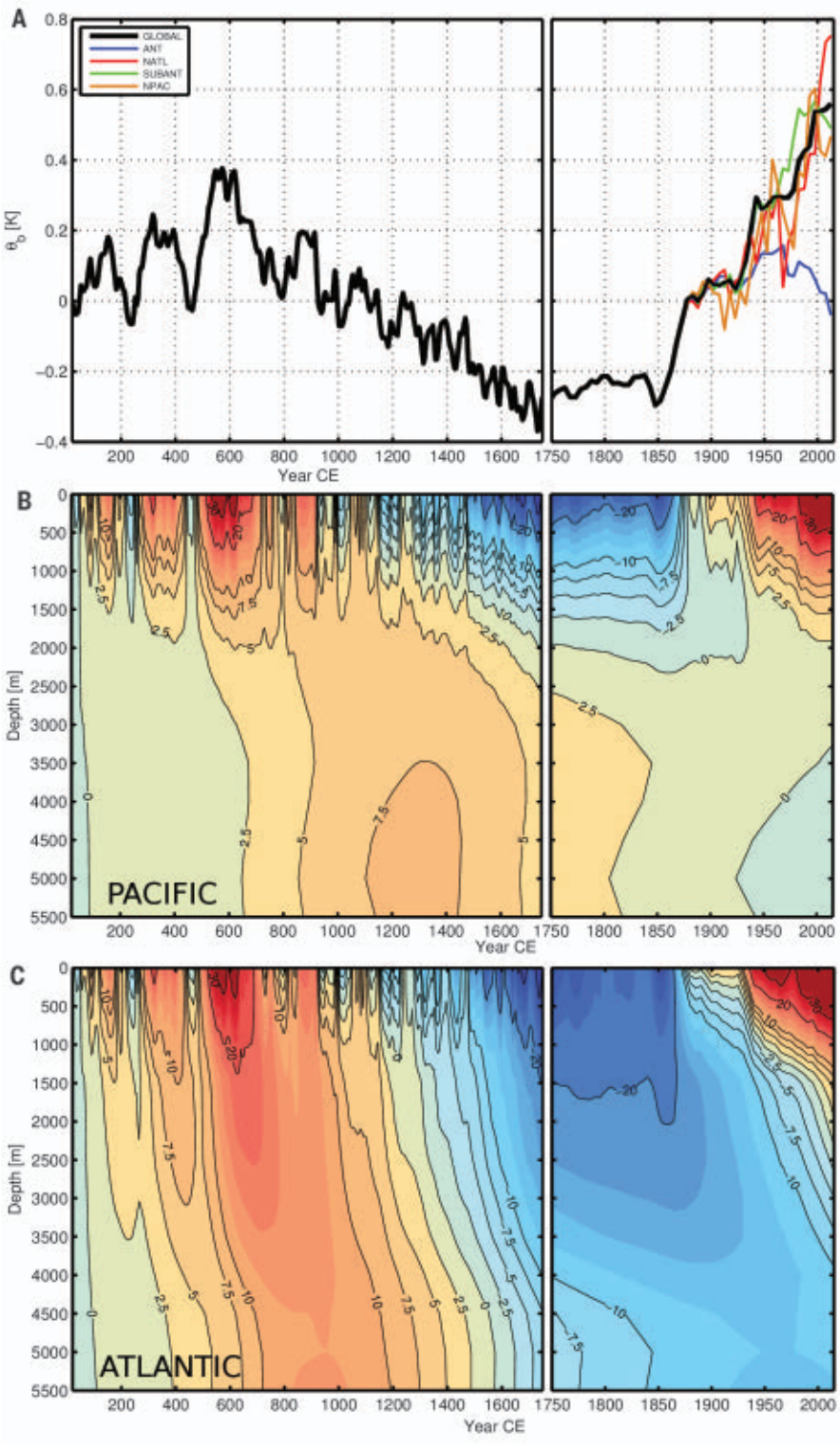


Fig. 1. Simulated interior ocean response to Common Era surface temperature anomalies. (**A**) Global average (black line) and regionally averaged (colored lines) surface temperature time series θ_b, for a simulation initialized from equilibrium in 15 CE (EQ-0015). Regional variations are plotted for the Antarctic (ANT), North Atlantic (NATL), sub-Antarctic (SUBANT), and North Pacific (NPAC). Prior to globally available instrumental surface temperatures beginning in 1870 CE, global changes are prescribed according to estimates from paleoclimate data. (**B**) Time evolution of the Pacific-average potential temperature profile from EQ-0015. (**C**) Similar to (B) but for the Atlantic-average profile. Atlantic and Pacific averages are taken north of 35°S and 45°S, respectively, and color shading has a 2.5-cK interval from −35 to 35 cK. Note the expanded time axis after 1750 CE.

combined to give a more detailed estimate. We first average Challenger-to-WOCE temperature trends over the Atlantic and Pacific basins as a function of depth. These basin-wide average trends are used to relax the assumption of globally uniform changes in surface conditions and to constrain regional temperature histories for 14 distinct regions over the Common Era by a control theory method (see supplementary materials). The result, referred to as OPT-0015, fits the observed vertical structure of Pacific cooling and Atlantic warming (Fig. 3). Global surface changes still explain the basic Atlantic-Pacific difference in OPT-0015, but greater Southern Ocean cooling between 600 and 1600 CE leads to greater rates of cooling in the deep Pacific over recent centuries. Regionally inferred variations in North Atlantic and sub-Antarctic surface temperatures also reproduce an Atlantic warming minimum at 800 m. Because OPT-0015 is constrained using only basin-wide averages, regional temperature patterns can be independently compared against observations. Notable in this regard is that OPT-0015 produces greater rates of cooling in the deep North Pacific and greater warming in the vicinity of the Atlantic deep western boundary current. Similar patterns are evident in the Challenger observations (fig. S7) as well as the average across multiple ocean reanalyses (*25*).

Regional surface temperatures in OPT-0015 can also be compared against ice-core borehole inversions. OPT-0015 places the coldest Antarctic conditions in the 1500s and the coldest North Atlantic in the 1800s, both of which are amplified relative to the global average (Fig. 4). This interhemispheric sequence of peak cooling aligns with the minimum surface temperatures estimated from boreholes in Antarctica (*2*) and Greenland (*1*). A second, weaker cool interval inferred from Greenland boreholes between 1400 and 1600 CE (*1*) is, however, not found for the North Atlantic in OPT-0015. The inference of amplified temperature anomalies in the Antarctic and North Atlantic oceans is also consistent with stronger positive feedbacks at high latitudes. Amplification of high-latitude signals could also stem from greater winter than summer cooling during the Little Ice Age (*28*) and from the greater sensitivity of deep-water formation to winter conditions (*29*). The combination of greater volatility in winter surface conditions and greater sensitivity of interior waters to these conditions may explain observations of amplified mid-depth temperature variability relative to the surface over the Holocene (*30*, *31*).

The OPT-0015 results provide an estimate of full-ocean changes in heat content over the Common Era. With regard to changes in heat content in the upper 700 m of the ocean (Fig. 4), there is excellent consistency between OPT-0015 and results from observational analyses (*32*) and model simulations (*33*), each indicating ~170 ZJ (1 ZJ = 10^{21} J) of heat uptake between 1970 and 2010 (Fig. 4). Over a longer period, 1875–2005, OPT-0015 gives 330 ZJ of global upper-ocean heat uptake, equal to the central estimate from an earlier analysis of upper-ocean heating using

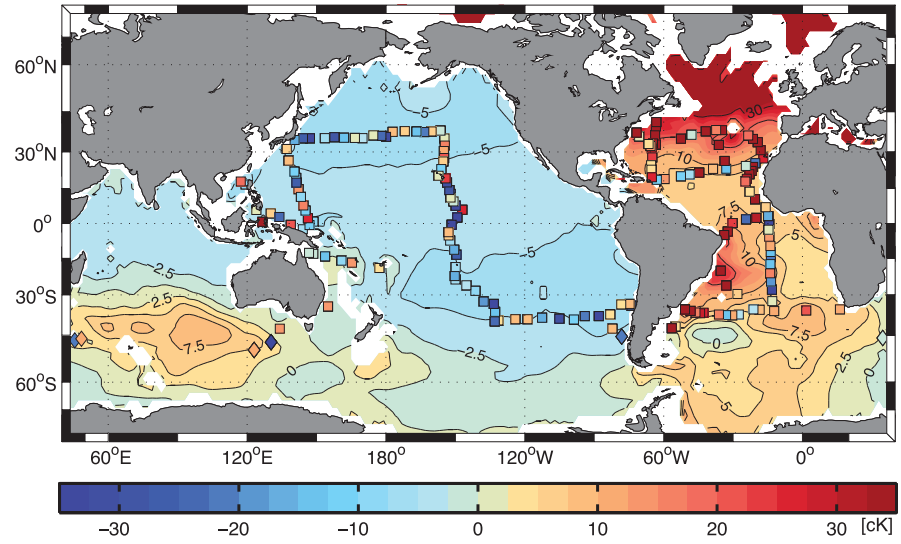


Fig. 2. Observed and simulated deep-ocean temperature changes. Observed ocean temperature changes are diagnosed by differencing WOCE and Challenger temperature measurements. WOCE temperatures are linearly interpolated to the location of Challenger temperatures, and differences are plotted after averaging between 1800 and 2600 m depth (colored markers). Simulated temperature changes for the same depth interval are diagnosed from OPT-0015. Color scaling is equivalent for observed and simulated temperature changes.

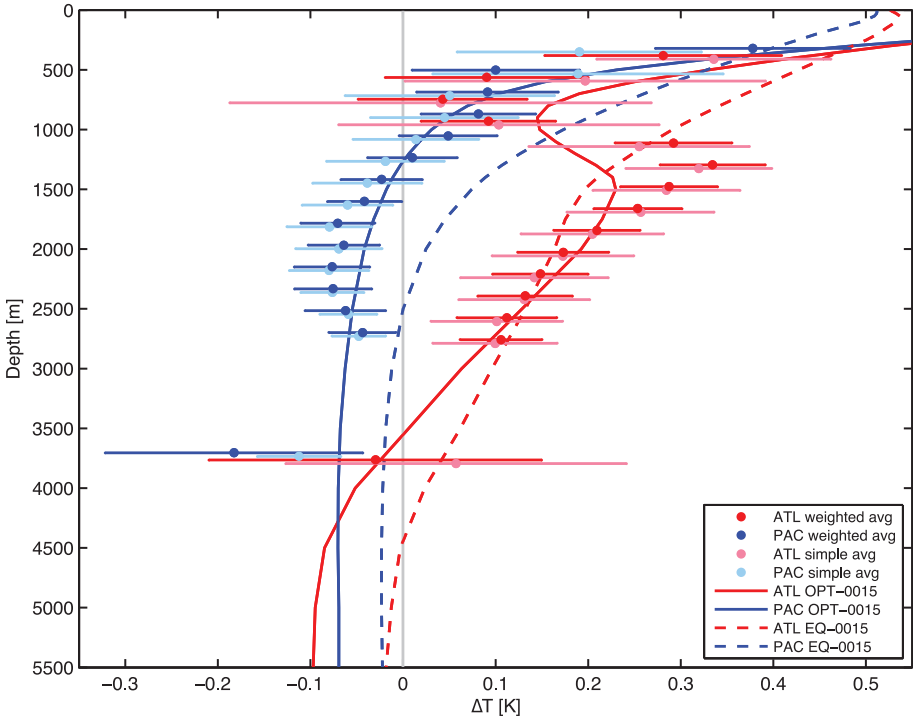


Fig. 3. Vertical profiles of temperature change. Difference between WOCE and Challenger temperatures is shown as a function of depth with 95% confidence intervals averaged over the Pacific (blue) and Atlantic (red). Features of the WOCE-Challenger temperature difference are reproduced in a simulation initialized from equilibrium at 15 CE (EQ-0015, dashed curves) and an inversion constrained by the observations (OPT-0015, solid curves). WOCE-Challenger temperature differences are calculated using a weighted average that accounts for the covariance of ocean temperatures and their uncertainties based on the expected effects of high-frequency oceanic variability (markers and error bars with darker colors). For comparison, a simple average for each basin and depth level is also shown with uncertainties that are empirically estimated (lighter colors).

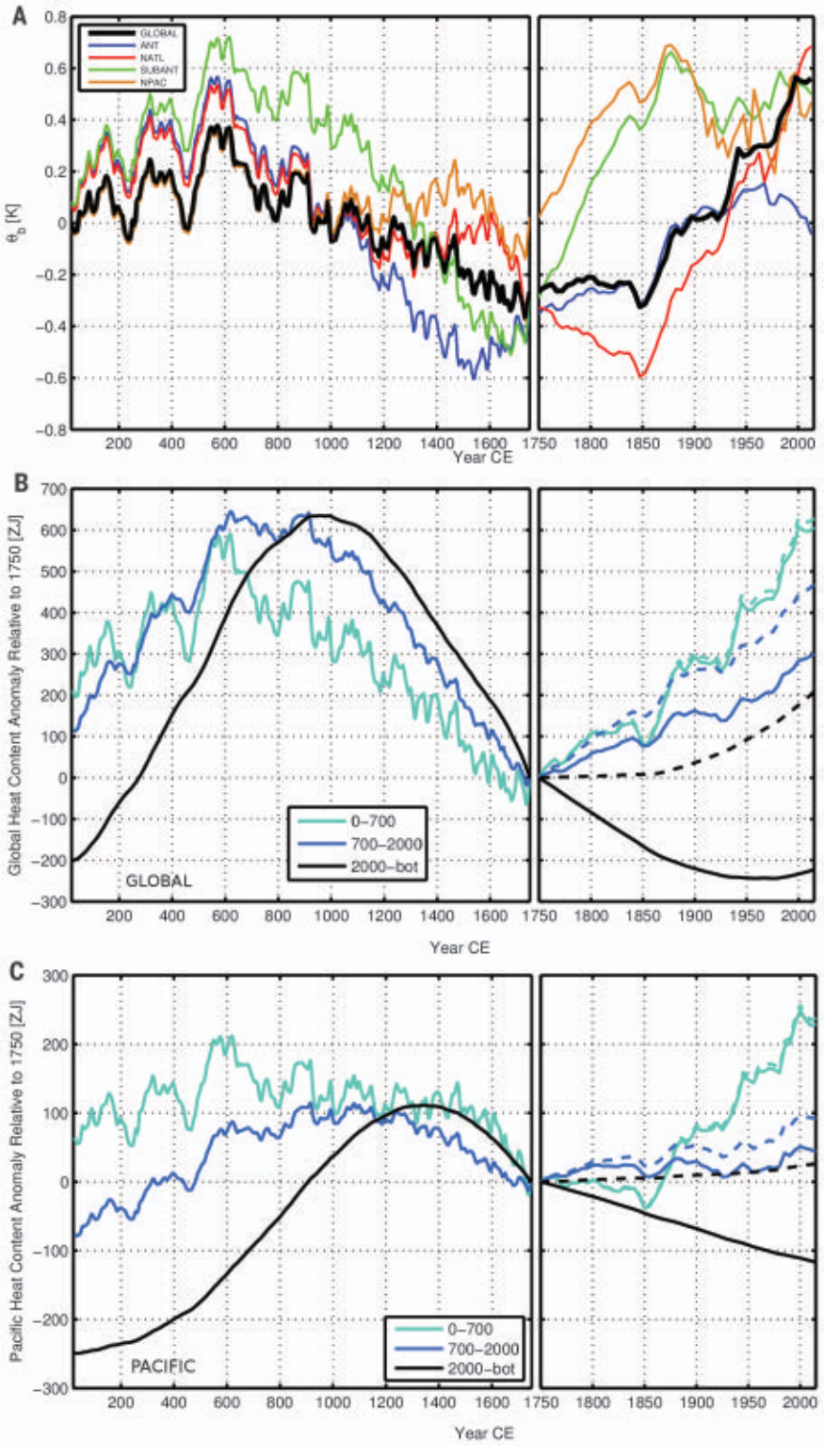


Fig. 4. Regional surface temperature variations and changes in ocean heat content over the Common Era. (**A**) Surface temperature time series after adjustment to fit the HMS Challenger observations (OPT-0015), including four major surface regions (colored lines) and the global area-weighted average (black line). (**B**) Time series of global oceanic heat content anomalies relative to 1750 CE from OPT-0015 as decomposed into upper (cyan, 0 to 700 m), mid-depth (blue, 700 to 2000 m), and deep (black, 2000 m to the bottom) layers. Heat content anomalies calculated from an equilibrium simulation initialized at 1750 (EQ-1750, dashed lines) diverge from the OPT-0015 solution in deeper layers. (**C**) Similar to (B) but for the Pacific. Heat content anomaly is in units of zettajoules (1 ZJ = 10^{21} J).

Challenger observations (*18*). More generally, OPT-0015 indicates that the upper 2000 m of the ocean has been gaining heat since the 1700s, but that one-fourth of this heat uptake was mined from the deeper ocean. This upper-lower distinction is most pronounced in the Pacific since 1750, where cooling below 2000 m offsets more than one-third of the heat gain above 2000 m.

The implications of the deep Pacific being in disequilibrium become more apparent when compared to a counterfactual scenario where the ocean is fully equilibrated with surface conditions in 1750 CE. That the deep Pacific gains heat in this scenario, referred to as EQ-1750, confirms that heat loss in OPT-0015 results from the cooling associated with entry into the Little Ice Age. Moreover, the EQ-1750 scenario leads to 85% greater global ocean heat uptake since 1750 because of excess warming below 700 m. It follows that historical model simulations are biased toward overestimating ocean heat uptake when initialized at equilibrium during the Little Ice Age, although additional biases are also likely to be present (*34*). Finally, we note that OPT-0015 indicates that ocean heat content was larger during the Medieval Warm Period than at present, not because surface temperature was greater, but because the deep ocean had a longer time to adjust to surface anomalies. Over multicentennial time scales, changes in upper and deep ocean heat content have similar ranges, underscoring how the deep ocean ultimately plays a leading role in the planetary heat budget.

REFERENCES AND NOTES

1. D. Dahl-Jensen *et al.*, *Science* **282**, 268–271 (1998).
2. A. J. Orsi, B. D. Cornuelle, J. P. Severinghaus, *Geophys. Res. Lett.* **39**, L09710 (2012).
3. PAGES 2k Consortium, *Nat. Geosci.* **6**, 339–346 (2013).
4. H. V. McGregor *et al.*, *Nat. Geosci.* **8**, 671–677 (2015).
5. F. Primeau, *J. Phys. Oceanogr.* **35**, 545–564 (2005).
6. G. Gebbie, P. Huybers, *J. Phys. Oceanogr.* **42**, 291–305 (2012).
7. C. Wunsch, P. Heimbach, *J. Phys. Oceanogr.* **44**, 2013–2030 (2014).
8. G. Gebbie, P. Huybers, *J. Phys. Oceanogr.* **40**, 1710–1728 (2010).
9. G. Gebbie, P. Huybers, *Geophys. Res. Lett.* **38**, L06604 (2011).
10. T. DeVries, F. Primeau, *J. Phys. Oceanogr.* **41**, 2381–2401 (2011).
11. R. M. Key *et al.*, *Global Biogeochem. Cycles* **18**, GB4031 (2004).
12. M. Holzer, F. Primeau, *J. Geophys. Res.* **115**, C12021 (2010).
13. É. Delhez, É. Deleersnijder, *Cont. Shelf Res.* **28**, 1057–1067 (2008).
14. D. C. Lund, J. Lynch-Stieglitz, W. B. Curry, *Nature* **444**, 601–604 (2006).
15. S. Rahmstorf *et al.*, *Nat. Clim. Chang.* **5**, 475–480 (2015).
16. J. Marshall *et al.*, *Clim. Dyn.* **44**, 2287–2299 (2015).
17. M. Kawase, *J. Phys. Oceanogr.* **17**, 2294–2317 (1987).
18. D. Roemmich, W. J. Gould, J. Gilson, *Nat. Clim. Chang.* **2**, 425–428 (2012).
19. P. Tait, *On the Pressure-Errors of the "Challenger" Thermometers*, Narrative of the Challenger Expedition, Vol. II, Appendix A (HM Stationery Office, 1882).
20. British Oceanographic Data Centre, *GEBCO Digital Atlas* (2003); www.bodc.ac.uk/projects/data_management/international/gebco/gebco_digital_atlas.
21. V. Gouretski, K. Koltermann, *WOCE Global Hydrographic Climatology* (Tech. Rep. 35, Berichte des Bundesamtes für Seeschifffahrt und Hydrographie, 2004).
22. D. Roemmich, C. Wunsch, *Nature* **307**, 447–450 (1984).

23. S. G. Purkey, G. C. Johnson, *J. Clim.* **23**, 6336–6351 (2010).
24. D. G. Desbruyères, S. G. Purkey, E. L. McDonagh, G. C. Johnson, B. A. King, *Geophys. Res. Lett.* **43**, 10356–10365 (2016).
25. M. Palmer *et al.*, *Clim. Dyn.* **49**, 909–930 (2017).
26. I. Fukumori, *Mon. Weather Rev.* **130**, 1370–1383 (2002).
27. A. Köhl, *Q. J. R. Meteorol. Soc.* **141**, 166–181 (2015).
28. A. Atwood, E. Wu, D. Frierson, D. Battisti, J. Sachs, *J. Clim.* **29**, 1161–1178 (2016).
29. H. Stommel, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 3051–3055 (1979).
30. Y. Rosenthal, B. K. Linsley, D. W. Oppo, *Science* **342**, 617–621 (2013).
31. Y. Rosenthal, J. Kalansky, A. Morley, B. Linsley, *Quat. Sci. Rev.* **155**, 1–12 (2017).
32. S. Levitus *et al.*, *Geophys. Res. Lett.* **39**, L10603 (2012).
33. P. J. Durack, P. J. Gleckler, F. W. Landerer, K. E. Taylor, *Nat. Clim. Chang.* **4**, 999–1005 (2014).
34. J. M. Gregory *et al.*, *Geophys. Res. Lett.* **40**, 1600–1604 (2013).

ACKNOWLEDGMENTS

We thank C. Wunsch for highlighting the potential influence of past climate events in the modern ocean; D. Halpern for pointing us to the HMS Challenger data; U. Ninnemann, K. Nisancioglu, T. Eldevik, T. Furevik, Y. Rosenthal, D. Roemmich, L. H. Smedsrud, and T. Stocker for discussions; and three anonymous reviewers for suggestions. **Funding:** Supported by the James E. and Barbara V. Moltz Fellowship and NSF grant OCE-1357121 (G.G.) and by NSF grant OCE-1558939 (P.H.). **Author contributions:** G.G. and P.H. performed the research and the writing. G.G. contributed as lead author; P.H. contributed as the co-author.

Competing interests: The authors declare that they have no competing financial interests. **Data and materials availability:** Data to reproduce the findings are available at the National Centers for Environmental Information, accession number 0178641.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/363/6422/70/suppl/DC1
Materials and Methods
Supplementary Text
Table S1
Figs. S1 to S9
Movie S1
References (*35*–*46*)

15 September 2018; accepted 12 November 2018
10.1126/science.aar8413

VIROLOGY

Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa fever outbreak

L. E. Kafetzopoulou[1,2,3], S. T. Pullan[1,2], P. Lemey[4], M. A. Suchard[5], D. U. Ehichioya[3,6], M. Pahlmann[3,6], A. Thielebein[3,6], J. Hinzmann[3,6], L. Oestereich[3,6], D. M. Wozniak[3,6], K. Efthymiadis[7], D. Schachten[3], F. Koenig[3], J. Matjeschk[3], S. Lorenzen[3], S. Lumley[1], Y. Ighodalo[8], D. I. Adomeh[8], T. Olokor[8], E. Omomoh[8], R. Omiunu[8], J. Agbukor[8], B. Ebo[8], J. Aiyepada[8], P. Ebhodaghe[8], B. Osiemi[8], S. Ehikhametalor[8], P. Akhilomen[8], M. Airende[8], R. Esumeh[8], E. Muoebonam[8], R. Giwa[8], A. Ekanem[8], G. Igenegbale[8], G. Odigie[8], G. Okonofua[8], R. Enigbe[8], J. Oyakhilome[8], E. O. Yerumoh[8], I. Odia[8], C. Aire[8], M. Okonofua[8], R. Atafo[8], E. Tobin[8], D. Asogun[8,9], N. Akpede[8], P. O. Okokhere[8,9], M. O. Rafiu[8], K. O. Iraoyah[8], C. O. Iruolagbe[8], P. Akhideno[8], C. Erameh[8], G. Akpede[8,9], E. Isibor[8], D. Naidoo[10], R. Hewson[1,2,11,12], J. A. Hiscox[2,13,14], R. Vipond[1,2], M. W. Carroll[1,2], C. Ihekweazu[15], P. Formenty[10], S. Okogbenin[8,9], E. Ogbaini-Emovon[8]*, S. Günther[3,6]*†, S. Duraffour[3,6]*

The 2018 Nigerian Lassa fever season saw the largest ever recorded upsurge of cases, raising concerns over the emergence of a strain with increased transmission rate. To understand the molecular epidemiology of this upsurge, we performed, for the first time at the epicenter of an unfolding outbreak, metagenomic nanopore sequencing directly from patient samples, an approach dictated by the highly variable genome of the target pathogen. Genomic data and phylogenetic reconstructions were communicated immediately to Nigerian authorities and the World Health Organization to inform the public health response. Real-time analysis of 36 genomes and subsequent confirmation using all 120 samples sequenced in the country of origin revealed extensive diversity and phylogenetic intermingling with strains from previous years, suggesting independent zoonotic transmission events and thus allaying concerns of an emergent strain or extensive human-to-human transmission.

Lassa fever is an acute viral hemorrhagic illness, first described in 1969 in the town of Lassa, Nigeria (*1*). It is contracted primarily through exposure to urine or feces of infected *Mastomys* spp. rodents or, less frequently, through the bodily fluids of infected humans. Lassa virus (LASV) is endemic in parts of West Africa, including Nigeria, Benin, Côte d'Ivoire, Mali, Sierra Leone, Guinea, and Liberia (*2*). The upsurge of Lassa fever cases during the 2018 endemic season in Nigeria—referred to here as the 2018 Lassa fever outbreak—has been the largest on record, reaching 1495 suspected cases and 376 confirmed cases and affecting more than 18 states by 18 March (fig. S1). This notably exceeds the 102 confirmed cases reported during the same period in 2017 (fig. S1) (*3*). The unprecedented scale of the outbreak raised fears of the emergence of a strain with a higher rate of transmission. Because of these concerns, on 28 February the Nigeria Centre for Disease Control (NCDC) and the World Health Organization (WHO) urgently requested sequencing information and preliminary results from our pilot-scale study, in which we used a metagenomic approach with the Oxford Nanopore MinION device (Oxford Nanopore Technologies) to conduct in-country, mid-outbreak viral genome sequencing. This instigated a major uptick in sequencing efforts, leading to the sequencing of 120 samples.

Nanopore sequencing is an emerging technology with great potential. The MinION is a small, robust sequencing device suited for the genetic analysis of pathogens in remote or resource-limited settings (*4*). Nanopore sequencing of polymerase chain reaction (PCR) amplicons of Ebola virus genomes provided important data from the field in real time during the 2014–2016 Ebola virus disease outbreak in West Africa (*5*), and a more sophisticated multiplex amplicon sequencing methodology (*6*) has been used effectively during recent Zika and yellow fever outbreaks in Brazil (*7*, *8*). However, highly variable pathogens such as LASV present a substantial challenge for this type of amplicon-based approach. Owing to an interstrain nucleic acid sequence variation of up to 32 and 25% for the L (large segment encoding the RNA polymerase and the zinc-binding protein) and S (small segment encoding the glycoprotein and the nucleoprotein) segments, respectively (*9*), even PCR-based laboratory

[1]Public Health England, National Infection Service, Porton Down, UK. [2]National Institute of Health Research (NIHR), Health Protection Research Unit in Emerging and Zoonotic Infections, University of Liverpool, Liverpool, UK. [3]Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany. [4]Department of Microbiology and Immunology, Rega Institute, KU Leuven – University of Leuven, Leuven, Belgium. [5]Departments of Biomathematics, Biostatistics, and Human Genetics, University of California, Los Angeles, CA, USA. [6]German Center for Infection Research (DZIF), partner site Hamburg, Germany. [7]Artificial Intelligence Laboratory, Vrije Universiteit Brussel, Brussels, Belgium. [8]Irrua Specialist Teaching Hospital, Irrua, Nigeria. [9]Faculty of Clinical Sciences, College of Medicine, Ambrose Alli University, Ekpoma, Nigeria. [10]World Health Organization, Geneva, Switzerland. [11]Faculty of Infectious and Tropical Diseases, Department of Pathogen Molecular Biology, London School of Hygiene and Tropical Medicine, London, UK. [12]Faculty of Clinical Sciences and International Public Health, Liverpool School of Tropical Medicine, Liverpool, UK. [13]Singapore Immunology Network, Agency for Science, Technology and Research (A*STAR), Singapore. [14]Institute of Infection and Global Health, University of Liverpool, Liverpool, UK. [15]Nigeria Centre for Disease Control, Abuja, Nigeria.
*These authors contributed equally to this work.
†Corresponding author. Email: guenther@bni.uni-hamburg.de

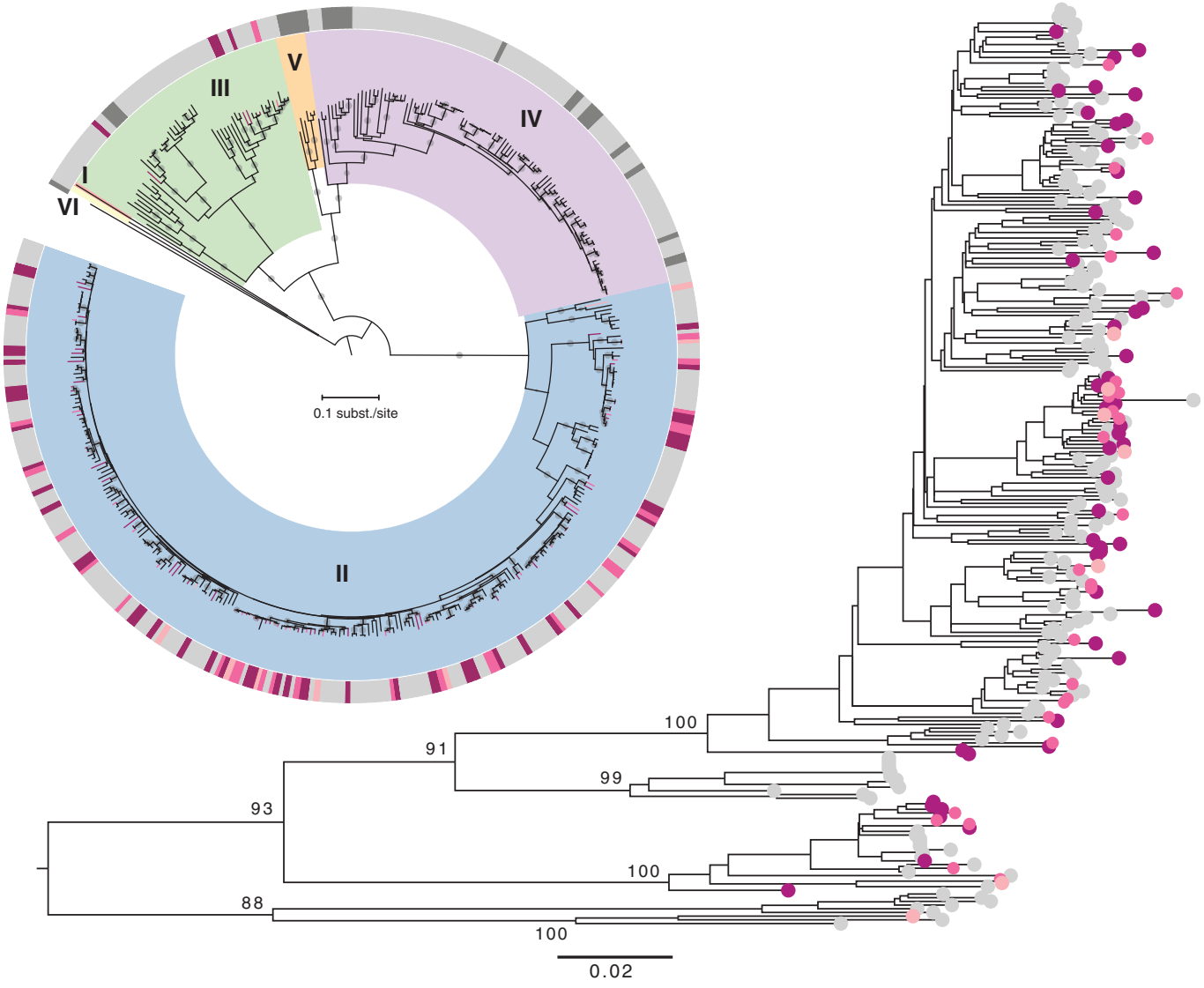


Fig. 1. Phylogenetic reconstruction of the S segment data. The circular tree includes 96 sequences from 2012 to 2017, 88 sequences from 2018, and sequences available from GenBank. The rectangular tree focuses on the genotype II clade (in blue in the circular tree), which includes most of the 2018 sequences. The six genotypes are indicated with different colors and roman numerals. Bootstrap support >90% is indicated with a small gray circle at the middle of their respective branches. The color strip highlights the human LASV sequences obtained from previous years (light gray); sequences obtained from rodent samples (dark gray); and, for 2018, the first seven sequences generated in Nigeria (light pink), the remaining 28 sequences analyzed on-site (medium pink), and the remaining sequences finalized in Europe (dark pink). The same color code is used in the genotype II rectangular tree. Bootstrap values >80% are shown for the major genotype II lineages.

diagnosis poses a serious challenge. Designing targeted whole-genome sequencing approaches, such as those using PCR amplicons or bait-and-capture probes, without prior knowledge of the targeted LASV lineage is therefore cumbersome. Random reverse-transcription (RT) and amplification by sequence-independent single primer amplification (SISPA) for metagenomic sequencing to identify RNA viruses has been demonstrated to work on the MinION (*10*), and our previous work highlighted the feasibility of retrieving complete viral genomes directly from patient samples at clinically relevant viral titers using this approach for dengue and chikungunya viruses (*11*). We describe here the application of field metagenomic sequencing of LASV at the Irrua Specialist Teaching Hospital (ISTH), Edo State, during the 2018 Lassa fever season.

A total of 120 LASV-positive samples were sequenced during a 7-week mission; these were selected on the basis of cycle threshold value and location of the 341 cases reported by ISTH between 1 January and 18 March 2018 (figs. S1 and S2). The majority of samples originated from Edo State followed by Ondo and Ebonyi (fig. S2). Selected samples covered the wide range of clinical viral loads observed, including several samples testing negative in one of the two real-time RT-PCR assays used (fig. S3 and data S1). Up to six samples were run in multiplex per MinION flow cell, along with a negative control. To produce high-confidence consensus sequences for phylogenetic inference, we chose to map both basecalled reads and raw signal data to a reference sequence and call variants using Nanopolish software, as developed for the West African Ebola virus disease outbreak (*5*); basecalled reads were then remapped to the consensus and a further round of correction was applied (fig. S4). Owing to the diversity of LASV, selection of an individual

reference genome for read alignment was required for each sample. To select the closest existing LASV reference genome, nonhuman reads from each sample were assembled de novo using Canu (*12*). A notable proportion of reads generated per sample were LASV at an average frequency of 4.26% with a maximum of 42.9%, allowing for sufficient genomic sequence (>70%) for phylogenetic comparison of at least one segment in 91 of the samples tested (figs. S3 to S6).

Additionally, sequences were validated by Illumina resequencing of 14 SISPA preparations, which matched with their Oxford Nanopore counterparts with little to no divergence, confirming the accuracy of the Oxford Nanopore approach (table S1).

Metagenomic classification using the Centrifuge software system (*13*) identified 0.10% of reads from sample 110 as originating from hepatitis A virus, providing 74% genome coverage at 20-fold depth. LASV accounted for 0.83% of reads in the same sample, providing 96% genome coverage. These findings demonstrate the potential of this simple approach to identify multiple RNA viruses, including those present as co-infections. In all other samples tested, LASV was the sole pathogen identified despite a small number of reads classified as other viruses (fig. S7 and data S1).

To dissect the molecular epidemiology of the 2018 Lassa fever outbreak in Nigeria, we performed phylogenetic analysis of all newly generated LASV sequences together with unpublished sequences from previous years (data S2) and sequences available in GenBank. We used this as a frame of reference to document how the genomic data generated in real time (made publicly available at virological.org) provided valuable epidemiological insights into the unfolding outbreak dynamics.

Maximum likelihood phylogenetic reconstruction of the S segment sequences indicates that all 2018 viruses fall within the Nigerian LASV diversity, specifically within genotypes II and

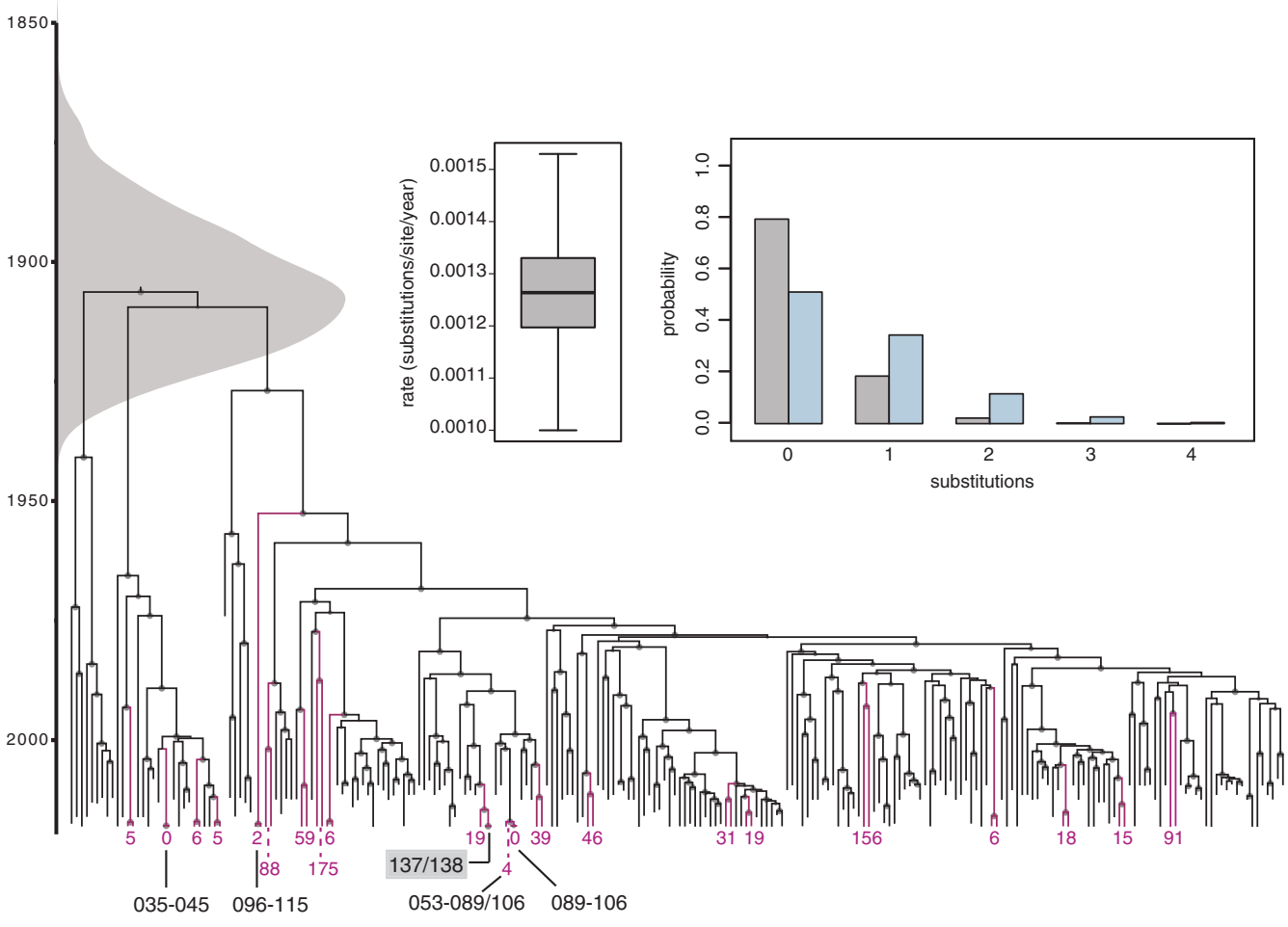


Fig. 2. Assessing the potential for direct linkage between pairs of 2018 sequences in the S segment. The maximum clade credibility tree summarizes a Bayesian evolutionary inference for the genotype II sequences in the S segment. A time scale and a marginal posterior distribution for the time to the most recent common ancestor are shown to the left. The size of the internal node circles reflects posterior probability support values. 2018 sequences clustering as pairs are indicated in dark pink; the number of substitutions between them is indicated at their respective tips. A posterior estimate of the evolutionary rate and probability distributions for observing a given number of substitutions during a human-to-human transmission event are shown as insets. The distribution represented by gray bars is based on the mean evolutionary rate estimate and a mean estimate for the generation time, whereas the light blue distribution is based on upper estimates and also incorporates an upper estimate for the MinION sequencing error (supplementary methods). At the bottom of the tree, clusters of sequences for which human-to-human transmission cannot be excluded according to the upper estimates of generation time are indicated. A pair of identical sequences (137-138) that was retrospectively found to be derived from the same patient is marked with a gray box. One pair (096-115) was disregarded as a potential transmission chain because of 21 differences in the L segment (fig. S9). The temporal signal before BEAST inference was explored in fig. S10.

III, and they are phylogenetically interspersed with Nigerian LASV sequences from previous years (Fig. 1). This phylogenetic pattern is mimicked by the L segment reconstruction (fig. S8). Only seven viruses in the entire genome dataset (n = 348) were identified as clustering significantly differently in the L and S segments (supplementary methods), which is in line with the small number of potential LASV reassortments identified previously (*9*). The phylogenetic pattern implicates independent spillover from rodent hosts as the major driver of Lassa fever incidence during the outbreak (Fig. 1 and fig. S8).

However, a number of sequences from the 2018 outbreak clustered as pairs in the phylogenetic reconstructions, raising concerns over human-to-human transmission. We illustrate such cluster pairs in a Bayesian time-measured tree estimated from genotype II S (Fig. 2) and L segment sequences (fig. S9). These analyses resulted in highly similar evolutionary rate estimates for both segments (mean, ~1.2×10^{-3} substitutions per site per year) (Fig. 2 and figs. S9 and S10), in agreement with previous estimates (*9*). We used these rate estimates together with an estimate of the time between successive cases in a transmission chain to assess how many substitutions can be expected between directly linked infections. We compared conservative to more liberal expectations, the latter accommodating an independent upper estimate of potential sequencing errors (Fig. 2 and fig. S9). In the S segment, for example, more than two substitutions between sequences from directly linked infections is highly unlikely ($P < 0.01$ and $P = 0.03$, respectively, for the conservative and liberal probability estimates). This expectation is consistent with the low number of substitutions observed in the coding region of human-to-human LASV transmission (*14*). Four clusters of sequences showing ≤4 and ≤12 nucleotide differences in the S and L segments, respectively, were identified (035-045, 035-058, 137-138, and 053-089-106; for some of them, only the S or L segment sequence was available). Retrospective tracing revealed that the sequences for pairs 137-138 and 035-058 were derived from the same patients. Epidemiological investigation of the remaining clusters did not provide evidence for transmission chains, though direct linkage cannot be excluded. Even when applying liberal assumptions for the number of mutations during human-to-human transmission, the vast majority of cases during the 2018 outbreak resulted from spillover from the natural reservoir.

A request for information on circulating strains was made on 28 February at the height of the outbreak; within 10 days, our pilot study was expedited and the initial analysis completed. The fact that the 2018 outbreak was fueled by the circulating LASV diversity and not by transmission of a new or divergent lineage was already evident from the first seven genomes generated by 10 March (fig. S1). This information was promptly communicated to the NCDC, forming the basis of its report released on 12 March 2018 (*15*). Whereas this small sample was restricted to genotype II, the final collection of 36 LASV genome sequences generated on-site also included a representative of genotype III (Fig. 1 and fig. S9), further supporting the spillover of long-standing LASV diversity in the outbreak. The conclusions drawn from the first set of genome sequences immediately eased fears of extensive human-to-human transmission and allowed public health resources to be allocated appropriately. The response was focused on intensified community engagement on rodent control, environmental sanitation, and safe food storage. Further research is needed to evaluate whether improved diagnostics and disease awareness and/or ecological and climate factors promoting transmission are the drivers behind the changing epidemiology of Lassa fever in Nigeria.

Portable metagenomic sequencing of genetically diverse RNA viruses on the MinION, direct from patient samples without the need to export material outside of the country of origin and with no pathogen-specific enrichment, is shown to be a feasible methodology enabling a real-time characterization of potential outbreaks in the field.

REFERENCES AND NOTES

1. J. D. Frame, J. M. Baldwin Jr., D. J. Gocke, J. M. Troup, *Am. J. Trop. Med. Hyg.* **19**, 670–676 (1970).
2. D. A. Asogun *et al*., *PLOS Negl. Trop. Dis.* **6**, e1839 (2012).
3. WHO, "Lassa Fever – Nigeria" (2018); www.who.int/csr/don/23-march-2018-lassa-fever-nigeria/en/.
4. M. Jain, H. E. Olsen, B. Paten, M. Akeson, *Genome Biol.* **17**, 239 (2016).
5. J. Quick *et al*., *Nature* **530**, 228–232 (2016).
6. J. Quick *et al*., *Nat. Protoc.* **12**, 1261–1276 (2017).
7. N. R. Faria *et al*., *Nature* **546**, 406–410 (2017).
8. N. R. Faria *et al*., *Science* **361**, 894–899 (2018).
9. K. G. Andersen *et al*., *Cell* **162**, 738–750 (2015).
10. A. L. Greninger *et al*., *Genome Med.* **7**, 99 (2015).
11. L. E. Kafetzopoulou *et al*., *Euro Surveill.* **23**, 1800228 (2018).
12. S. Koren *et al*., *Genome Res.* **27**, 722–736 (2017).
13. D. Kim, L. Song, F. P. Breitwieser, S. L. Salzberg, *Genome Res.* **26**, 1721–1729 (2016).
14. S. L. M. Whitmer *et al*., *Emerg. Infect. Dis.* **24**, 599–602 (2018).
15. Nigeria Centre for Disease Control, "Early Results of Lassa Virus Sequencing & Implications for Current Outbreak Response in Nigeria" (2018); https://ncdc.gov.ng/news/121/early-results-of-lassa-virus-sequencing-%26-implications-for-current-outbreak-response-in-nigeria.
16. P. Lemey, ISTH-BNITM-PHE/LASVsequencing: LASVrelease, Zenodo (2018); http://doi.org/10.5281/zenodo.1481015.

ACKNOWLEDGMENTS

We thank the health authorities of Nigeria for their cooperation during the outbreak response. **Funding:** L.E.K., S.T.P., R.H., R.V., M.W.C., and J.A.H. acknowledge funding by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Emerging and Zoonotic Infections at the University of Liverpool in partnership with Public Health England (PHE), in collaboration with Liverpool School of Tropical Medicine. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health, or Public Health England. L.E.K. has received travel expenses and accommodation from Oxford Nanopore to speak at conferences regarding this work. L.E.K. has received some reagents free of charge from Oxford Nanopore in support of her Ph.D. project. M.W.C. has received reagents free of charge from Oxford Nanopore in support of previous projects not related to the work presented in this manuscript. L.E.K. and M.W.C. have not received other financial compensation nor hold shares. P.L. and M.A.S. acknowledge funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (grant 725422-ReservoirDOCS) and from the Wellcome Trust Collaborative Award, 206298/Z/17/Z. P.L. acknowledges support by the Special Research Fund, KU Leuven ("Bijzonder Onderzoeksfonds," KU Leuven, OT/14/115), and the Research Foundation–Flanders ("Fonds voor Wetenschappelijk Onderzoek – Vlaanderen," G066215N, G0D5117N, and G0B9317N). M.A.S. acknowledges support under National Science Foundation grant DMS 1264153. This study was supported by the German Federal Ministry of Health through support of the WHO Collaborating Centre for Arboviruses and Hemorrhagic Fever Viruses at the Bernhard Nocht Institute for Tropical Medicine (agreements ZMV I 1-2517WHO005 and ZMV I 1-2517WHO010) and through the Global Health Protection Program (agreement ZMVI1-2517-GHP-704), the German Federal Ministry for Economic Cooperation and Development through the Rapid Deployment Expert Group to Combat Threats (SEEG), the European Union's Horizon 2020 research and innovation program to S.G. (grant 653316-EVAg), and the German Research Foundation (DFG) to S.G. and D.U.E. (GU 883/4-1). D.U.E. acknowledges fellowships from Alexander von Humboldt Foundation and Kirmser Foundation. The funders had no role in the design and interpretation of the data and preparation of the manuscript. **Author contributions:** L.E.K., S.G., S.D., S.T.P., and P.L. conceptualized the study; L.E.K., S.T.P., and P.L. set up the methodology; L.E.K., J.H., A.T., S.D., and D.U.E. performed sequencing and data validation; L.E.K., P.L., M.A.S., S.T.P., D.S., F.K., J.M., and S.Lo. performed the formal sequencing data analysis; L.E.K., S.D., J.H., A.T., M.P., and L.O. performed sample selection, data collection, and organization of sequencing datasets; D.M.W., K.E., D.S., F.K., and J.M. set up and assisted with the bioinformatics pipeline; M.A.S., D.U.O., M.P., L.O., Y.I., D.I.A., T.O., E.O., R.O., J.Ag., B.E., J.Ai., P.E., B.O., S.E., P.A., M.A., R.Es., E.M., R.G., A.E., G.I., G.Od., G.Ok., R.En., J.O., E.O.Y., I.O., C.A., M.O., R.A., E.T., D.A., N.A., P.O.O., M.O.R., K.O.I., C.O.I., P.A., C.E., G.A., and E.I. performed diagnostic analysis; L.E.K., S.T.P., P.L., and S.D. visualized data presentation; L.E.K., S.T.P., P.L., and S.D. wrote the manuscript; all authors reviewed and edited the manuscript; S.G., M.W.C., J.A.H., R.H., and R.V. supervised the study; M.P., R.V., A.T., C.I., P.F., D.N., S.O., E.O.E., S.G., S.D., and S.Lu. performed project administration and implementation; S.G., P.L., M.W.C., R.V., R.H., J.A.H., L.E.K., and D.U.E. were involved in funding acquisition. **Competing interests:** C.I. is a member of the WHO Strategic Technical Advisory Group on Infectious Diseases; D.A. serves as an expert for the WHO R&D Blueprint for action to prevent epidemics (the Blueprint); S.G. is a member of the Scientific Advisory Group (SAG) to advise WHO on the implementation of the Blueprint, including a plan for international coordination of the R&D effort in the event of a highly infectious pathogen epidemic; S.O. serves as an expert for the Blueprint. All other authors declare no competing interests. **Data and materials availability:** LASV sequences from 2018 are deposited in GenBank under BioProject PRJNA482058 (data S1); sequences from 2012 to 2017 are deposited under BioProjects PRJNA482054 and PRJNA482058 (data S2). Alignments, trees, and BEAST xml files are available at https://github.com/ISTH-BNITM-PHE/LASVsequencing and in (*16*).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/363/6422/74/suppl/DC1
Materials and Methods
Figs. S1 to S10
Table S1
References (*17–30*)
Data S1 and S2

3 August 2018; accepted 12 November 2018
10.1126/science.aau9343

PALEONTOLOGY

An elephant-sized Late Triassic synapsid with erect limbs

Tomasz Sulej[1] and Grzegorz Niedźwiedzki[2]*

Here, we describe the dicynodont *Lisowicia bojani*, from the Late Triassic of Poland, a gigantic synapsid with seemingly upright subcursorial limbs that reached an estimated length of more than 4.5 meters, height of 2.6 meters, and body mass of 9 tons. *Lisowicia* is the youngest undisputed dicynodont and the largest nondinosaurian terrestrial tetrapod from the Triassic. The lack of lines of arrested growth and the highly remodeled cortex of its limb bones suggest permanently rapid growth and recalls that of dinosaurs and mammals. The discovery of *Lisowicia* overturns the established picture of the Triassic megaherbivore radiation as a phenomenon restricted to dinosaurs and shows that stem-group mammals were capable of reaching body sizes that were not attained again in mammalian evolution until the latest Eocene.

The time interval from the end of the Permian to the beginning of the Jurassic (252 to 201 million years ago) coincided with a trophic and taxonomic restructuring of terrestrial ecosystems (*1*). During the middle Permian, the predominant large herbivores were dinocephalians—stem-group mammals (in the clade Synapsida, "mammal-like reptiles") such as *Tapinocephalus*, which was up to 3 m in length and weighed an estimated 2 tons (*2*). These animals were associated with faunas of nonsynapsid pareiasaurian parareptiles and a rich variety of other synapsids. In the late Permian, the large-size synapsid herbivores (rhachiocephalid dicynodonts) and the last pareiasaurians played the role of largest terrestrial herbivores (*2*). By the Early Jurassic, no stem-group mammals were larger than a half meter in length, and most were much smaller (*3*, *4*), whereas herbivorous dinosaurs exceeded 15 m in length and 10 tons in weight; virtually all mid-sized to large tetrapods were archosaurs. Until now, this pattern has appeared clear-cut, with Triassic synapsids approaching the largest Permian forms in size but with no forms reaching sizes of the first large sauropodomorphs.

Dicynodonts were among the most abundant and diverse synapsids from the middle Permian to the early Late Triassic (*5*, *6*). There were two episodes of body size increase in their evolution: the first in the late Permian, exemplified by *Rhachiocephalus*, and the second in the Middle to Late Triassic, when dicynodonts such as *Placerias* were the dominant herbivores in some ecosystems (*6*). There have been comparatively few rigorous estimates of dicynodont or dinocephalian body masses presented in the literature. The largest of these dicynodonts (such as kannemeyeriiforms) are estimated previously to have reached lengths of 3 to 3.5 m and weights of up to 1 to 2 tons (*2*, *7*), which is very similar to the largest middle Permian dinocephalians, such as *Tapinocephalus* or *Moschops* (*2*).

Fossils of Triassic dicynodonts are extremely abundant in African, Asian, and North and South American deposits but are comparatively poorly known from other regions (*5*, *6*). Dicynodonts were seemingly rare in the European Late Triassic, being known only from a single mandible (*8*) and questionable isolated bones (*9*). The Late Triassic dicynodont fossils described here, of *Lisowicia bojani*, are the first substantial finds from Europe (*10*, *11*).

Many previous authors have analyzed the structure of the pelvic girdle of Triassic dicynodonts, resulting in the unanimous opinion that the posture of the hindlimbs was erect (*12*). By contrast, most authors agree that Triassic dicynodonts had sprawling forelimbs with the horizontal position of the humerus (*13*). *Lisowicia* has a relatively conventional dicynodont hindlimb construction but departs from the standard forelimb posture (Fig. 1). In many respects, its forelimb position resembles that of large quadrupedal dinosaurs, but forelimb elements of *Lisowicia* are morphologically similar to other dicynodonts (Fig. 2). The result is a subcursorial tetrapod with upright limb posture, unlike any other known stem-group mammal but comparable with that of large crown-group mammals such as rhinoceroses and hippopotami, as well as quadrupedal dinosaurs such as sauropodomorphs and ceratopsians.

Histological data from limb bones of *Lisowicia* provide life history data that complement its distinctness. Like other large dicynodonts, *Lisowicia* shows a large area of bone resorption in the inner cortex. However, unlike in other genera, there is no clear sign of growth-slowing later in life. The presence of potential lines of arrested growth in the studied tibia might be indicative of slower, more episodic growth. However, the lack of characteristic slow-growing tissue on the periphery of

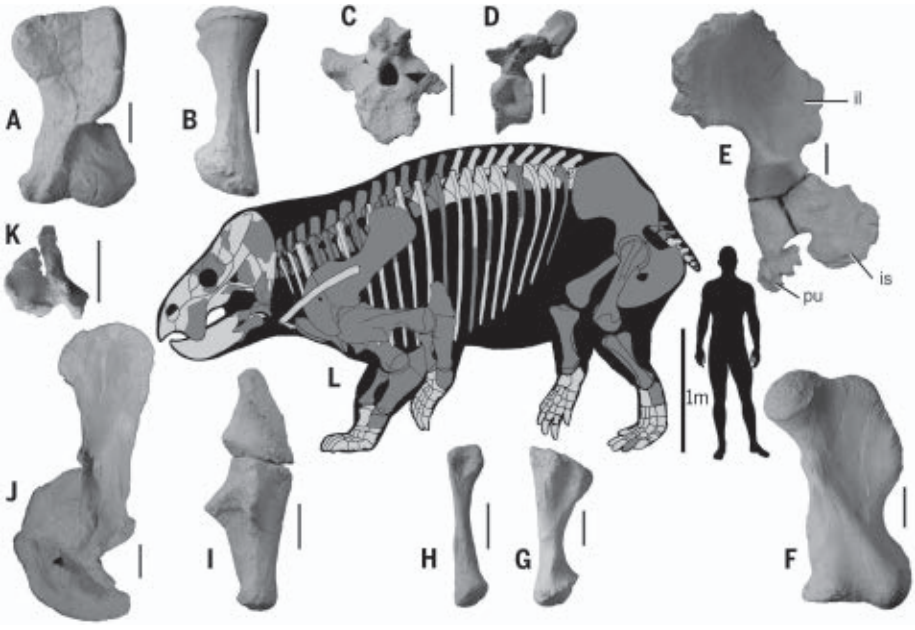


Fig. 1. The skeleton restoration of *Lisowicia bojani*. (**A**) Left humerus (ZPAL V.33/96) in ventral view. (**B**) Left radius (ZPAL V.33/665) in lateral view. (**C**) Cervical vertebrae (ZPAL V.33/720) in posterior view. (**D**) Dorsal vertebrae (ZPAL V.33/720) in lateral view. (**E**) Left pelvis (ZPAL V.33/720; ilium, pubis, and ischium) in lateral view. (**F**) Left femur (ZPAL V.33/75) in anterior view. (**G**) Left tibia (ZPAL V.33/75) in lateral view. (**H**) Left fibula (ZPAL V.33/75) in medial view. (**I**) Left ulna (ZPAL V.33/470) in lateral view. (**J**) Left scapulocoracoid (ZPAL V.33/468) in lateral view. (**K**) Fused quadrate and quadratojugal (ZPAL V.33/735) in posterior view. Scale bars, 10 cm (A) to (K); 1 m for the skeleton. (**L**) Light gray bones represent missing elements. il, ilium; pu, pubis; is, ischium.

[1]Institute of Paleobiology, Polish Academy of Sciences (PAS), Twarda 51/55, 00-818 Warsaw, Poland. [2]Department of Organismal Biology, Uppsala University, Norbyvägen 18A, 752 36 Uppsala, Sweden.
*Corresponding author. Email: grzegorz.niedzwiedzki@ebc.uu.se

the largest bones of *Lisowicia* suggests that the studied material represents either a fast-growing taxon or juvenile/subadult individuals of extremely large body size. However, the second explanation is rather unlikely because of its size and that these two bones are well ossified.

Lisowicia demonstrates that Late Triassic dicynodonts became specialized herbivores. It displays several features in the limb skeleton that suggest that this group evolved new postural adaptations. The massive scapula of *Lisowicia* lacks a distinct acromion process for articulation with the clavicle, the scapula articulates with the humerus on its posteroventrally (instead of posterolaterally) located glenoid, and distal articulation surfaces of the humerus are in the same plane instead of being rotated (formal taxonomic description is provided in the supplementary materials).

On the basis of published scaling relationships (*14*), we estimate an adult body mass of 9000 kg, which approaches that of an African elephant [the largest recorded individual stood 4 m at the shoulders and weighed 10,000 kg (*15*)]. This confirms that *Lisowicia* was certainly the largest Triassic land nondinosaur tetrapod. Gigantism in herbivorous dinosaurs first emerged in the Late Triassic, with the evolution of the first large sauropodomorphs (*16*, *17*) and then the earliest true sauropods (*18*). Until now, gigantism in the Triassic appeared to be entirely a dinosaur adaptation (*19*), and previously known Triassic dicynodonts were substantially smaller. The discovery of *Lisowicia* suggests that general ecological factors may have been driving the process, rather than clade-specific attributes of dinosaurs (*20*).

The find of *Lisowicia* shows that at least one dicynodont lineage also participated in the "push for gigantism" at the same time as the sauropodomorphs (*20*) but also suggests that their evolutionary history in the Late Triassic is poorly documented (Fig. 3A). In addition, recognition of *Lisowicia* as a placeriine dicynodont together with the resurrection and recent description of *Pentasaurus* from South Africa (*21*) alters our understanding of the Late Triassic fossil record of dicynodonts. Although their relatively lower abundance and richness compared with those of Middle and early Late Triassic faunas suggest evolutionary decline, the concept of the Late Triassic kannemeyeriiforms as highly geographically restricted relicts is no longer valid (*22*). The recognition of dicynodonts in the late Norian–earliest Rhaetian of Europe (*10*) and Karoo Basin (*21*) conflicts with some ideas on early Late Triassic dicynodont extinction and survival, namely their supposed absence during the radiation of early sauropodomorphs (Fig. 3B).

Upright posture has been associated with decreased joint stress and energetic cost of locomotion (*23*). Selection pressures on some aspects of lifestyle or ecology were likely drivers of the evolution of the distinct posture of *Lisowicia* among dicynodonts. Increase in the body size of dicynodonts across the Late Triassic may have been driven by selection pressure to reach a size refuge from large predators (*24*). It is possible also that the gigantism of the latest dicynodonts was a metabolic adaptation that allowed these animals to maximize food retention time and consequently the energy gain (*25*). It took Late Triassic dicynodonts some 20 million years to produce giant forms (Fig. 3, B and C), and it was a rather gradual size increase and a similar pace

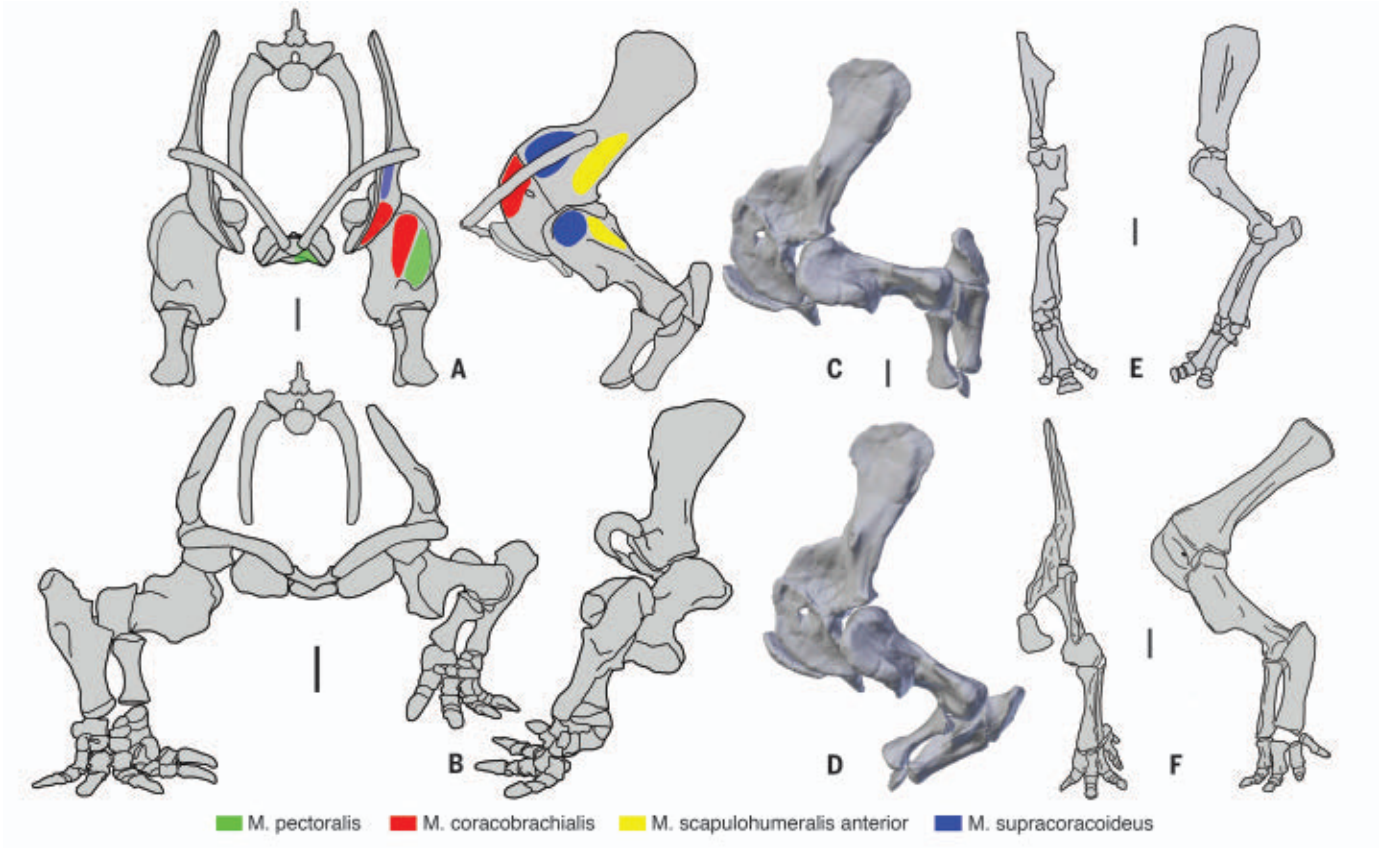


Fig. 2. Comparison of the reconstructed pectoral girdle of *Lisowicia bojani* with another dicynodont, dinosaur, and recent mammal. (**A**) Position of bones of *L. bojani* in anterior and lateral views. Some proportions of the bones were estimated by means of comparison with articulated skeletons of *Parakannemeyeria* (IVPP V. 979) and *Sinokannemeyeria* (IVPP V.974), but most were inferred from the size of articulation areas. (**B**) Reconstruction of large dicynodont *Stahleckeria* (GPIT/RE/8001) in anterior and lateral views. (**C** and **D**) Hypothetical flexibility of the humerus in protraction-retraction. (**E**) Reconstruction of rhinoceros *Diceros* in anterior and lateral views based on MPUWr 502223. (**F**) Reconstruction of *Triceratops* in anterior and lateral views based on (*27*). Scale bars, 10 cm.

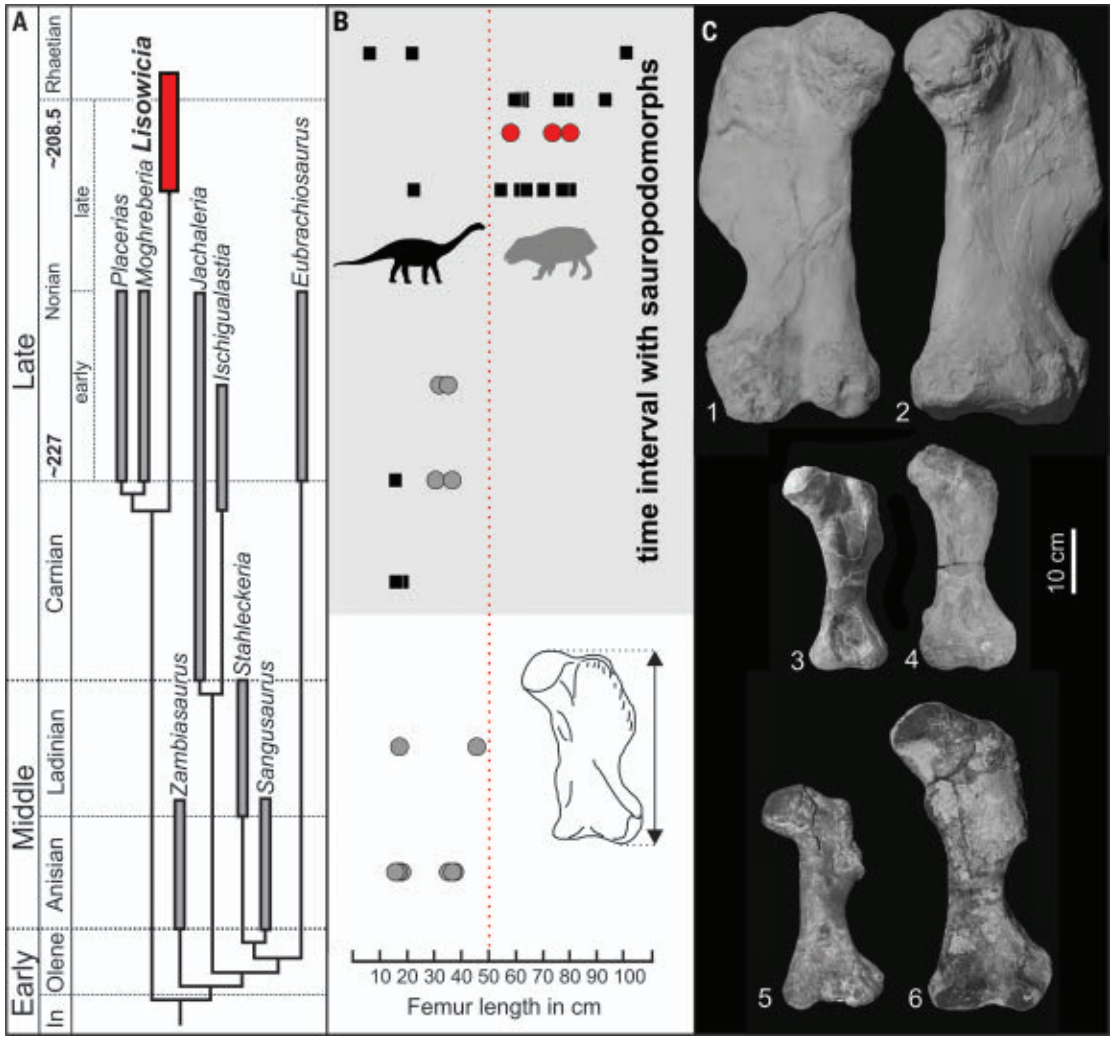


Fig. 3. Phylogeny of kannemeyeriiform dicynodonts and its relationship with the changes in femur length of dicynodonts and sauropodomorphs. (**A**) Time-calibrated phylogeny of the Triassic dicynodonts simplified after (*22*) (numerical ages for the base and top of Norian are based on the Chronostratigraphic Chart of the ICS v. 2018/8) with position of *L. bojani*. (**B**) Femur length (body size proxy) of sauropodomorph (black squares) and dicynodont (gray circles) taxa from the Middle to Late Triassic plotted at the stratigraphic range midpoints for each taxon. (**C**) Comparison of dicynodont femur bones (1, 2, ZPAL V.33/763, *Lisowicia*; 3, UCMP 32394, *Placerias*; 4, MCZ 378 58M, *Ischigualastia*; 5, MCN PV 3600, *Dinodontosaurus*; 6, GPIT/RE/8002, *Stahleckeria*).

to that observed in the evolution of herbivorous dinosaurs in the Mesozoic and mammalian lineages in the mid-late Paleogene (*26*). All suggest that in the Late Triassic, there was a substantial temporal overlap in the occurrence of very large herbivores: the previously dominant dicynodonts and their emerging ecological analogs among archosaurs, the sauropodomorph dinosaurs.

REFERENCES AND NOTES

1. H.-D. Sues, N. C. Fraser, *Triassic Life on Land: The Great Transition* (Columbia Univ. Press, 2010).
2. G. M. King, *Anomodontia, Encyclopedia of Paleoherpetology (Part 17C)* (Gutsav Fischer Verlag, 1988).
3. Z. Kielan-Jaworowska, R. L. Cifelli, Z.-X. Luo, *Mammals from the Age of Dinosaurs: Origins, Evolution, and Structure* (Columbia Univ. Press, 2004).
4. T. S. Kemp, *The Origin and Evolution of Mammals* (Oxford Univ. Press, 2005).
5. J. Fröbisch, *Earth Sci. Rev.* **95**, 119–157 (2009).
6. J. Fröbisch, *PLOS ONE* **3**, e3733 (2008).
7. A. R. Fiorillo, K. Padian, C. Musikasinthorn, *Palaios* **15**, 373–386 (2000).
8. R. R. Schoch, *Neues Jahrb. Geol. Palaontol. Abh.* **263**, 119–123 (2012).
9. M. W. Maisch, C. S. Vega, R. R. Schoch, *Palaeodiversity* **2**, 271–278 (2009).
10. J. Dzik, T. Sulej, G. Niedźwiedzki, *Acta Palaeontol. Pol.* **53**, 733–738 (2008).
11. T. Sulej, R. Bronowicz, M. Tałanda, G. Niedzwiedzki, *Proc. R. Soc. Edinb.* **101**, 261–269 (2011).
12. J. Fröbisch, *Can. J. Earth Sci.* **43**, 1297–1308 (2006).
13. L. R. Walter, in *The Beginning of the Age of Dinosaurs* (Cambridge Univ. Press, 1986), pp. 89–97.
14. N. E. Campione, D. C. Evans, *BMC Biol.* **10**, 60 (2012).
15. D. Macdonald, *The New Encyclopedia of Mammals* (Oxford Univ. Press, 2001).
16. A. M. Yates, *Palaeontology* **46**, 317–337 (2003).
17. P. M. Sander *et al.*, *Biol. Rev. Camb. Philos. Soc.* **86**, 117–155 (2011).
18. A. M. Yates, J. W. Kitching, *Proc. Biol. Sci.* **270**, 1753–1758 (2003).
19. P. C. Sereno, *Science* **284**, 2137–2147 (1999).
20. R. B. Sookias, R. J. Butler, R. B. J. Benson, *Proc. Biol. Sci.* **279**, 2180–2187 (2012).
21. Ch. F. Kammerer, *Palaeontologia Africana* **52**, 102–128 (2018).
22. C. F. Kammerer, J. Fröbisch, K. D. Angielczyk, *PLOS ONE* **8**, e64203 (2013).
23. A. A. Biewener, *Science* **245**, 45–48 (1989).
24. G. Niedźwiedzki, P. Gorzelak, T. Sulej, *Lethaia* **44**, 87–92 (2011).
25. P. Bajdek, K. Owocki, G. Niedźwiedzki, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **411**, 1–17 (2014).
26. D. Prothero, *Rhinoceros Giants: The Palaeobiology of Indricotheres* (Indiana Univ. Press, 2013).
27. S. Fujiwara, *J. Vertebr. Paleontol.* **29**, 1136–1147 (2009).

ACKNOWLEDGMENTS

We thank J. Dzik (Institute of Paleobiology, PAS) for proposing this research and support, M. Dec (Institute of Paleobiology, PAS) for help with preparation of dicynodont shoulder girdle virtual model, P. E. Ahlberg (Uppsala University) for discussion, D. Snitting (Uppsala University) for help with phylogenetic and computed tomography data, N. Campione (University of New England) for his help with body mass estimate, and K. Zaremba-Niedźwiedzka (Uppsala University) for help during preparation of the manuscript. We are very thankful to anonymous reviewers whose comments radically improved the final version of the paper. **Funding:** The study was supported by Polish grant (2012/07/B/NZ8/02707) and Swedish Vetenskapsrådet grant (2017-05248). **Author contributions:** T.S. designed the study. T.S. and G.N. performed the comparative and analytical work, participated in morphological studies, and wrote the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All described specimens are accessioned at the Institute of Paleobiology, PAS (Poland).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/363/6422/78/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S16
Tables S1 to S7
References (*28*–*74*)
Movies S1 to S4

28 November 2016; resubmitted 4 August 2018
Accepted 19 October 2018
Published online 22 November 2018
10.1126/science.aal4853

EVOLUTION

DNA fragility in the parallel evolution of pelvic reduction in stickleback fish

Kathleen T. Xie[1,2,3], Guliang Wang[4], Abbey C. Thompson[1,5], Julia I. Wucherpfennig[1], Thomas E. Reimchen[6], Andrew D. C. MacColl[7], Dolph Schluter[8], Michael A. Bell[9]*, Karen M. Vasquez[4], David M. Kingsley[1,2]†

Evolution generates a remarkable breadth of living forms, but many traits evolve repeatedly, by mechanisms that are still poorly understood. A classic example of repeated evolution is the loss of pelvic hindfins in stickleback fish (*Gasterosteus aculeatus*). Repeated pelvic loss maps to recurrent deletions of a pelvic enhancer of the *Pitx1* gene. Here, we identify molecular features contributing to these recurrent deletions. *Pitx1* enhancer sequences form alternative DNA structures in vitro and increase double-strand breaks and deletions in vivo. Enhancer mutability depends on DNA replication direction and is caused by TG-dinucleotide repeats. Modeling shows that elevated mutation rates can influence evolution under demographic conditions relevant for sticklebacks and humans. DNA fragility may thus help explain why the same loci are often used repeatedly during parallel adaptive evolution.

Many phenotypic traits evolve repeatedly in organisms adapting to similar environments, and studying these cases can reveal ecological and genetic factors shaping parallel evolution (*1*, *2*). For example, loss of pelvic appendages has evolved repeatedly in mammals, amphibians, reptiles, and fishes. Marine stickleback fish (*Gasterosteus aculeatus*) develop a robust pelvic apparatus, whereas many freshwater populations have lost pelvic structures (*3*). Pelvic reduction is associated with particular ecological conditions, is likely adaptive, and maps to recurrent and independent deletions of a pelvic enhancer (*Pel*) upstream of the homeodomain transcription factor gene (*Pitx1*) that show repeatable molecular signatures of positive selection (*4*–*7*). This unusual spectrum of regulatory deletions contrasts with the accumulation of single-nucleotide changes in other studies (*6*, *8*, *9*), hinting that special DNA features may shape adaptive variation at the *Pitx1* locus (*6*).

Pel enhancer sequences show high predicted helical twist flexibility (*6*), a DNA feature associated with delayed replication and fragile site instability (*10*). To examine whether *Pel* forms alternative DNA structures in vitro, we used two-dimensional (2D) electrophoresis to analyze distributions of plasmid topoisomers (*11*) (Fig. 1A). A control stickleback genomic region showed smooth curves characteristic of B-DNA (Fig. 1B). In contrast, *Pel* sequences from marine populations showed mobility shifts characteristic of alternative DNA structure formation (Fig. 1B). Structural transitions started at a negative superhelical density of $-\sigma = 0.043$ and changed apparent linking numbers by 10 to 16 helical turns, similar to shifts produced by Z-DNA (left-handed DNA, starting $-\sigma = 0.046$) of ~105 to 170 base pairs (bp) (*12*, *13*). *Pel* sequences from pelvic-reduced populations did not show unusual electrophoretic transitions (Fig. 1B), suggesting that natural *Pel* mutations remove sequences forming alternative DNA structures.

To test the effect of *Pel* sequences on chromosome stability in vivo, we measured the rate of DNA double-strand breaks in yeast artificial chromosomes (Fig. 2A). Constructs without added test regions broke at background rates of 3.37 breaks per 10^6 divisions (Fig. 2B), consistent with previous reports (*14*). Chromosomes containing marine *Pel* broke ~25 to 50 times more frequently (Fig. 2B), a rate even higher than that of previously analyzed human fragile sites (*14*). *Pel* from freshwater pelvic-reduced populations [but not freshwater pelvic-complete populations (fig. S1)] broke at rates similar to that of the control (Fig. 2B), suggesting that natural *Pel* mutations remove breakage-prone regions.

Reverse complements of marine *Pel* broke ~10 to 20 times less frequently than identical sequences in the forward orientation (Fig. 2B). RNA transcription can influence fragile site breakage (*15*), but reversing transcription orientation of the nearby *URA3* marker did not significantly affect *Pel* fragility (Fig. 2C). In contrast, adding a replication origin on the opposite side of *Pel* did switch fragility, making the forward sequence stable and the reverse complement fragile (Fig. 2C). Thus, *Pel* fragility is markedly dependent on DNA replication direction.

Pel contains abundant runs of alternating pyrimidine-purine repeats (Fig. 3A and data S1), which can adopt alternative structures, such as Z-DNA, previously associated with deletions in bacteria, mice, and humans (*16*, *17*). Three stretches of ~15, ~20, and ~50 TG-dinucleotide repeats in marine *Pel* total ~170 bp (consistent with linking

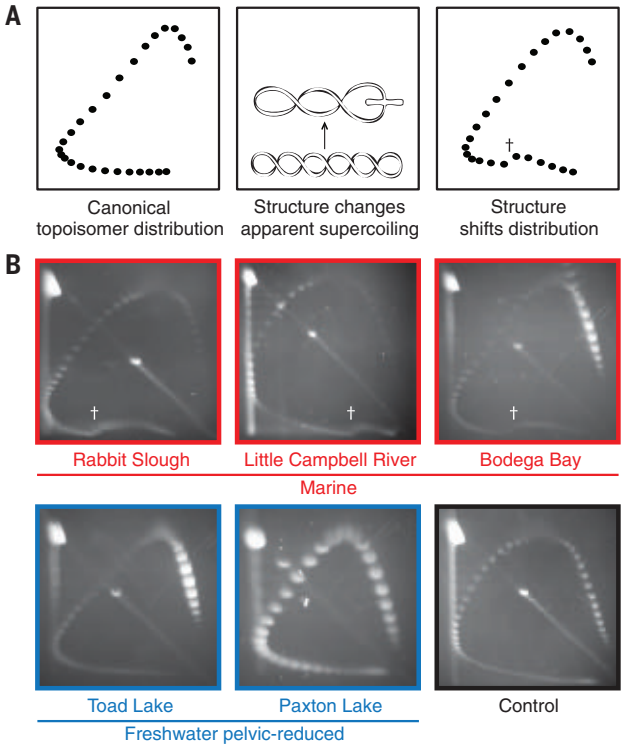


Fig. 1. Marine but not freshwater *Pel* alleles form alternative structures in vitro. (**A**) 2D electrophoresis of circular DNA topoisomers. A distribution of plasmid topoisomers is separated on an agarose gel; each topological class forms one spot. Canonical B-DNA forms a smooth distribution. Alternative structures cause mobility shifts. Distribution shifts at the linking number that induces alternative structure. Dagger symbol, mobility shift. (**B**) *Pel* from marine and freshwater pelvic-reduced populations. Control, *Atp1a1*.

[1]Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA, USA. [2]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA, USA. [3]Department of Biochemistry, Stanford University School of Medicine, Stanford, CA, USA. [4]Division of Pharmacology and Toxicology, University of Texas at Austin, Austin, TX, USA. [5]Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. [6]Department of Biology, University of Victoria, Victoria, BC, Canada. [7]School of Life Sciences, University of Nottingham, Nottingham, UK. [8]Department of Zoology, University of British Columbia, Vancouver, BC, Canada. [9]Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY, USA.
*Present address: University of California Museum of Paleontology, Berkeley, CA, USA.
†Corresponding author. Email: kingsley@stanford.edu

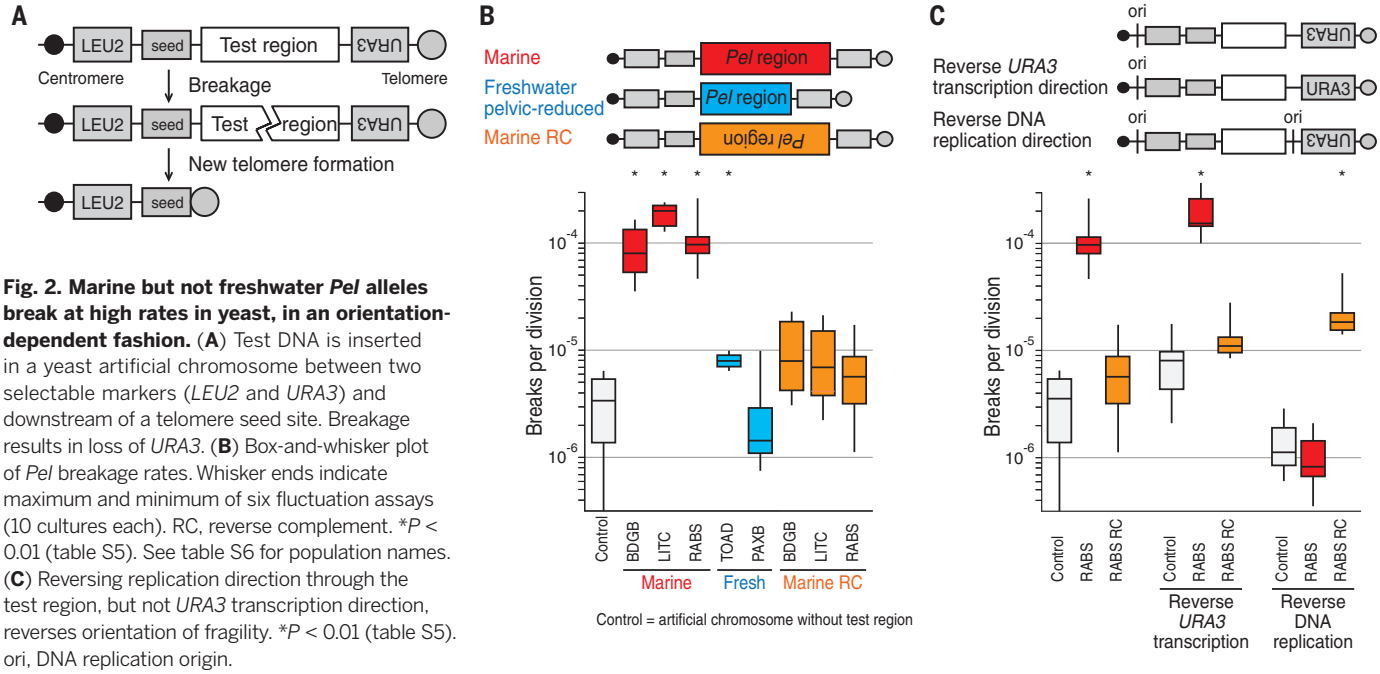


Fig. 2. Marine but not freshwater *Pel* alleles break at high rates in yeast, in an orientation-dependent fashion. (**A**) Test DNA is inserted in a yeast artificial chromosome between two selectable markers (*LEU2* and *URA3*) and downstream of a telomere seed site. Breakage results in loss of *URA3*. (**B**) Box-and-whisker plot of *Pel* breakage rates. Whisker ends indicate maximum and minimum of six fluctuation assays (10 cultures each). RC, reverse complement. *P < 0.01 (table S5). See table S6 for population names. (**C**) Reversing replication direction through the test region, but not *URA3* transcription direction, reverses orientation of fragility. *P < 0.01 (table S5). ori, DNA replication origin.

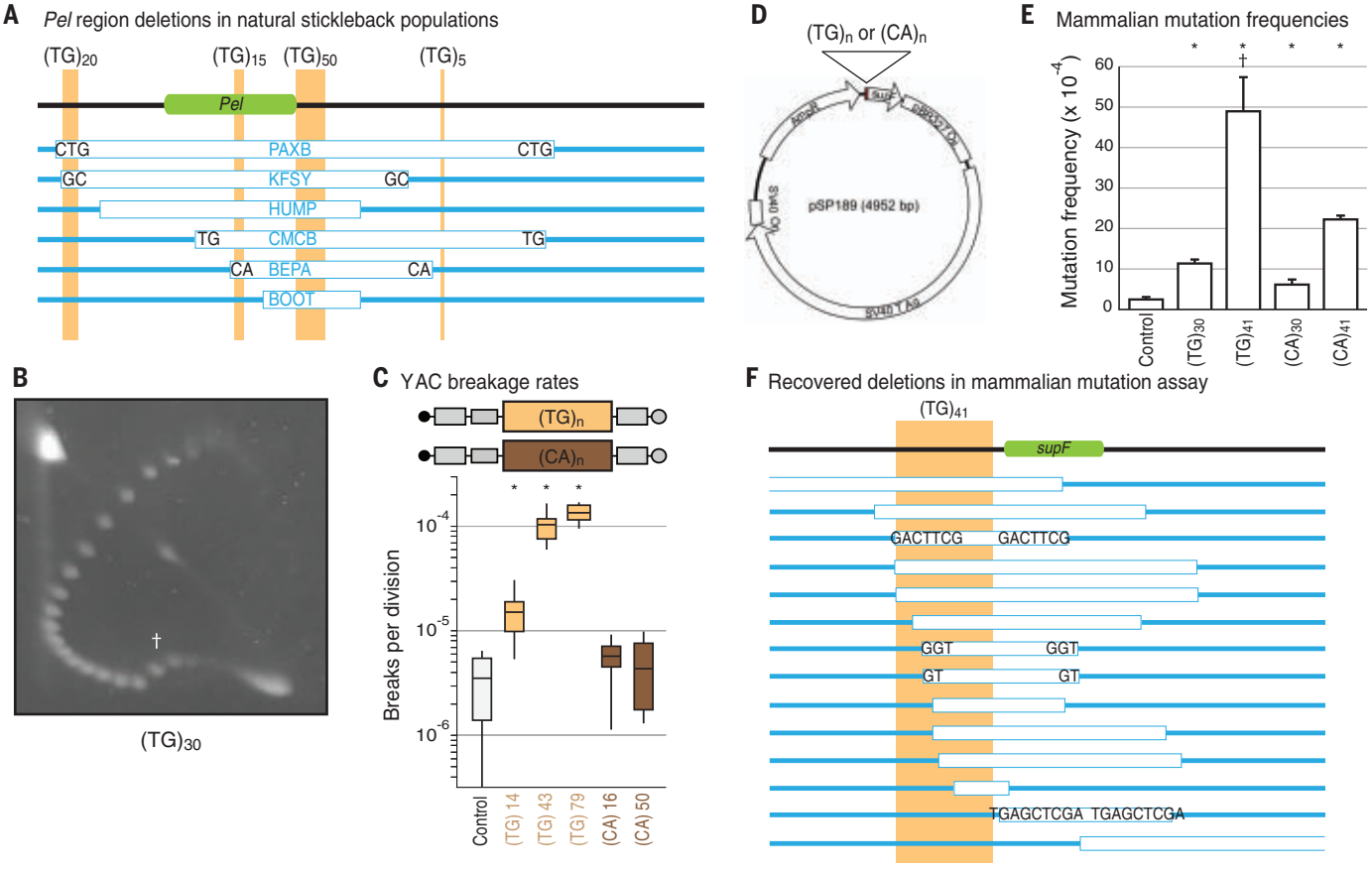


Fig. 3. TG-dinucleotide repeats recapitulate structure formation, high breakage rate, orientation dependence, and deletion spectrum. (**A**) To-scale maps of *Pel* in different freshwater pelvic-reduced populations (table S6). Green, *Pel* sequence driving pelvis expression (*6*). Tan, TG-repeats. White boxes, DNA deletions in indicated populations. Blue, DNA remaining. Letters indicate microhomologies at deletion junctions. (**B**) 2D gel for $(TG)_{30}$. Dagger symbol, mobility shift. (**C**) Yeast artificial chromosome (YAC) breakage rates for TG- or CA-repeats of varying lengths. *P < 0.01 (table S5). (**D**) Reporter shuttle plasmid schematic. (**E**) Mammalian mutation frequencies. Error bars indicate SEM of four or five independent experiments. *P < 0.05 (Student's t test). Dagger symbol, deletions dominate mutation spectrum (fig. S2A). (**F**) To-scale map of $(TG)_{41}$-induced deletions in mammalian cells.

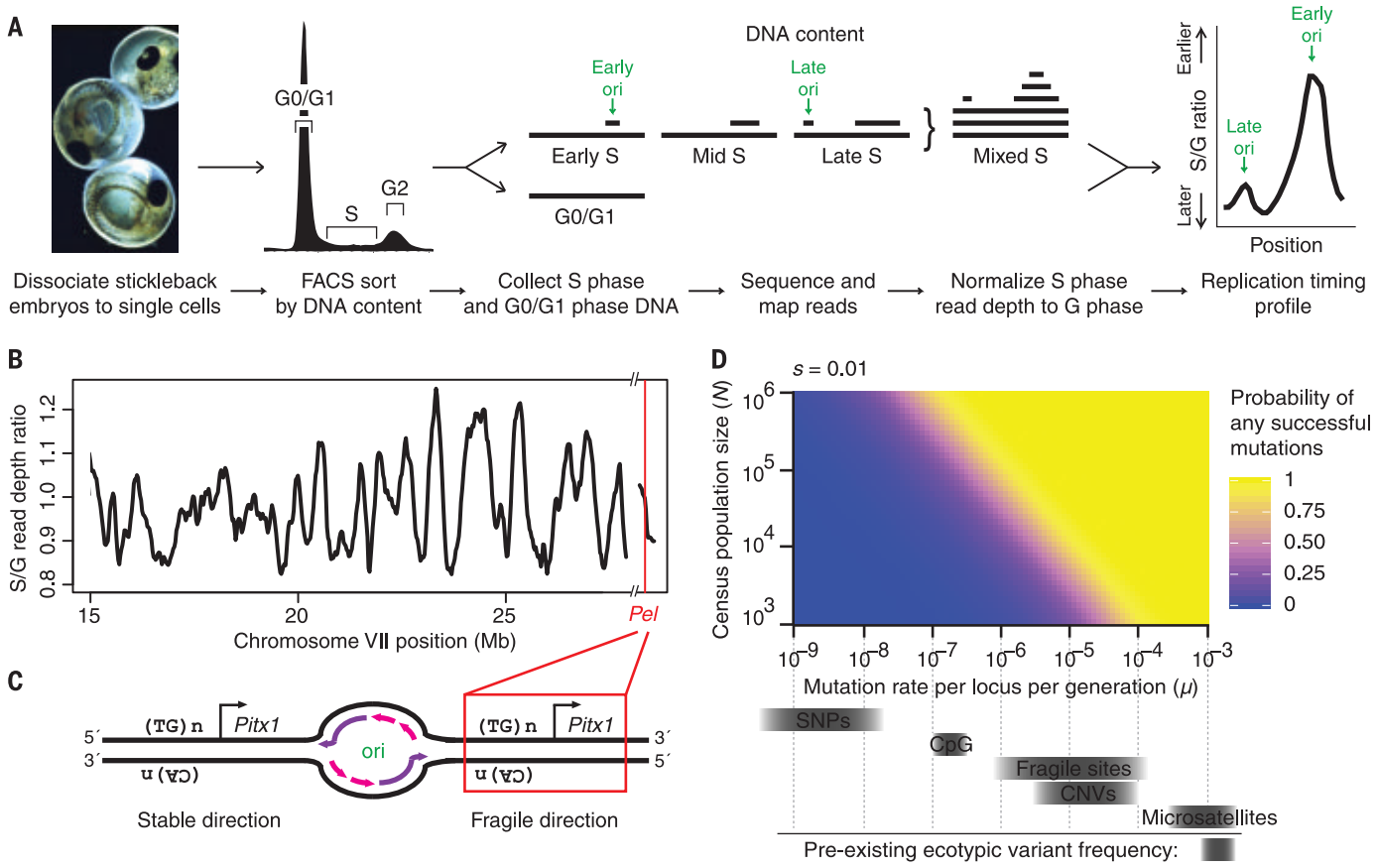


Fig. 4. *Pel* is located in the breakage-prone orientation in sticklebacks, generating a fragile site likely to contribute to parallel evolution in natural populations. (**A**) Workflow for profiling genome-wide replication timing. FACS, fluorescence-activated cell sorting. (**B**) Stickleback chromosome VII replication timing. Red line indicates the *Pel* locus, which is subtelomeric. Hash marks indicate reference genome assembly gap. (**C**) Diagram of stable and fragile replication orientations. Purple, newly synthesized leading strand; pink, newly synthesized lagging strand. (**D**) Probability of at least one de novo mutation arising at a particular locus in 10,000 generations and eventually becoming fixed, as a function of typical stickleback population sizes (N) and mutation rates (μ, gray bars) for single-nucleotide polymorphisms (SNPs), copy number variants (CNVs), and fragile sites. De novo point mutations are unlikely to occur and become fixed in small vertebrate populations, even when conferring a selective advantage (s = 0.01, modeled here). In contrast, mutations occurring at fragile sites are likely to arise and contribute to repeated evolution when conferring a selective advantage. For additional parameters, including neutrality (s = 0), see figs. S6 and S7.

number changes seen in the topoisomer assays above). TG-repeats alone induced mobility shifts in topoisomer assays (Fig. 3B) (*18*) and elevated chromosome breakage in yeast, with longer repeats stimulating more breaks (Fig. 3C). In contrast, both long and short versions of the reverse complement sequence (CA-repeats) were stable (Fig. 3C), recapitulating the orientation dependence of *Pel* fragility.

We also tested the effect of TG- and CA-repeats in mammalian COS-7 cells (Fig. 3D) (*19*). Dinucleotide repeats elevated mutation frequencies, with TG-repeats being more mutagenic than CA-repeats of comparable length, and longer repeats being more mutagenic than shorter repeats (Fig. 3E), in accordance with results from yeast assays. Mutations stimulated by the most mutagenic sequence, $(TG)_{41}$, were predominantly >100-bp deletions that removed part or all of the repeat and adjacent reporter gene (Fig. 3F and fig. S2A). Approximately 70% of deletion junctions contained microhomologies and insertions (Fig. 3F and fig. S2, A and B), consistent with error-prone microhomology-mediated end-joining repair and similar to junctions seen in stickleback pelvic-reduction alleles (*6*) (Fig. 3A). Ligation-mediated polymerase chain reaction suggested that breaks initiated near the dinucleotide repeats (fig. S2C). Taken together, our results indicate that TG-repeats form alternative DNA structures in vitro and can recapitulate the high mutation rates, orientation dependence, and propensity to stimulate breaks and deletions of the full *Pel* region.

To determine the orientation of *Pel* sequences relative to DNA replication in sticklebacks (Fig. 4A and fig. S3), we sequenced S- and G-phase cells from developing embryos and calculated S/G read-depth ratios to determine replication timing (*20*). *Pel* is located in a timing transition region (Fig. 4B and fig. S4), consistent with unidirectional replication. The replication direction through *Pel* matches the fragile orientation (Fig. 4C), suggesting that *Pel* would form a TG-repeat–associated fragile site in vivo. Experimental CRISPR targeting confirmed that initiation of breaks in *Pel* was sufficient to trigger local DNA deletions and macroscopic loss of pelvic structures in genetic crosses (fig. S5).

Could elevated mutation rates contribute to reuse of *Pel* deletions in parallel evolution? Population genetic modeling indicates that new mutations occurring at the low rates of typical single-nucleotide changes (~10^{-9} mutations per site per generation) would rarely arise at a particular locus in postglacial stickleback populations, whereas mutations occurring at elevated rates (~10^{-5} mutations per site per generation, for fragile sites) would arise often. When new mutations do occur, their subsequent fate is controlled by drift and selection (*21*). Neutral or small-effect point mutations will usually be lost or rise to fixation slowly, whereas deletions may cause larger phenotypic effects and can sweep if environmental conditions favor pelvic reduction (Fig. 4D and figs. S6 and S7). The combined effects on both the "arrival of the fittest" and the "survival of the fittest" may explain why recurrent *Pel* deletions are the predominant mechanism for evolving stickleback pelvic reduction. For other traits,

ancient standing variants provide an alternative way to overcome the demographic constraints of waiting for de novo mutations in small populations and can also lead to reuse of similar alleles in different populations (*22*, *23*).

The demographic parameters typical of sticklebacks apply to many vertebrates evolving with small population sizes or facing rapid environmental changes. For example, migration of modern humans out of Africa occurred with relatively small populations adapting to new environments in 3000 generations or fewer (*24*). Notably, nearly half of currently known mutations underlying adaptive traits in modern humans also appear to be produced by mechanisms with elevated mutation rates (table S1).

High mutation rates have been described at contingency loci in bacteria and other systems (*25*–*30*). Our study reveals an example of DNA fragility contributing to repeated morphological evolution in vertebrates. Our data also highlight several mechanisms that could alter local mutation rates, including expansion and contraction of TG-repeats, changes in sequence orientation, or changes in DNA replication. Natural variation in such parameters may affect the evolvability of different loci and the particular genetic paths likely to be taken when ecological conditions favor a given phenotype. The sequence features associated with DNA fragility in the *Pel* region are also found in thousands of other positions in stickleback and human genomes (fig. S8). Notably, TG-repeats are enriched in other loci that have undergone recurrent ecotypic deletions during marine-to-freshwater stickleback evolution (*31*) (table S2 and fig. S9) and are enriched near DNA breakage sites in humans (fig. S10). As causative changes are identified for a greater number of phenotypic traits, it will be interesting to see the extent to which DNA fragility has influenced the genes and mutations that underlie evolutionary change in nature.

REFERENCES AND NOTES

1. D. Schluter, E. A. Clifford, M. Nemethy, J. S. McKinnon, *Am. Nat.* **163**, 809–822 (2004).
2. D. L. Stern, V. Orgogozo, *Science* **323**, 746–751 (2009).
3. M. A. Bell, *Biol. J. Linn. Soc. Lond.* **31**, 347–382 (1987).
4. T. E. Reimchen, *Can. J. Zool.* **58**, 1232–1244 (1980).
5. M. A. Bell, G. Ortí, J. A. Walker, J. P. Koenings, *Evolution* **47**, 906–914 (1993).
6. Y. F. Chan *et al.*, *Science* **327**, 302–305 (2010).
7. M. Karhunen, J. Merilä, T. Leinonen, J. M. Cano, O. Ovaskainen, *Mol. Ecol. Resour.* **13**, 746–754 (2013).
8. B. Prud'homme *et al.*, *Nature* **440**, 1050–1053 (2006).
9. D. L. Stern, N. Frankel, *Philos. Trans. R. Soc. London Ser. B* **368**, 20130028 (2013).
10. R. G. Thys, C. E. Lehman, L. C. T. Pierce, Y.-H. Wang, *Curr. Genomics* **16**, 60–70 (2015).
11. R. Bowater, F. Aboul-Ela, D. M. Lilley, *Methods Enzymol.* **212**, 105–120 (1992).
12. A. Nordheim, A. Rich, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1821–1825 (1983).
13. A. Rich, A. Nordheim, A. H. Wang, *Annu. Rev. Biochem.* **53**, 791–846 (1984).
14. H. Zhang, C. H. Freudenreich, *Mol. Cell* **27**, 367–379 (2007).
15. A. Helmrich, M. Ballarino, L. Tora, *Mol. Cell* **44**, 966–977 (2011).
16. G. Wang, L. A. Christensen, K. M. Vasquez, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 2677–2682 (2006).
17. G. Wang, S. Carbajal, J. Vijg, J. DiGiovanni, K. M. Vasquez, *J. Natl. Cancer Inst.* **100**, 1815–1817 (2008).
18. H. Hamada, M. G. Petrino, T. Kakunaga, M. Seidman, B. D. Stollar, *Mol. Cell. Biol.* **4**, 2610–2621 (1984).
19. G. Wang, K. M. Vasquez, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 13448–13453 (2004).
20. N. Rhind, D. M. Gilbert, *Cold Spring Harb. Perspect. Biol.* **5**, a010132 (2013).
21. M. Kimura, *Genetics* **47**, 713–719 (1962).
22. P. F. Colosimo *et al.*, *Science* **307**, 1928–1933 (2005).
23. R. D. Barrett, D. Schluter, *Trends Ecol. Evol.* **23**, 38–44 (2008).
24. 1000 Genomes Project Consortium, *Nature* **526**, 68–74 (2015).
25. R. Moxon, C. Bayliss, D. Hood, *Annu. Rev. Genet.* **40**, 307–333 (2006).
26. A. Stoltzfus, L. Y. Yampolsky, *J. Hered.* **100**, 637–647 (2009).
27. X. Du *et al.*, *Nucleic Acids Res.* **42**, 12367–12379 (2014).
28. S. C. Galen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13958–13963 (2015).
29. A. Bacolla, J. A. Tainer, K. M. Vasquez, D. N. Cooper, *Nucleic Acids Res.* **44**, 5673–5688 (2016).
30. A. D. Hargreaves *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 7677–7682 (2017).
31. C. B. Lowe *et al.*, *Genome Res.* **28**, 256–265 (2018).

ACKNOWLEDGMENTS

We thank V. Tien, J. Le, M. Yau, M. Thakur, A. Muralidharan, M. Whitlock, B. Belotserkovskii, R. Driscoll, K. Cimprich, J. Wang, S. Quake, and A. Casper for experimental assistance or advice; R. Daugherty, J. Rollins, B. Lohman, R. Mollenhauer, M. Reyes, and F. von Hippel for help with fieldwork; C. Freudenreich for yeast strains; and Z. Weng and B. Carter for help with high-throughput sequencing and cell sorting. **Funding:** NIH grants 5P50HG2568 (D.M.K.), CA093729 (K.M.V.), and 2T32GM007790 (J.I.W.); NSF grant DEB0919184 (M.A.B.); NSF and Stanford CEHG Graduate Fellowships (K.T.X.); NIH Predoctoral Fellowship (A.C.T.); HHMI investigator (D.M.K.). **Author contributions:** K.T.X. and D.M.K. designed the study. K.T.X., G.W., A.C.T., and J.I.W. performed experiments. K.T.X., G.W., A.C.T., D.S., K.M.V., and D.M.K. analyzed data. T.E.R., A.D.C.M., D.S., and M.A.B. provided key populations and comments. K.T.X. and D.M.K. wrote the paper with input from all authors. **Competing interests:** None declared. **Data and materials availability:** Raw sequencing data and processed S/G read-depth ratio data have been deposited at GEO accession GSE121537.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/363/6422/81/suppl/DC1
Materials and Methods
Figs. S1 to S11
Tables S1 to S6
References (*32*–*61*)
Data S1

10 March 2017; resubmitted 18 April 2018
Accepted 28 November 2018
10.1126/science.aan1425

PROTEIN TRANSLOCATION

Structure of the posttranslational Sec protein-translocation channel complex from yeast

Samuel Itskanov[1] and Eunyong Park[2]*

The Sec61 protein-conducting channel mediates transport of many proteins, such as secretory proteins, across the endoplasmic reticulum (ER) membrane during or after translation. Posttranslational transport is enabled by two additional membrane proteins associated with the channel, Sec63 and Sec62, but its mechanism is poorly understood. We determined a structure of the Sec complex (Sec61-Sec63-Sec71-Sec72) from *Saccharomyces cerevisiae* by cryo–electron microscopy (cryo-EM). The structure shows that Sec63 tightly associates with Sec61 through interactions in cytosolic, transmembrane, and ER-luminal domains, prying open Sec61's lateral gate and translocation pore and thus activating the channel for substrate engagement. Furthermore, Sec63 optimally positions binding sites for cytosolic and luminal chaperones in the complex to enable efficient polypeptide translocation. Our study provides mechanistic insights into eukaryotic posttranslational protein translocation.

The eukaryotic Sec61 or prokaryotic SecY complex forms a universally conserved protein-conducting channel that is essential for biogenesis of many proteins (*1*–*3*). The channel mediates transport of soluble (e.g., secretory) proteins across the eukaryotic endoplasmic reticulum (ER) membrane or the prokaryotic plasma membrane through its water-filled pore and integration of membrane proteins into the lipid phase through its lateral gate. The Sec61/SecY channel consists of an hourglass-shaped α subunit, which contains 10 transmembrane segments (TMs 1 to 10), and two small β and γ subunits, which are single-pass membrane proteins in eukaryotes (*4*). Often, translocation is coupled with translation (i.e., cotranslational translocation) by direct docking of a translating ribosome onto the channel. The channel also translocates many proteins in a posttranslational manner, the mechanisms of which differ between eukaryotes and prokaryotes. In eukaryotes, posttranslational translocation requires two essential membrane proteins, Sec63 and Sec62, which associate with the channel (*5*–*8*), and the ER-resident Hsp70 chaperone BiP, which grasps the substrate polypeptide in the ER lumen and prevents it from backsliding to the cytosol (*9*–*12*). In fungal species, the complex (hereafter referred to as the Sec complex) is further associated with the nonessential Sec71

[1]Biophysics Graduate Program, University of California, Berkeley, Berkeley, CA 94720, USA. [2]Department of Molecular and Cell Biology and California Institute for Quantitative Biosciences, University of California, Berkeley, Berkeley, CA 94720, USA.
***Corresponding author. Email: eunyong_park@berkeley.edu**

and Sec72 subunits (*10*, *11*, *13*). The molecular architecture of the Sec complex and the functions of its subunits are poorly defined.

To gain insight into Sec-mediated protein translocation, we determined a structure of the *Saccharomyces cerevisiae* Sec complex at 3.7-Å resolution by cryo–electron microscopy (cryo-EM) (Fig. 1 and figs. S1 and S2). Many side chains are clearly visible in the density map, enabling modeling of an accurate atomic structure (Fig. 1B and fig. S2C). The map also allowed us to improve the model for the eukaryotic Sec61 channel, which was previously built into maps at ~4- to 5-Å local resolutions (*14*, *15*). However, Sec62 and the ER-luminal J domain of Sec63, which transiently interacts with BiP (*9*–*11*, *16*), were not sufficiently resolved for model building, likely because of their flexible motions (Fig. 1A). The structure reveals that Sec63 together with Sec71-Sec72 forms a large soluble domain, which sits on the cytosolic side of the Sec61 channel (Fig. 1). Sec63 consists of an N-terminal domain containing three TMs and a J domain between the second and third TMs and a C-terminal cytosolic domain (Fig. 2, A and B). The cytosolic domain contains two α helical domains (HD1 and HD2) and an immunoglobulin-like [fibronectin type-III (FN3)] domain, which are arranged similarly to the homologous region of the Brr2 RNA helicase (*17*) (fig. S3). Sec71-Sec72, the structure of which is similar to a recent crystal structure of *Chaetomium thermophilum* Sec71-Sec72 (*18*), clamps Sec63's cytosolic domain like tongs (fig. S4).

Sec63 makes extensive contacts with the channel through its transmembrane, cytosolic, and luminal domains, indicative of a major role in regulating the channel's function (Fig. 2, C to E). In the membrane region, the TMs of Sec63 are located at the back (opposite from the lateral gate) of the Sec61 channel, interacting with the TMs of Sec61β and Sec61γ as well as TM1 and TM5 of Sec61α (Fig. 2C). Considering the extensive interactions between these elements, the TMs of Sec63 likely make a main contribution to the association between Sec61 and the rest of the Sec complex. In the cytosolic region, the FN3 domain of Sec63 interacts with the loop between TM6 and TM7 (L6/7) of Sec61α through antigen-antibody–like binding. Like other FN3 domains, FN3 of Sec63 has a canonical β-sandwich fold composed of seven β strands (referred to as A to G) but contains unusually long A-B, B-C, and D-E interstrand loops (fig. S3, B and C). With both A-B and B-C loops, FN3 creates a binding surface for L6/7, which uses a combination of surface complementarity and electrostatic and hydrophobic interactions (Fig. 2E and fig. S3D). Although sequence conservation is not obvious, metazoan Sec63s have similar extensions in the A-B and B-C loops. We expect analogous interactions between Sec63 and Sec61 in other eukaryotes. The interaction between FN3 and L6/7 is noteworthy because L6/7, together with L8/9, forms a docking site for the ribosome (*14*, *19*, *20*) (fig. S5A). Accordingly, superimposition of the Sec complex with a ribosome-bound Sec61 structure shows massive steric clashes between the ribosome and the cytosolic domains of Sec63 and Sec62 (fig. S5B), explaining why Sec61 in the Sec complex cannot bind to the ribosome (*7*, *11*). In the ER luminal side, a segment preceding TM3 of Sec63 is directed into the luminal funnel of the Sec61 channel through the crevice present between TM5 of Sec61α and the TM of Sec61γ (Fig. 2D). This segment makes an antiparallel β sheet together with a β hairpin looping out in the middle of Sec61α's TM5. This β-augmentation is further buttressed by hydrophobic interactions with the N-terminal segment of Sec63. These features are highly conserved throughout eukaryotes and thus likely play an important role in optimal positioning of the J domain.

One pronounced feature of the Sec complex structure is a fully open channel (Fig. 3, A and B). The Sec61/SecY channel has a characteristic clamshell-like topology, in which its central pore can open toward the lipid phase through the lateral gate formed between TM2 and TM7. Compared with previous Sec61/SecY structures (*4*, *14*, *21*–*24*), the channel in the Sec complex displays a substantially wider opening at its lateral gate, through which a signal sequence can readily pass as an α helix (Fig. 3 and fig. S6). This contrasts with structures of channels associated with the ribosome or the bacterial posttranslational translocation motor SecA (*14*, *21*–*24*), in which the channel shows an only partially open lateral gate (Fig. 3, C to F), which was proposed to be further opened by interaction with the hydrophobic signal sequence during the initial substrate insertion. The opening is achieved by a largely rigid-body movement between the two halves (TMs 1 to 5 and 6 to 10) of Sec61α and additional motions of the lateral gate helices. The fully open conformation appears to be a result of the extensive interactions with Sec63. For example, binding between FN3 and L6/7 perhaps pulls the C-terminal half of Sec61α to open the lateral gate. However, further investigation will be necessary to understand the precise mechanism and the dynamics of channel gating in the native membrane environment. At the open lateral gate slit, there is a weak density feature, which likely represents bound detergent molecules (Fig. 3, A and B). In the native membrane, lipid molecules may occupy this site and facilitate initial binding of signal sequences.

Our channel structure likely also represents a fully open state of the translocation pore (Fig. 3B and fig. S7). The radius of the pore constriction is ~3 Å, large enough to pass an extended polypeptide chain. The opening would also permit passage of small hydrated ions and polar molecules in the absence of a translocating polypeptide (*25*, *26*), although the relatively positive electrostatic potential around the pore may disfavor permeation of positively charged species (fig. S7C). Yeast Sec61 has a relatively less hydrophobic pore constriction compared with nonfungal Sec61 and prokaryotic SecY (fig. S7D). In prokaryotes, reduction of hydrophobicity in the pore constriction has been shown to lead to membrane potential dissipation (*26*), and similarly, in higher eukaryotes it might cause calcium leakage from the ER. However, yeast may tolerate ion leakage because calcium is stored primarily in the vacuole. In resting or primed channels, the pore is closed or narrow (<2 Å in radius) and further blocked

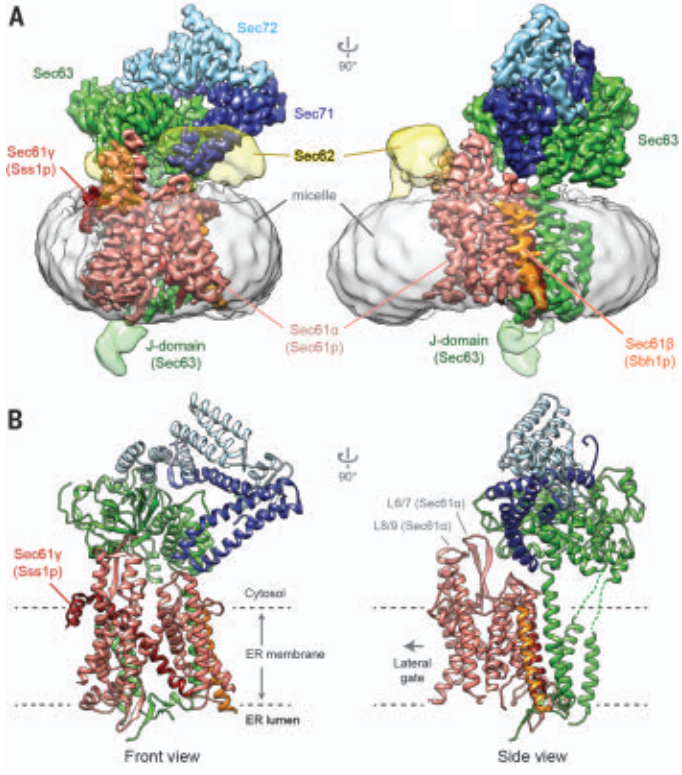


Fig. 1. Structure of the yeast Sec complex. (**A**) Cryo-EM density map and (**B**) atomic model of the yeast post-translational protein translocation complex. The front view is a view into the lateral gate.

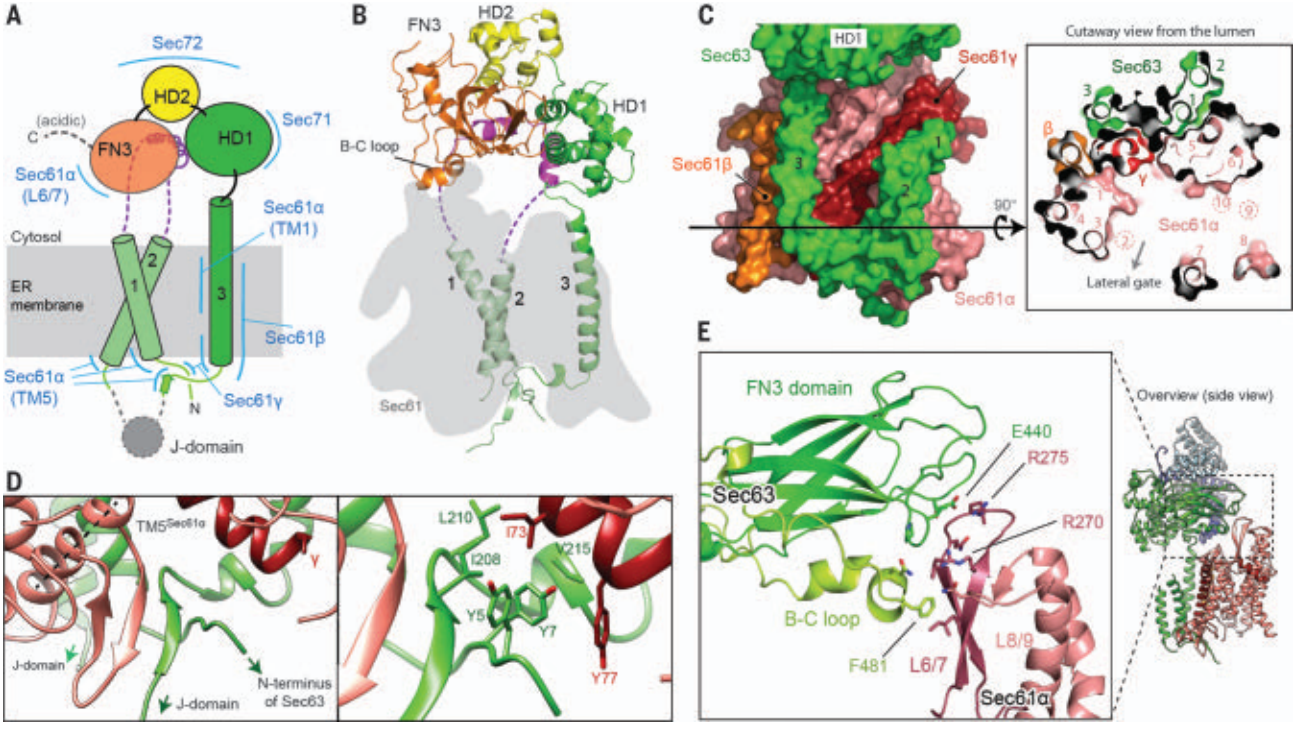


Fig. 2. Structure of Sec63 and its interactions with the channel. (**A**) Schematic of Sec63 domains. Regions interacting with other parts of the complex are indicated by blue lines. Unmodeled regions are shown with dashed lines. (**B**) Structure of Sec63 (front view). The position of Sec61 is shaded in gray. (**C**) Interactions between TMs of Sec63 and Sec61. On the left is a view from the back; on the right is a cutaway view from the ER lumen. The black arrowed line represents the cross-sectional plane. TMs 2, 9, and 10 of Sec61α are located above the cross-sectional plane. (**D**) Interactions between Sec63 and Sec61 in the luminal side. On the left is a β sheet formed between Sec61α (TM5 indicated by a dashed line) and the segment between Sec63 TM3 and the J domain. On the right is a magnified view with side chains shown as sticks. (**E**) Interactions between the FN3 domain and the cytosolic loop L6/7 of Sec61α (also see Fig. 1B). L, Leu; I, Ile; V, Val; Y, Tyr; E, Glu; R, Arg; F, Phe.

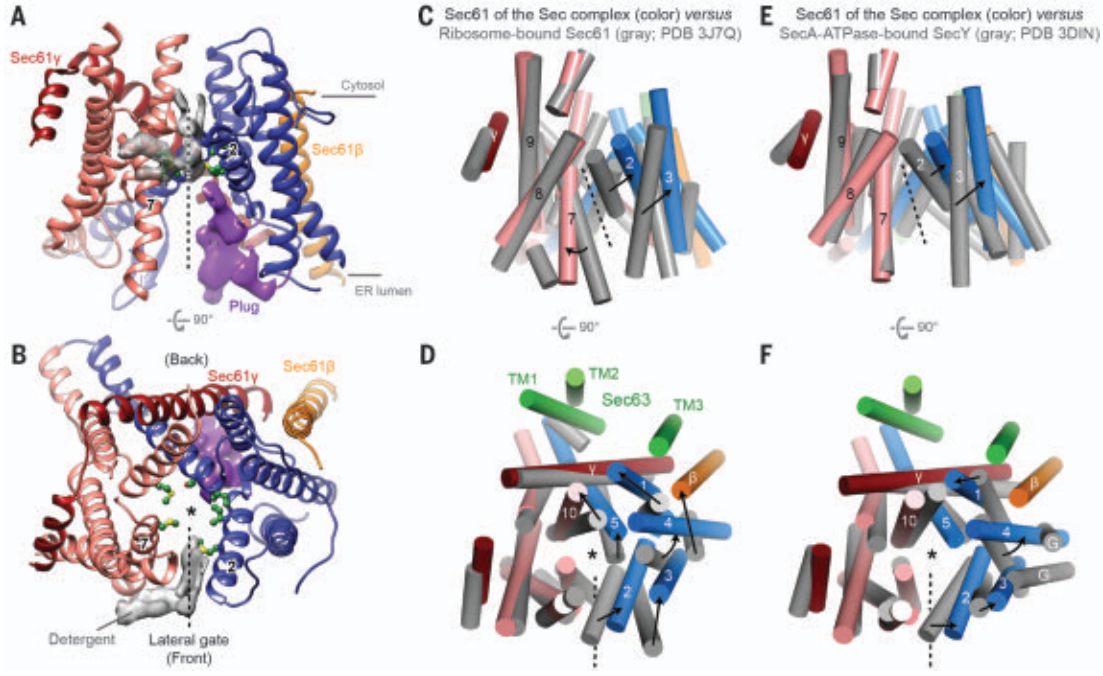


Fig. 3. A fully opened Sec61 channel in the Sec complex. (**A** and **B**) Structure of the Sec61 channel. The N- and C-terminal halves of Sec61α are in blue and salmon, respectively. The gray density feature is presumed detergent molecules. Pore-lining residues are shown as green balls and sticks. The density feature for the plug is in purple. Numbers "2" and "7" indicate TM2 and TM7, respectively. (**C** to **F**) Comparison of Sec61 of the Sec complex (colored) with Sec61 of the cotranslational ribosome-Sec61 complex [gray; (C) and (D)] or SecY of a bacterial post-translational SecA-SecY channel complex [gray; (E) and (F)]. The structures are aligned with respect to the C-terminal half of Sec61α [(C) to (F)]. Shown are the front [(A), (C), and (E)] and cytosolic [(B), (D), and (F)] views. Numbers indicate corresponding TMs. Dashed lines represent the lateral gate. Asterisks indicate the translocation pore. For simplicity, L6/7 and L8/9 of Sec61α are not shown. In (D) and (F), TMs of Sec63 are also shown (green). In (F), TMs of SecG are indicated by "G." Also see fig. S6 for comparisons to archaeal SecY and substrate-engaged channels. ATPase, adenosine triphosphatase; PDB, Protein Data Bank.

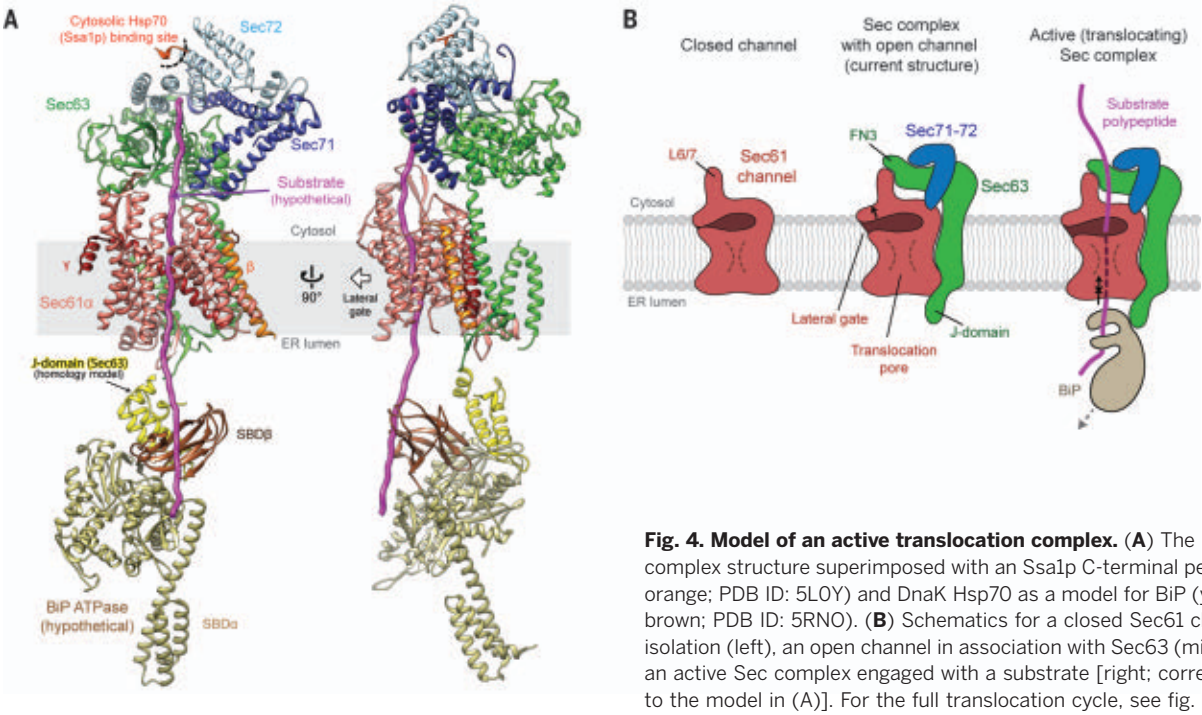


Fig. 4. Model of an active translocation complex. (**A**) The Sec complex structure superimposed with an Ssa1p C-terminal peptide (red orange; PDB ID: 5L0Y) and DnaK Hsp70 as a model for BiP (yellow and brown; PDB ID: 5RNO). (**B**) Schematics for a closed Sec61 channel in isolation (left), an open channel in association with Sec63 (middle), and an active Sec complex engaged with a substrate [right; corresponding to the model in (A)]. For the full translocation cycle, see fig. S8.

by a small α-helical plug in the luminal funnel (*4*, *14*, *21*). By contrast, in our structure, the plug seems flexible and displaced from the pore (Fig. 3, A and B).

The spatial arrangement of Sec63 and Sec71-Sec72 with respect to the Sec61 channel suggests how these components play roles in accepting a polypeptide substrate from a cytosolic chaperone and handing it over to the channel and subsequently to BiP. Studies of *C. thermophilum* Sec72 have suggested that Sec72 provides a docking site for the cytosolic Hsp70 chaperone Ssa1p (*18*), which prevents substrates from premature folding or aggregation before translocation (*6*). Superimposition of the cocrystal structure of Sec72 and an Ssa1p C-terminal tail shows that the Ssa1p-binding site is ~60 Å above the channel's pore (Fig. 4A). While the cytosolic domain of Sec63-71-72 sits on top of Sec61, its position is tilted such that the polypeptide can insert straight down to the pore. Similarly, Sec62 is also positioned off the translocation path (Fig. 1A). Thus, upon release from Ssa1p, a substrate would efficiently engage with the pore without obstruction. The structure also allows us to propose how BiP Hsp70 may catch the substrate in the ER lumen. Despite the low resolution of the J domain (Fig. 1A), we could dock a homology model into the EM density map based on the shape of the feature and the orientations of the flanking segments (Fig. 4A). We then superimposed a recent crystal structure of a bacterial J domain–Hsp70 complex (*27*) to our EM structure (Fig. 4A). This modeling exercise showed that a peptide-binding cleft of the Hsp70 [called substrate-binding domain β (SBDβ)] would be placed directly below the translocation pore. Thus, the J domain seems optimally positioned to allow BiP to grasp the substrate polypeptide as it emerges from the channel.

Our structure offers a model for how Sec63 enables posttranslational translocation (Fig. 4B and fig. S8) and provides a more complete picture of how the Sec61/SecY channel works together with different binding partners (i.e., ribosomes, Sec63, or SecA) to enable transport of a range of substrates. Association of Sec63 seems to induce full opening of the channel, a conformation in which the channel can readily accept a substrate polypeptide. Such a conformation, compared with a partially open channel seen with the other modes, is likely advantageous for many posttranslational-specific substrates, which tend to have a less hydrophobic signal sequence (*28*–*30*).

REFERENCES AND NOTES

1. E. Park, T. A. Rapoport, *Annu. Rev. Biophys.* **41**, 21–40 (2012).
2. R. M. Voorhees, R. S. Hegde, *Curr. Opin. Cell Biol.* **41**, 91–99 (2016).
3. E. C. Mandon, S. F. Trueman, R. Gilmore, *Cold Spring Harb. Perspect. Biol.* **5**, a013342 (2013).
4. B. Van den Berg *et al.*, *Nature* **427**, 36–44 (2004).
5. J. A. Rothblatt, R. J. Deshaies, S. L. Sanders, G. Daum, R. Schekman, *J. Cell Biol.* **109**, 2641–2652 (1989).
6. R. J. Deshaies, S. L. Sanders, D. A. Feldheim, R. Schekman, *Nature* **349**, 806–808 (1991).
7. H. A. Meyer *et al.*, *J. Biol. Chem.* **275**, 14550–14557 (2000).
8. J. Tyedmers *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7214–7219 (2000).
9. D. Feldheim, J. Rothblatt, R. Schekman, *Mol. Cell. Biol.* **12**, 3288–3296 (1992).
10. J. L. Brodsky, R. Schekman, *J. Cell Biol.* **123**, 1355–1363 (1993).
11. S. Panzner, L. Dreier, E. Hartmann, S. Kostka, T. A. Rapoport, *Cell* **81**, 561–570 (1995).
12. K. E. Matlack, B. Misselwitz, K. Plath, T. A. Rapoport, *Cell* **97**, 553–564 (1999).
13. N. Green, H. Fang, P. Walter, *J. Cell Biol.* **116**, 597–604 (1992).
14. R. M. Voorhees, I. S. Fernández, S. H. Scheres, R. S. Hegde, *Cell* **157**, 1632–1643 (2014).
15. K. Braunger *et al.*, *Science* **360**, 215–219 (2018).
16. K. E. Matlack, K. Plath, B. Misselwitz, T. A. Rapoport, *Science* **277**, 938–941 (1997).
17. T. H. Nguyen *et al.*, *Structure* **21**, 910–919 (2013).
18. A. Tripathi, E. C. Mandon, R. Gilmore, T. A. Rapoport, *J. Biol. Chem.* **292**, 8007–8018 (2017).
19. Z. Cheng, Y. Jiang, E. C. Mandon, R. Gilmore, *J. Cell Biol.* **168**, 67–77 (2005).
20. T. Becker *et al.*, *Science* **326**, 1369–1373 (2009).
21. J. Zimmer, Y. Nam, T. A. Rapoport, *Nature* **455**, 936–943 (2008).
22. P. F. Egea, R. M. Stroud, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 17182–17187 (2010).
23. R. M. Voorhees, R. S. Hegde, *Science* **351**, 88–91 (2016).
24. L. Li *et al.*, *Nature* **531**, 395–399 (2016).
25. D. Heritage, W. F. Wonderlin, *J. Biol. Chem.* **276**, 22655–22662 (2001).
26. E. Park, T. A. Rapoport, *Nature* **473**, 239–242 (2011).
27. R. Kityk, J. Kopp, M. P. Mayer, *Mol. Cell* **69**, 227–237.e4 (2018).
28. D. T. Ng, J. D. Brown, P. Walter, *J. Cell Biol.* **134**, 269–278 (1996).
29. M. A. Smith, W. M. Clemons Jr., C. J. DeMars, A. M. Flower, *J. Bacteriol.* **187**, 6454–6465 (2005).
30. S. F. Trueman, E. C. Mandon, R. Gilmore, *J. Cell Biol.* **199**, 907–918 (2012).

ACKNOWLEDGMENTS

We thank D. Toso for help with electron microscope operation and J. Hurley, S. Brohawn, and K. Tucker for critical reading of the manuscript. **Funding:** This work was funded by UC Berkeley (E.P.) and an NIH training grant (T32GM008295; S.I.). **Author contributions:** S.I. and E.P. performed experiments, interpreted results, and wrote the manuscript; E.P. conceived and supervised the project. **Competing interests:** None declared. **Data and materials availability:** The cryo-EM density maps and atomic model have been deposited in EM Data Bank (accession code: EMD-0336) and Protein Data Bank (accession code: 6N3Q), respectively.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/363/6422/84/suppl/DC1
Materials and Methods
Figs. S1 to S8
Table S1
References (*31*–*40*)

9 October 2018; accepted 21 November 2018
Published online 13 December 2018
10.1126/science.aav6740

PROKARYOTIC IMMUNITY

Functionally diverse type V CRISPR-Cas systems

Winston X. Yan[1]*, Pratyusha Hunnewell[1]*, Lauren E. Alfonse[1], Jason M. Carte[1], Elise Keston-Smith[1], Shanmugapriya Sothiselvam[1], Anthony J. Garrity[1], Shaorong Chong[1], Kira S. Makarova[2], Eugene V. Koonin[2], David R. Cheng[1], David A. Scott[1]†

Type V CRISPR-Cas systems are distinguished by a single RNA-guided RuvC domain-containing effector, Cas12. Although effectors of subtypes V-A (Cas12a) and V-B (Cas12b) have been studied in detail, the distinct domain architectures and diverged RuvC sequences of uncharacterized Cas12 proteins suggest unexplored functional diversity. Here, we identify and characterize Cas12c, -g, -h, and -i. Cas12c, -h, and -i demonstrate RNA-guided double-stranded DNA (dsDNA) interference activity. Cas12i exhibits markedly different efficiencies of CRISPR RNA spacer complementary and noncomplementary strand cleavage resulting in predominant dsDNA nicking. Cas12g is an RNA-guided ribonuclease (RNase) with collateral RNase and single-strand DNase activities. Our study reveals the functional diversity emerging along different routes of type V CRISPR-Cas evolution and expands the CRISPR toolbox.

Competition between prokaryotes and viruses has led to the evolution of diverse defense strategies, with more being identified through the mining of growing genomic and metagenomic sequence databases (*1*–*3*). Class 2 CRISPR-Cas systems are of particular interest, because their programmable single-effector nucleases have enabled genome engineering and nucleic acid detection tools (*4*–*8*). Class 2 systems include types II, V, and VI, which are based on Cas9, Cas 12, and Cas13 effectors, respectively (*9*–*11*). Cas9 contains an HNH nuclease domain inserted into a RuvC nuclease domain (*12*–*14*), and the two domains together cleave double-stranded DNA (dsDNA). Cas12 contains a single RuvC nuclease domain that cleaves dsDNA adjacent to protospacer adjacent motif (PAM) sequences (*15*) and single-stranded DNA (ssDNA) nonspecifically (*16*). Cas13 contains two HEPN domains that cleave RNA exclusively (*10*, *17*, *18*).

We aggregated more than 10 terabytes of sequence data and generated a database of 293,985 putative CRISPR-Cas systems (*19*). From this database we identified type V systems with predicted effectors ranging in size from 720 to 1093 amino acids, each of which contained a C-terminal RuvC domain (fig. S1). The classification tree of type V effectors splits into three major branches: (i) Cas12a, -c, -d, and -e; (ii) Cas12b and its distant homologs; and (iii) subtype V-U variants closely related to transposon-encoded TnpB. Predicted type V effectors in this study showed weak sequence similarity ($E > 10^{-3}$) with previously characterized ones. Combined with differences in locus organization and subsequently uncovered functional differences, our work supports the assignment of separate subtypes, V-G, V-H, and V-I (Fig. 1A, fig. S1, and table S1). Whereas Cas12h and -i cluster with Cas12b, albeit at a large evolutionary distance, Cas12g clusters with the predicted subtype V-U effectors and TnpBs (Fig. 1A). The subtype V-U effectors, including the recently identified Cas14a, -b, and -c (subtype V-F), are much smaller than the typical CRISPR effectors and show greater similarity to TnpB (*11*, *20*). Cas14a and Cas12g appear to have evolved from distinct TnpB ancestors (Fig. 1A). Thus, experimental characterization of subtype V-G is of particular interest to elucidate the routes of evolution of TnpB proteins into functional CRISPR effectors.

To functionally characterize the type V-G, -H, and -I systems, we used an *Escherichia coli* negative selection screen, in which RNA-guided interference activity of reconstituted CRISPR-Cas systems reduces bacterial viability at 37°C (*19*). Each screen included: (i) an effector plasmid carrying predicted Cas genes; (ii) a CRISPR array library targeting pACYC184 and *E. coli* essential genes; and (iii) a noncoding plasmid containing concatenated *cas* gene-flanking noncoding sequences for the unbiased detection of trans-activating crRNA (tracrRNA) elements (Fig. 1, B and C).

In vivo screening of the compact subtype V-G effector, Cas12g1 (767 amino acids), revealed interference activity that specifically targeted the sense DNA strand of actively transcribed substrate regions (Fig. 2A, fig. S2A, and table S2). Analysis of target-flanking sequences revealed no PAM requirements for interference (fig. S2, B to D). Mutation of the RuvC-I motif of Cas12g1 [Asp^{513}→Ala (D513A)] or omission of the noncoding plasmid substantially decreased interference activity (Fig. 2B and fig. S2, E to G). RNA sequencing of screen samples revealed a tracrRNA expressed from the noncoding plasmid and a mature crRNA from the CRISPR array library (Fig. 2, C and D, and fig. S3). However, purified Cas12g1 was incapable of processing its pre-crRNA in vitro with or without tracrRNA, suggesting that additional endogenous factors are required for in vivo crRNA biogenesis (figs. S4 and S5 and table S3).

By investigating the mechanism of in vivo interference by subtype V-G systems, we found that ternary complexes consisting of Cas12g1, tracrRNA, and in vivo screen–validated crRNAs showed no cleavage of cognate ssDNA or dsDNA substrates at 37°C (fig. S6, A and B, and tables S4 to S8). The Cas12g1 locus originates from a hot spring metagenome but, although the ternary complex is thermostable [complex melting temperature (T_m) = 74°C] (fig. S7), we observed no ssDNA or dsDNA cleavage at 42°C, 50°C, or 60°C (fig. S6, C to H). Given the transcriptional association of Cas12g1 in vivo interference indicated by the preferential targeting of sense-strand DNA (Fig. 2A), we assessed cleavage of sense ssDNA (containing crRNA spacer-complementary target) or antisense ssDNA substrates in the presence of sense or antisense RNA transcripts. The Cas12g1 ternary complex efficiently cleaved the sense ssDNA in the presence of sense RNA (hereafter, target RNA), and this activity increased in efficiency from 37° to 50°C (Fig. 2E and fig. S8A). No ssDNA cleavage was observed for any other DNA-RNA substrate combination (fig. S8, B to D). In the presence of target RNA, Cas12g1 ternary complex also cleaved unrelated collateral ssDNA (Fig. 2E and fig. S9), demonstrating that target RNA activates nonspecific collateral ssDNA cleavage in trans by Cas12g1.

The weak ssDNA cleavage observed at 37°C is likely not responsible for the robust Cas12g1 interference activity observed in vivo. Thus, we investigated the intrinsic ribonuclease (RNase) activity of Cas12g1 and observed strong target RNA cleavage with the ternary complex at 37°C (Fig. 2F), and this was further enhanced at 50°C (fig. S10 and tables S4 to S8). At 50°C, detectable target RNA cleavage was observed at ternary complex concentrations as low as 125 pM (fig. S11A), with no background cleavage of nontarget RNA at the highest complex concentration tested (250 nM) (figs. S10B and S11B). Cas12g1 ternary complex also cleaved dye-labeled collateral RNA accompanying unlabeled target RNA at target concentrations as low as 100 pM, demonstrating that the stand-alone RNA detection sensitivity of Cas12g1 is comparable to that of the highest performing Cas13 variants (Fig. 2F and fig. S11, C and D) (*21*). Both RNA and ssDNA cleavage by Cas12g1 are metal ion dependent and require an intact RuvC domain that was previously known to cleave only DNA (Fig. 2, G and H, and figs. S4 and S12). The thermostability and nucleic acid detection sensitivity of Cas12g1 has the potential to enhance the performance and durability of nucleic acid diagnostic methods, such as SHERLOCK and DETECTR (*16*, *18*, *22*). Additionally, the small size of Cas12g1 is likely to facilitate delivery for diverse in vivo transcriptome engineering applications (*23*, *24*).

We next investigated subtype V-H and V-I systems containing effectors Cas12h (870 to 933

[1]Arbor Biotechnologies, Cambridge, MA 02139, USA.
[2]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: dscott@arbor.bio

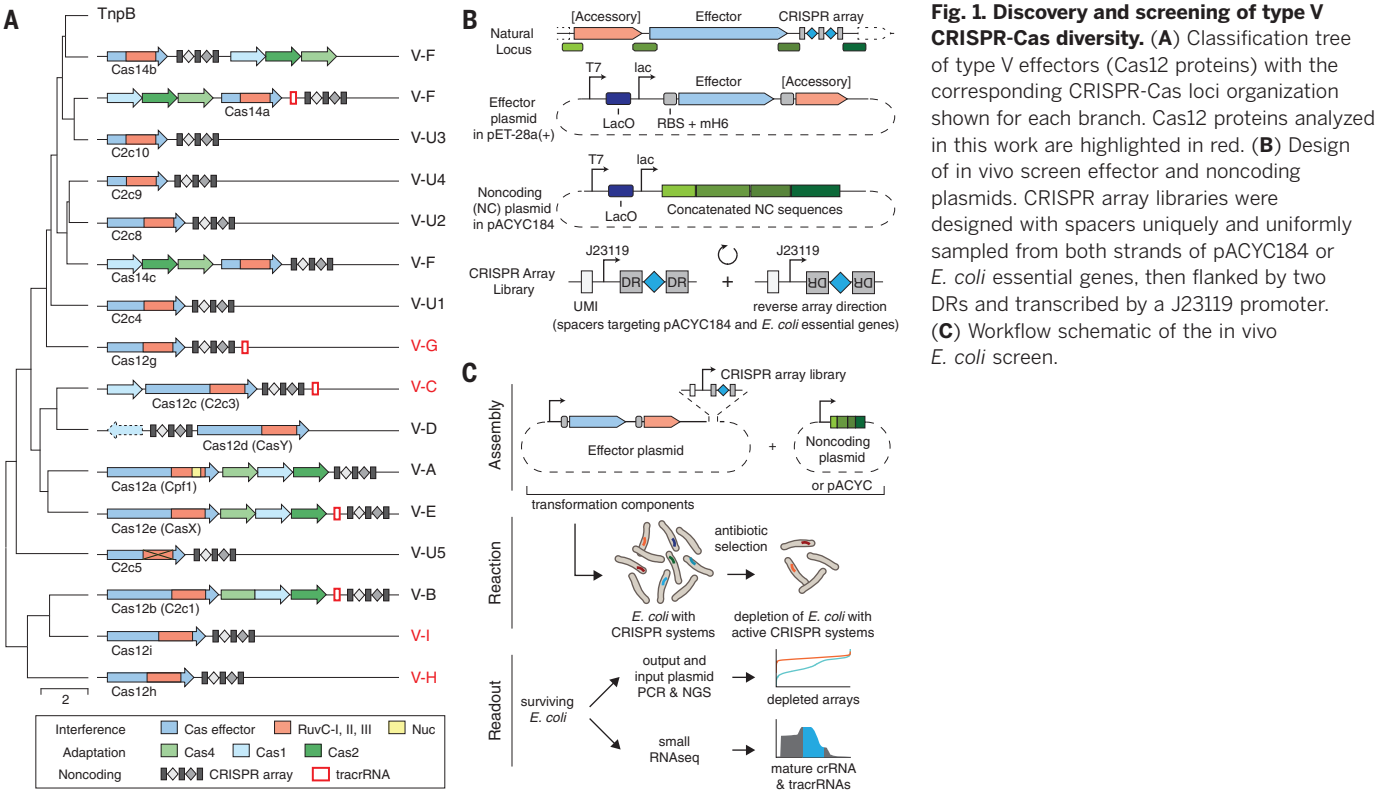


Fig. 1. Discovery and screening of type V CRISPR-Cas diversity. (**A**) Classification tree of type V effectors (Cas12 proteins) with the corresponding CRISPR-Cas loci organization shown for each branch. Cas12 proteins analyzed in this work are highlighted in red. (**B**) Design of in vivo screen effector and noncoding plasmids. CRISPR array libraries were designed with spacers uniquely and uniformly sampled from both strands of pACYC184 or *E. coli* essential genes, then flanked by two DRs and transcribed by a J23119 promoter. (**C**) Workflow schematic of the in vivo *E. coli* screen.

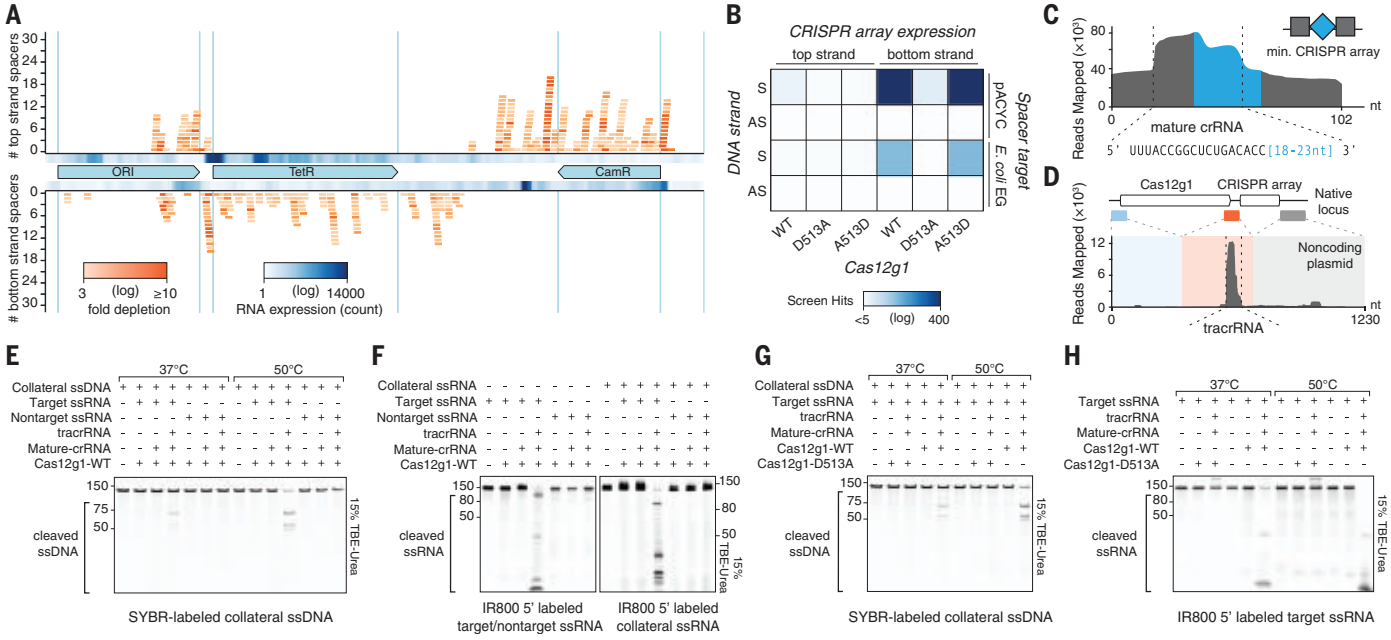


Fig. 2. Cas12g displays RNA-activated target cleavage of RNA and collateral trans-cleavage of RNA and ssDNA. (**A**) Strongly depleted CRISPR arrays from in vivo screening of Cas12g1 and its noncoding plasmid mapped to pACYC184. (**B**) Heatmap showing strongly depleted CRISPR arrays (screen hits) to evaluate RuvC and substrate strand dependencies of Cas12g1 (S, sense; AS, antisense; EG, essential genes). A513D was cloned from the D513A construct to rescue its activity. Strongly depleted CRISPR arrays in negative control screens without the effector were subtracted from this and similar analyses. (**C** and **D**) Mature crRNA (C) and tracrRNA (D) identified from small RNA sequencing of in vivo screen samples containing Cas12g1 and noncoding plasmid. The schematic above tracrRNA shows construction of noncoding plasmid from native locus sequences. (**E** and **F**) Target ssRNA activated collateral ssDNA cleavage at 37°C and 50°C (E) and target and collateral ssRNA cleavage at 37°C (F). (**G** and **H**) Cleavage assays targeting collateral ssDNA (G) and ssRNA (H) with purified RuvC mutant dCas12g1 D513A.

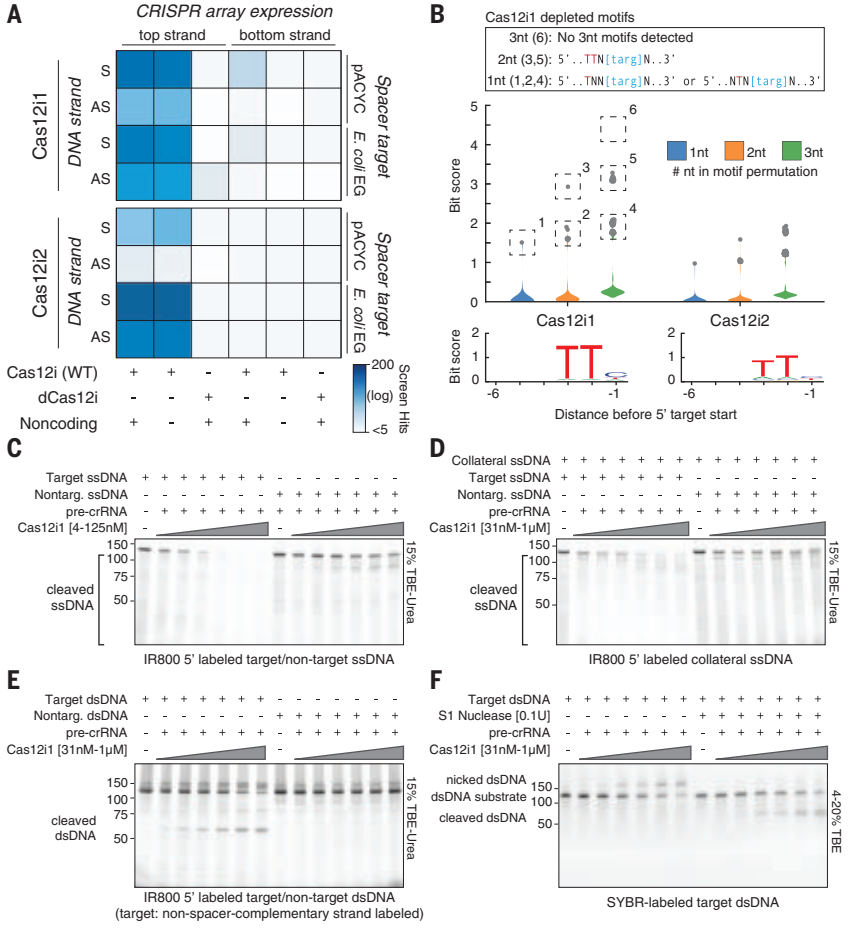


Fig. 3. In vivo and in vitro activity of Cas12i. (**A**) Evaluation of a minimal active system for Cas12i, with heatmaps showing strongly depleted CRISPR arrays from in vivo screening in different Cas12i system compositions (S, sense; AS, antisense; EG, essential genes). (**B**) (Top) Distribution of bit scores for all permutations of 1- to 3-nucleotide (nt) motifs within the target and 15-nt flanking sequences corresponding to strongly depleted in vivo arrays, calculated as described in (*19*). The box above describes motif analysis for Cas12i1 as an example. (Bottom) Web logos from target-flanking sequences. (**C** to **E**) Titration of a Cas12i1 binary complex on target and nontarget ssDNA (C), collateral ssDNA with target and nontarget ssDNA (D), and target and nontarget dsDNA (E). (**F**) S1 nuclease treatment to resolve dsDNA nicks (induced by Cas12i1) into dsDNA breaks.

amino acids) and Cas12i (1033 to 1093 amino acids), respectively. These effectors show distant similarity to Cas12b, with substantial truncation of N-terminal regions responsible for PAM recognition and DNA unwinding (*25*, *26*). In vivo screening of Cas12h1 (870 amino acids), Cas12i1 (1093 amino acids), and Cas12i2 (1054 amino acids) demonstrated robust and broadly distributed targeting of both strands of dsDNA substrates that was dependent on an intact RuvCI domain (Fig. 3A and figs. S13A and S14, A to F). The noncoding plasmid was not required, indicating that, unlike subtype V-B systems, the minimal V-H and V-I interference modules include only the effector and crRNA (Fig. 3A and fig. S14, G and H). Analysis of target-flanking sequences corresponding to strongly depleted arrays from in vivo screens showed that dsDNA interference by Cas12h1 depends on a 5′ RTR PAM (fig. S13B), whereas Cas12i1 and Cas12i2 prefer a 5′ TTN PAM (Fig. 3B).

Small RNA sequencing of Cas12i1 in vivo screen samples demonstrated biogenesis of a mature crRNA (fig. S15), which was confirmed in vitro using purified Cas12i1 and a minimal pre-crRNA (DR-spacer-DR-spacer-DR) (fig. S16 and table S3). Binary complexes containing Cas12i1 and pre-crRNAs efficiently cleaved target containing ssDNA substrates (Fig. 3C) as well as labeled collateral ssDNA in the presence of unlabeled target ssDNA, consistent with collateral ssDNA cleavage activity (Fig. 3D).

We observed Cas12i1-mediated cleavage of dsDNA under denaturing conditions, which was suggestive of dsDNA nicking. While reactions containing dsDNA with a labeled non-spacer-complementary strand showed robust DNA cleavage over a wide range of binary complex concentrations, those containing dsDNA with the spacer-complementary strand labeled showed only small amounts of cleavage at the highest concentrations tested (Fig. 3E and fig. S17, A and B). Under nondenaturing conditions, Cas12i1 cleavage reactions yielded products with lower electrophoretic mobility than the input dsDNA, and these products were then converted to double-strand breaks by S1 nuclease treatment, consistent with nicking of dsDNA substrates (Fig. 3F). These results suggested that Cas12i1 preferentially nicks the non-spacer-complementary strand, and it cleaves the spacer-complementary strand with a lower efficiency to yield a dsDNA break. Together, the small size, autonomous processing of multiplexed crRNAs, and nicking activity of Cas12i could enhance double-nicking applications for high-fidelity genome editing (*27*).

Subtype V-C loci have been previously observed but never characterized due to incomplete genomic data (*10*). With our expanded database, we detected and synthesized in vivo screen plasmids for complete subtype V-C systems containing the effectors OspCas12c (from *Oleiphilus* sp. HI0009), Cas12c1, and Cas12c2. All these systems showed broad and symmetrical targeting of both DNA strands, consistent with autonomous dsDNA interference (Fig. 4A and fig. S18). RNA sequencing of screening samples for the minimal subtype V-C systems demonstrated pre-crRNA processing and highly expressed tracrRNAs (fig. S19). A 5′ TG PAM was required for Cas12c1 and OspCas12c, and a minimal 5′ TN PAM was required for Cas12c2 (Fig. 4B). The single-nucleotide TN PAM for Cas12c2 dsDNA targeting complements recently engineered Cas9 effectors with minimal PAMs (*28*), potentially expanding the target space for genome editing.

We have presented here a framework for systematic discovery, screening, and characterization of class 2 CRISPR-Cas systems, and we demonstrated a range of activities for four type V CRISPR-Cas subtypes, including target and collateral cleavage of ssRNA and ssDNA as well as dsDNA nicking and cleavage (Fig. 4C). These findings reveal the transition in the properties of Cas12 proteins along the proposed evolutionary path from TnpB to large type V effectors. Additionally, future applications could include

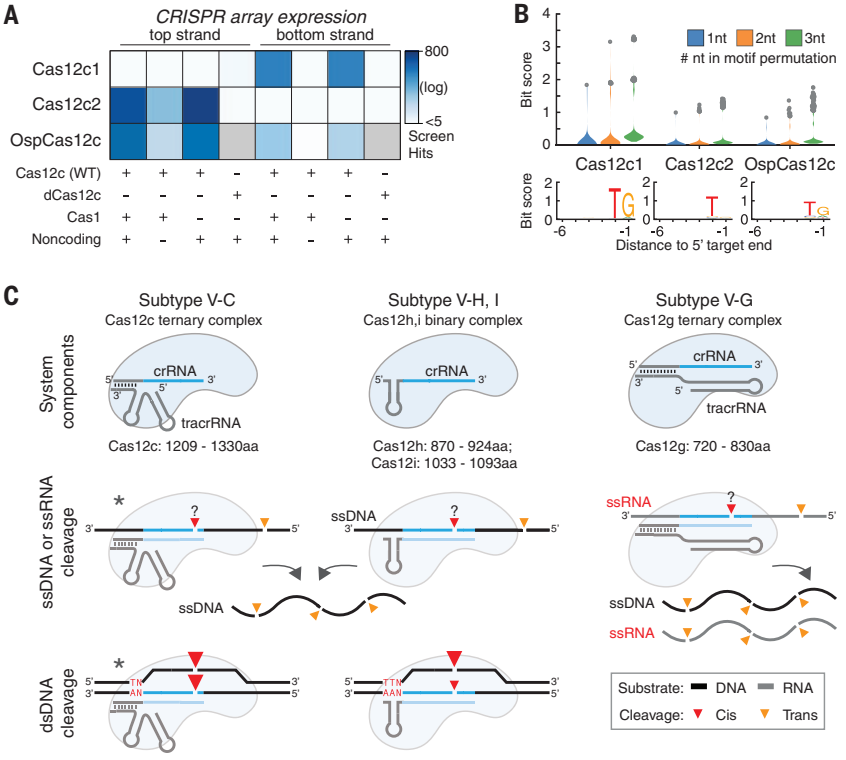


Fig. 4. In vivo dsDNA interference by Cas12c. (**A**) Evaluation of a minimal active system for Cas12c, with heatmaps showing strongly depleted CRISPR arrays from in vivo screening in different Cas12c system compositions. Gray boxes indicate data not available. (**B**) (Top) Distribution of bit scores for all permutations of 1- to 3-nt motifs within the target and 15-nt flanking sequences corresponding to strongly depleted arrays. (Bottom) Web logos from target-flanking sequences. (**C**) Overview of minimal components and interference mechanisms of Cas12g, -h, -i, and -c. Asterisks denote putative mechanisms subject to additional validation.

expanded genomic targeting via the minimal Cas12c2 PAM, high-fidelity genome editing using Cas12i nicking (*27*), or sensitive and durable nucleic acid detection via collateral cleavage by the thermostable Cas12g1 (*18*, *22*). We anticipate that our discovery framework will yield new CRISPR-Cas variants as genomic and metagenomic sequence databases grow, expanding the understanding of CRISPR biology and the nucleic acid manipulation toolbox.

REFERENCES AND NOTES

1. D. Burstein *et al.*, *Nature* **542**, 237–241 (2017).
2. S. Doron *et al.*, *Science* **359**, eaar4120 (2018).
3. E. V. Koonin, K. S. Makarova, Y. I. Wolf, *Annu. Rev. Microbiol.* **71**, 233–261 (2017).
4. R. Barrangou, P. Horvath, *Nat. Microbiol.* **2**, 17092 (2017).
5. R. Barrangou *et al.*, *Science* **315**, 1709–1712 (2007).
6. D. B. T. Cox, R. J. Platt, F. Zhang, *Nat. Med.* **21**, 121–131 (2015).
7. S. E. Klompe, S. H. Sternberg, Harnessing "A Billion Years of Experimentation": The Ongoing Exploration and Exploitation of CRISPR–Cas Immune Systems. *CRISPR J.* **1**, 141–158 (2018).
8. G. J. Knott, J. A. Doudna, *Science* **361**, 866–869 (2018).
9. K. S. Makarova *et al.*, *Nat. Rev. Microbiol.* **13**, 722–736 (2015).
10. S. Shmakov *et al.*, *Mol. Cell* **60**, 385–397 (2015).
11. S. Shmakov *et al.*, *Nat. Rev. Microbiol.* **15**, 169–182 (2017).
12. G. Gasiunas, R. Barrangou, P. Horvath, V. Siksnys, *Proc. Natl. Acad. Sci. U.S.A.* **109**, E2579–E2586 (2012).
13. M. Jinek *et al.*, *Science* **337**, 816–821 (2012).
14. E. V. Koonin, K. S. Makarova, F. Zhang, *Curr. Opin. Microbiol.* **37**, 67–78 (2017).
15. B. Zetsche *et al.*, *Cell* **163**, 759–771 (2015).
16. J. S. Chen *et al.*, *Science* **360**, 436–439 (2018).
17. O. O. Abudayyeh *et al.*, *Science* **353**, aaf5573 (2016).
18. A. East-Seletsky *et al.*, *Nature* **538**, 270–273 (2016).
19. W. X. Yan *et al.*, *Mol. Cell* **70**, 327–339.e5 (2018).
20. L. B. Harrington *et al.*, *Science* **362**, 839–842 (2018).
21. J. S. Gootenberg *et al.*, *Science* **360**, 439–444 (2018).
22. J. S. Gootenberg *et al.*, *Science* **356**, 438–442 (2017).
23. D. B. T. Cox *et al.*, *Science* **358**, 1019–1027 (2017).
24. S. Konermann *et al.*, *Cell* **173**, 665–676.e14 (2018).
25. D. Dong *et al.*, *Nature* **532**, 522–526 (2016).
26. T. Yamano *et al.*, *Cell* **165**, 949–962 (2016).
27. F. A. Ran *et al.*, *Cell* **154**, 1380–1389 (2013).
28. J. H. Hu *et al.*, *Nature* **556**, 57–63 (2018).

ACKNOWLEDGMENTS

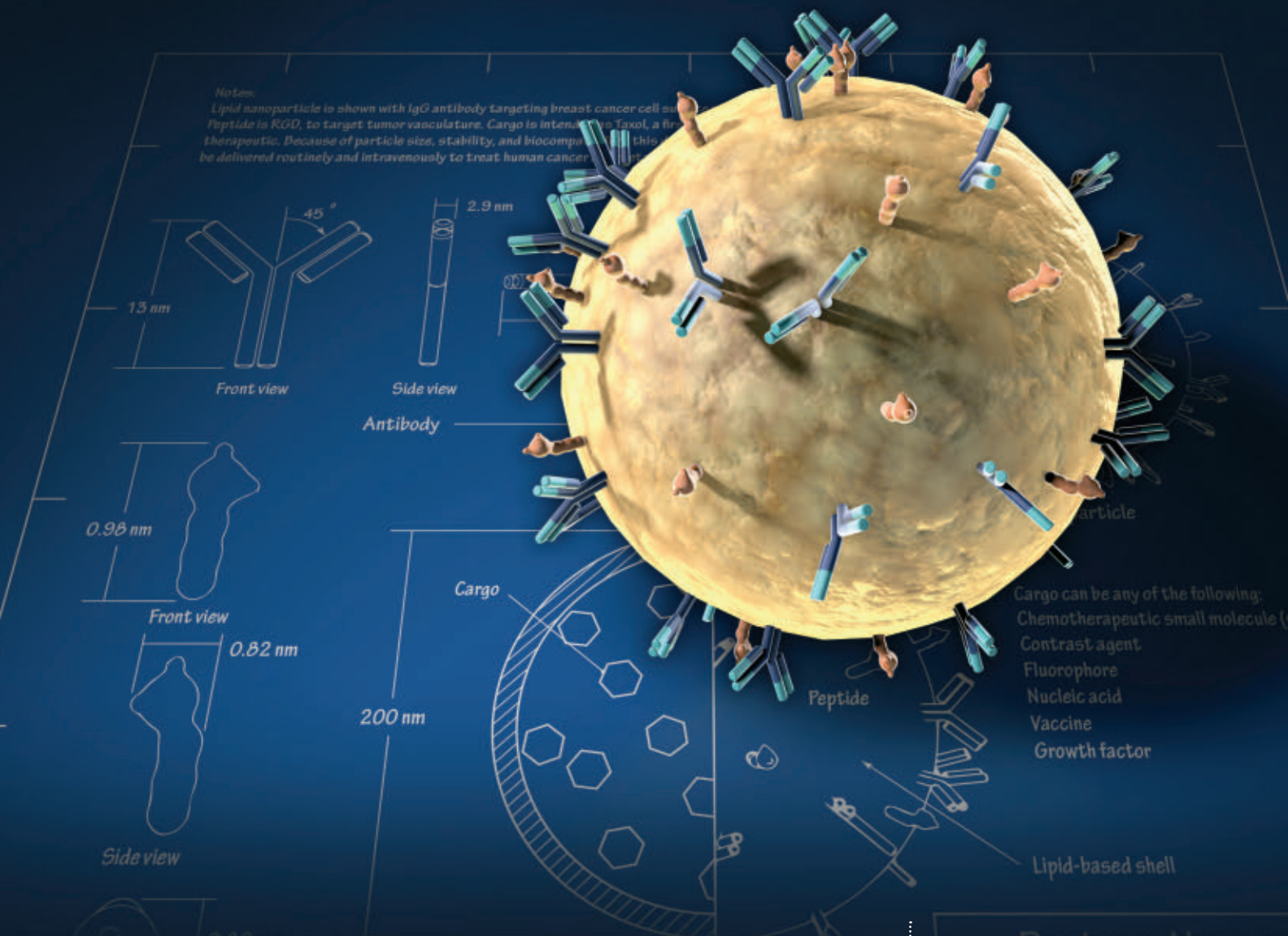
We thank the entire Arbor Biotechnologies team for support and comments on this work. **Funding:** Arbor Biotechnologies is a privately funded company. K.S.M. and E.V.K. are supported by the intramural program of the U.S. Department of Health and Human Services (to the National Library of Medicine). **Author contributions:** W.X.Y. and D.A.S., with input from P.H., D.R.C., L.E.A., J.M.C., E.K.S., S.S., S.C., and A.J.G., conceived and designed the study. D.R.C. and D.A.S. designed and implemented the computational searches, with additional input from K.S.M. and E.V.K., including phylogenetic analysis and classification. W.X.Y., D.A.S., P.H., L.E.A., J.C., E.K.S., S.S., S.C., and A.J.G. performed all of the experimental work and analyzed the data. W.X.Y. and D.A.S. wrote the manuscript with input from E.V.K. and help from all authors. **Competing interests:** W.X.Y., P.H., L.E.A., J.M.C., E.K.S., S.S., S.C., A.J.G., D.R.C., and D.A.S. are employees and shareholders of Arbor Biotechnologies, Inc. W.X.Y., D.R.C., and D.A.S. are current or former officers and D.R.C. is a director of Arbor Biotechnologies. Arbor Biotechnologies has filed patents related to this work. **Data and materials availability:** All data are available in the manuscript or the supplementary material. All reagents are available to the academic community through Addgene. Sequencing data are available on the NCBI Sequence Read Archive under Bioproject ID PRJNA496291.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/363/6422/88/suppl/DC1
Materials and Methods
Figs. S1 to S19
Tables S1 to S10
References (*29*–*33*)

14 October 2018; accepted 20 November 2018
Published online 6 December 2018
10.1126/science.aav7271

Make Your Research Hit the Target



Science Translational Medicine publishes peer-reviewed, cutting-edge biomedical research in the fields of cardiology, cancer, immunotherapy, infectious diseases and vaccines, bioengineering and devices, neurology and neurodegenerative diseases, obesity, diabetes and metabolic disorders, drug discovery, genomic medicine, imaging, stem cell therapy and regenerative medicine.

Submit your research today

Learn more at: **ScienceTranslationalMedicine.org**

Science Translational Medicine

AAAS

SCIENCE CAREERS ADVERTISING

For full advertising details, go to ScienceCareers.org and click For Employers, or call one of our representatives.

AMERICAS
+1 (202) 326 6577
+1 (202) 326 6578
advertise@sciencecareers.org

EUROPE, INDIA, AUSTRALIA, NEW ZEALAND, REST OF WORLD
+44 (0) 1223 326527
advertise@sciencecareers.org

CHINA, KOREA, SINGAPORE, TAIWAN, THAILAND
+86 (131) 4114 0012
advertise@sciencecareers.org

JAPAN
+81 (3) 6459 4174
advertise@sciencecareers.org

JOB BOARD SUPPORT
support@sciencecareers.org

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

ScienceCareers.org

All ads submitted for publication must comply with applicable U.S. and non-U.S. laws. *Science* reserves the right to refuse any advertisement at its sole discretion for any reason, including without limitation for offensive language or inappropriate content, and all advertising is subject to publisher approval. *Science* encourages our readers to alert us to any ads that they feel may be discriminatory or offensive.

10 ways that *Science* Careers can help advance your career

1. Register for a free online account on **ScienceCareers.org**.
2. Search thousands of job postings and find your perfect job.
3. Sign up to receive e-mail alerts about job postings that match your criteria.
4. Upload your resume into our database and connect with employers.
5. Watch one of our many webinars on different career topics such as job searching, networking, and more.
6. Download our career booklets, including Career Basics, Careers Beyond the Bench, and Developing Your Skills.
7. Complete an interactive, personalized career plan at "my IDP."
8. Visit our Career Forum and get advice from career experts and your peers.
9. Research graduate program information and find a program right for you.
10. Read relevant career advice articles from our library of thousands.

Visit **ScienceCareers.org** today — all resources are free

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

SCIENCECAREERS.ORG

By **Barbara A. Wanchisen**

Lessons from the 'real world'

When I left my tenured professorship for a nonacademic job, I thought I had already done the hard part: making the decision to leave the professional world that had been my home for many years. I had little inkling of the culture shock that awaited me in my new sphere of science policy and how disorienting it would be. In the 17 years since, I've seen that I'm far from the only one who struggles with this transition. Regardless of career stage, leaving academia requires some adjusting. For those who, like me, make the move later in their careers, after decades as established academics, it can be even harder. Knowing what to expect beforehand can take some of the shock out of the transition to the world outside academia.

YOU AREN'T THE CENTER OF THE WORLD, AND YOU WILL HAVE AN ACTUAL BOSS. If you're a senior academic, teaching, advising, and managing a research group can turn your world into a solipsistic universe where what you say goes. Most students are attentive and agreeable, no matter what they really think. The dean and department chair are loosely construed as your supervisors, but they generally want you to focus on your own ideas so that you can bring in grant money and prestige.

Outside of academia, on the other hand, collaboration in service of a common goal is far more important than any one person's ideas. That is true even if you are in a very senior leadership role. There are goals to achieve and relationships to manage; you can't pursue just your own interests. Having a more formal boss might feel odd, but be open to it. Bosses can offer support in ways that university structures can't.

"I had little inkling of the culture shock that awaited me."

YOU WON'T HAVE NEARLY ENOUGH TIME TO DO THE BACKGROUND WORK YOU'D LIKE. Completeness and thorough examination are hallmarks of a serious scholar. In academia, you can usually take all the time you need to do plenty of background research, talk to colleagues, and cogitate before producing a decision or a publication. But in the outside world, you may have 30 minutes to come up with an answer. This was a particularly hard lesson for me to learn. Sage advice from someone more junior than I but with more "real world" work experience helped immensely. They said that sometimes I just had to go with my gut and call it a day.

SUMMER IS NOT SPECIAL. After a long academic year, summer feels different and, in some ways, special. It can be a time to regroup and refresh—write, develop ideas, even spend a few months away from the distractions of campus. Outside of academia, summer is generally three more months in the workplace. It might be punctuated by vacation, but by and large, work goes on as usual for 12 months a year.

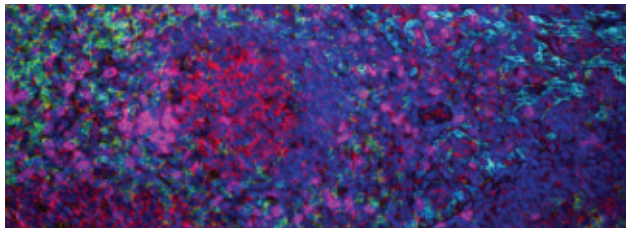
THERE IS NO TENURE. This is a tough one! Many in the professoriate value this near-guarantee of steady employment for a lifetime. But leaving tenure can actually be liberating for a go-getter with a good work ethic. Employers want to keep strong, creative employees around, and they realize that individuals are not necessarily there for life. That means that organizations can be very loyal to their long-term, high-performing employees and treat them well. And if you don't enjoy the work or feel badly treated, there's no reason to stay.

If you decide to take the plunge into the outside world, take some time to acclimate to the new environment. You'll probably experience some initial surprises, but give yourself a chance to see whether things get better. And if this article convinces you that academia is the right place for you and that you should stay and flourish there, then that is a win, too. ■

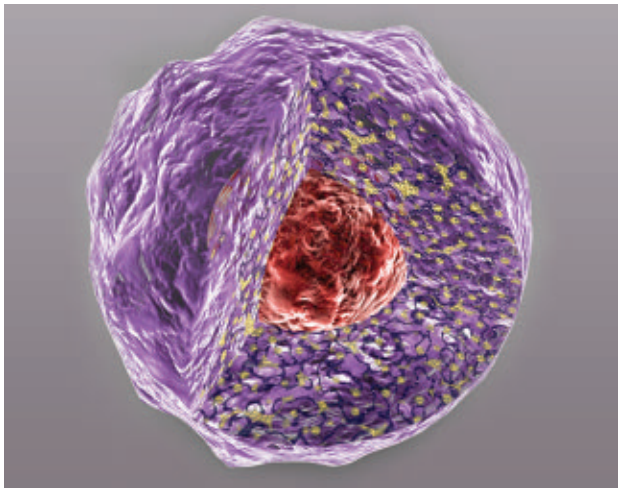
Barbara A. Wanchisen is the director of the Board on Behavioral, Cognitive, and Sensory Sciences at the National Academies of Sciences, Engineering, and Medicine in Washington, D.C. Send your story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

want new technologies?



antibodies
apoptosis
biomarkers
cancer
cytometry
data
diseases
DNA
epigenetics
genomics
immunotherapies
medicine
microbiomics
microfluidics
microscopy
neuroscience
proteomics
sequencing
toxicology
transcriptomics



watch our webinars

Learn about the latest break-throughs, new technologies, and ground-breaking research in a variety of fields. Our expert speakers explain their quality research to you and answer questions submitted by live viewers.

VIEW NOW!
webinar.sciencemag.org

Science
AAAS

Brought to you by the *Science*/AAAS Custom Publishing Office

@SciMagWebinars

AACR
American Association for Cancer Research®

ANNUAL MEETING 2019 \\ ATLANTA

March 29-April 3 | Georgia World Congress Center | Atlanta, GA

Join us in Atlanta for the latest innovative and inspiring cancer research from around the world

Register Today!
Become a Member! Join the AACR and receive a discount on registration

The AACR Annual Meeting highlights the work of the greatest minds in cancer science and medicine from institutions all over the world. This meeting presents the many scientific discoveries across the breadth of cancer research—from prevention, early detection, and interception; to cancer biology, translational, and clinical studies; to survivorship, population science, and advocacy. This year's program, with the theme of "Integrative Cancer Science • Global Impact • Individualized Patient Care," will be a comprehensive, cutting-edge scientific event that you will not want to miss!

INTEGRATIVE CANCER SCIENCE
GLOBAL IMPACT
INDIVIDUALIZED PATIENT CARE

#AACR19

AACR.org/AnnualMeeting