Science

$15
11 OCTOBER 2024
SPECIAL ISSUE
science.org

AAAS
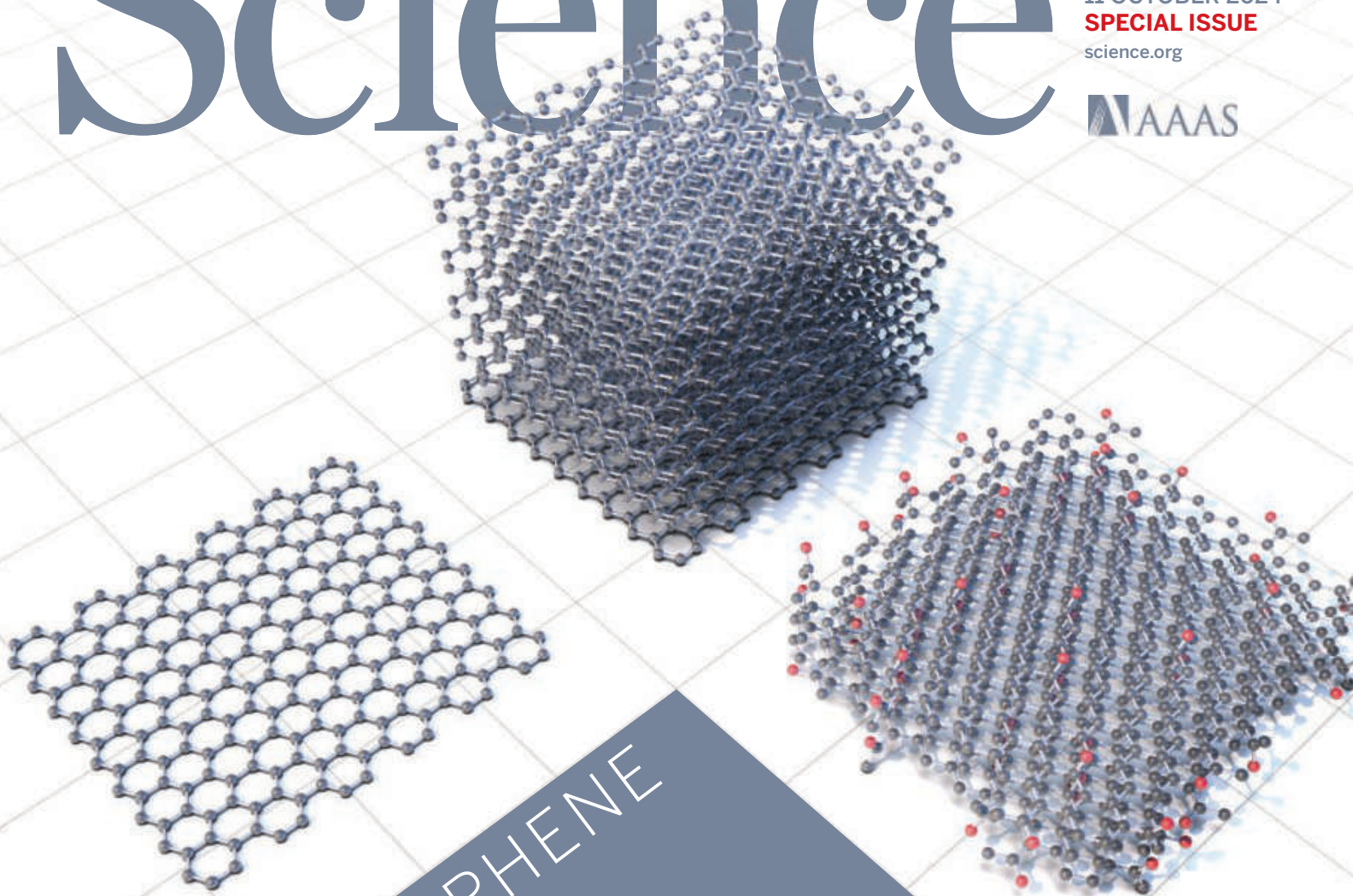
GRAPHENE AT 20

New forms— and a brighter future—for the storied material

CONTENTS

INSIGHTS

SPECIAL SECTION
Graphene at 20

ON THE COVER

It is still tricky to make the perfect, atom-thin sheets of carbon known as graphene (left). Instead, companies trying to commercialize the material often use stacks of graphene nanoplatelets (middle) and layers of graphene oxides (right). These cheaper forms of graphene can contain defects but approximate its superpowers as a strong and electrically sensitive material. See pages 138 and 144.
Illustration: C. Bickel/ Science

NEWS

RESEARCH

Forest plant species in Europe are shifting westward more than toward the poles, in response to nitrogen and sulfur deposition.

Galapagos finch (*Geospiza fortis*) bills, and thus songs, change with drought, potentially leading to speciation.

211

Is it cancer?
Delivering PRECISION in the determination of breast cancer risk

An intact surgical resection of Ductal Carcinoma In Situ (DCIS) stained with fluorescent labels for different markers. Each color represents a different cell type.

I n 2009, a single mammogram changed the trajectory of Donna Pinto's life. Though the test did not produce a breast cancer diagnosis, it revealed a growth known as ductal carcinoma in situ (DCIS). She faced a grueling treatment regimen. "[The] treatment options are a mastectomy or a partial mastectomy with 7 weeks of radiation," says Pinto, who is now a patient advocate based in San Diego, California.

Although sometimes referred to as "stage 0" breast cancer, most of the time—in up to 80% of cases—these small clumps of abnormal cells within the mammary ducts will never progress to a malignant state. But because clinicians have no way to distinguish those growths from those that pose a real threat, the default treatment protocol for women like Pinto is to handle DCIS like an actively growing cancer. This means many women are at risk of undergoing costly and stressful overtreatment that will not deliver any meaningful health benefit.

Developing better ways to tell lethal and nonlethal cancers apart was one of the first challenges set by Cancer Grand Challenges, a research initiative co-founded by Cancer Research UK and the National Cancer Institute in the U.S. and supported by a network of partners. DCIS fit the bill perfectly, and team PRECISION was awarded the opportunity to identify its underlying causes and why some forms of DCIS are aggressive while others pose no threat. Team lead Jelle Wesseling, a pathologist at the Netherlands Cancer Institute, has struggled with how to handle DCIS diagnoses for his entire career. "I didn't feel comfortable about the treatment of these women," says Wesseling, noting that the financial, psychological, and physical costs of potentially unnecessary surgeries and radiotherapy can be tremendous. "But to de-escalate medical practice is quite a difficult thing to do."

Armed with ~$19 million in Cancer Grand Challenges funding from Cancer Research UK and the Dutch Cancer Society (KWF), team PRECISION assembled an international team of researchers, clinicians, and advocates—including Pinto. Since 2017, the team has taken a deep dive into DCIS biology, working from the scale of genes and biomolecules all the way up to the real-world challenges of people grappling with a difficult diagnosis. And thanks to these unceasing efforts, Wesseling now sees opportunities to meaningfully move the needle in DCIS care.

Deep learning

One of PRECISION's priorities was to learn more about how DCIS arises and about which cellular processes determine whether it progresses, as these aspects were poorly understood. They achieved a major victory on this front—described in a 2023 *Cancer Cell* article—developing an extensive collection of more than 100 patient-derived xenograft models, in which DCIS cells from donors are surgically implanted into the mammary tissue of mice. This allows researchers to observe these cells in real time as they establish themselves in their new hosts, learn why certain cells flourish or stagnate, and gain a dynamic view of DCIS progression that would otherwise be inaccessible in humans.

"Basically, at the two extremes, you see two growth patterns: what we call replacement growth versus expansive growth," explains Wesseling. Importantly, these behaviors were shown to be innate properties of the abnormal cells themselves, such that the original cells can be repeatedly cultured and used to continuously produce new models. The researchers behind this effort, led by molecular pathologist Jos Jonkers at the Netherlands Cancer Institute, have made their collection available for other researchers interested in characterizing DCIS. Recent work from team members Jacco van Rheenen and Colinda Scheele shows multiple layers of protection exist to stop mutated clones from spreading within the breast, identifying the molecular mechanisms behind why having fewer menstrual cycles is associated with decreased breast cancer risk.

Another front of the team's research efforts was characterizing the cellular and molecular "signatures" of DCIS and assessing how they correlate with the

risk of future progression. This work was spearheaded by researchers including Elinor Sawyer, a clinical oncologist at King's College London specialized in breast cancer.

"We were essentially asking the question: which DCIS do you not need to treat?" says Sawyer. "Ideally, what we want is a really good biomarker to say, 'Actually, your risk of recurrence is so low that you don't need to even have an operation.'" She contrasts this with prior efforts that took the opposite approach, asking which growths pose the most imminent danger—an approach that will not reduce the risk of overtreatment.

Among other tools, the team used RNA-seq to broadly analyze genome-wide gene expression in DCIS samples. This allowed them to identify patterns that can discriminate low- and high-risk forms of DCIS, and according to Sawyer, they've begun using this screening method with additional cohorts. But there are also informative features of DCIS waiting to be uncovered in the data delivered by mammography. Clinicians may not know how to spot them, but the powerful pattern recognition capabilities of deep learning algorithms can make associations based on subtle features of the shape and organization of DCIS growths. An initial iteration of this artificial intelligence–based approach could discriminate low-risk DCIS from high-risk with an accuracy of 76%, as described in a 2024 article from *Cancer Imaging*. Wesseling is encouraged by the progress on this front, and says discussions are underway with prospective industry partners who might help validate and further develop this algorithm to be implemented in the clinic.

Testing active surveillance

PRECISION's work benefits from separate clinical trials running parallel to their research on preventing DCIS overtreatment. COMET—a randomized controlled trial that has recruited 1200 people diagnosed with DCIS—is one example. The COMET study compares the traditional, all-out DCIS treatment protocol to a less aggressive approach called "active surveillance." In the latter scenario, clinicians monitor progression closely but only intervene if there is evidence of malignant growth and invasion. The goal of COMET is to determine whether active surveillance can improve quality of life without worsening cancer risk.

Patient advocates played a critical role in these trials, and Pinto is closely involved with COMET. "A lot of women preferred active surveillance, but ... you have to be willing to be randomized," she says. This meant she had to work with the researchers to develop strategies for communicating to participants the importance of randomized controlled trials for objectively assessing clinical interventions. This preference for active surveillance was also apparent in the Netherlands-

Donna Pinto **(right)** with Deborah Collyar **(left),** team PRECISION patient advocates during the team's final annual research meeting in Amsterdam.

based LORD trial, where slow recruitment compelled Wesseling and his study team to change their study design to allow the choice between active surveillance and conventional treatment. To date, Wesseling says that 80% of participants opted for the former option after weighing the pros and cons.

The combination of good prognostic biomarkers and good strategies for communicating risk could fundamentally change the standard of care for DCIS treatment, favoring active surveillance in many cases currently slated for surgery—but there are still important unknowns. For example, a 2022 PRECISION study published in *Nature Genetics* analyzed recurring growths that appear after surgical DCIS treatment and whether these new growths were related to the original DCIS lesion. "About 20% of the so-called recurrences were not actually true recurrences," says Sawyer. "They're actually new primary tumors," says Sawyer. This introduces complications in terms of follow-up treatment planning, but it could also confound efforts to develop prognostic biomarkers if a meaningful number of tumors that arise after DCIS detection are not directly related to the DCIS. "They're kind of muddying the waters," says Sawyer. Importantly, these findings also establish DCIS as a broader risk factor for subsequent breast cancer even if the DCIS growth is not necessarily a tumor precursor itself.

Despite these complexities, team PRECISION's members are enthusiastic about the headway they've made toward improving DCIS care for future generations with this condition. Wesseling praises Cancer Grand Challenges for making this possible through the initiative's flexibility and scale of funding, which allowed the team to respond quickly and bring on additional expertise and resources as new questions arose. And Pinto's experience with the team has given her hope that future diagnoses of DCIS will be far less disruptive and draining than her own experience. "Worse than surgery is the mental anguish," she says. "It's really going to save a lot of women from just an emotional terror."

Sponsored by

Science
AAAS

Science in a shifting policy system

n June, the US Supreme Court overturned a 40-year-old legal principle known as the *Chevron* doctrine, which deferred to federal agencies on how regulations are interpreted in legal cases when the legislation enacted by Congress is unclear. Striking a blow against this precedent, Supreme Court Chief Justice John Roberts indicated that *Chevron* was "fundamentally misguided" and that only the nation's courts have the prerogative to interpret the laws they administer. Judges across the United States now anticipate an onslaught of litigation seeking to overturn federal regulatory actions, and many of these lawsuits will turn on scientific evidence that guided those actions. Instead of deferring to the scientific and technical expertise of federal agencies, courts must now independently review the scientific record behind a challenged regulation for its persuasiveness, along with scientific findings produced by the litigating parties. It's too early to fully unpack the impact of overturning *Chevron*, but the scientific community must step forward to support the crafting of clear and evidence-informed legislation, and the use of scientific evidence in the courts.

Federal agency staff of the executive branch of the US government, such as the Department of Health and Human Services, the Department of Energy, and the Department of Agriculture, are trained to interpret and implement the often-complex technicalities of legislation passed by Congress. Recognizing that federal courts lack the relevant scientific and technical expertise, the *Chevron* doctrine directed them to defer to the executive branch's interpretation of ambiguous legislation.

Congress often drafts bills with broad language, allowing laws to be flexibly applied to unanticipated and emergent issues. However, vague language can open the way to challenges. For example, the Environmental Protection Agency's interpretation of the US Clean Water Act has raised numerous controversies about the distinction between wetlands and waterways. Last year, the Supreme Court curtailed federal protection of wetlands, leaving an estimated 17 to 90 million acres unprotected without regulations at the state level. This is a preview of what lies ahead. Instead of entrusting agency professionals with interpreting the fine points of the law, Congress will have to ramp up its own scientific expertise and draft legislation with more specific language. The alternative would be years of litigation by those intent on opposing government regulations.

Judges themselves also will increasingly need accessible scientific information and analysis as courts get called upon more frequently to review the science behind laws and regulations involving public health, the environment, education, energy, commerce, and more. Currently, the Federal Judicial Center provides judges and court staff with a *Reference Manual on Scientific Evidence*, which describes the basic tenets of scientific fields from which legal evidence is derived. Such reference material may not be sufficient now that judges no longer give the benefit of the doubt to agency expertise.

Scientific associations that provide summaries of scientific consensus to legislators and their staff will also have to work harder to ensure the drafting of more precise laws. Organizations such as the American Educational Research Association and the American Psychological Association already monitor lawsuits to identify those where a judge could benefit from a "friend of the court" brief on the relevant science, without taking a position on how the case should be decided. For example, information submitted to the court on adolescent brain development from neuroscience and psychology societies has been cited in youth justice cases. Judges rarely invite these interventions, so scientific associations will have to take even more initiative.

> "...the scientific community must become...more involved in helping lawmakers and the courts..."

In this new era, the next generation of scientists should possess a stronger understanding of the regulations relevant to their fields and the impacts of laws not only on their own research but on science as a whole and science's benefit to society. Establishing new courses and training for students and early career scientists in this vein may inspire them to bring their scientific expertise to legislative and judicial decision-making.

Government policy has always been fraught with politics and conflicting interests. The *Chevron* doctrine attempted to rise above the fray by giving priority to the expertise of the professionals working at federal agencies. With that off the table, the scientific community must become ever more involved in helping lawmakers and the courts, ensuring that legislation, policy, legal interpretation, and oversight are informed by the best available science.

–Theresa Harris and Rashada Alexander

Theresa Harris
is director of the AAAS Center for Scientific Responsibility and Justice*, Washington, DC, USA. tharris@aaas.org

Rashada Alexander
is director of the AAAS Science and Technology Policy Fellowships* program, Washington, DC, USA. ralexander@aaas.org

Infusing new energy into the field of aging research

The day we are born, the aging clock starts ticking. It may tick faster or slower depending on our genes, the good and bad habits we pick up in life, or just plain luck—but it always moves inexorably forward until the day we die.

The aging process is not fully outside of human control, however. And over the past few decades, researchers have uncovered some intriguing mechanisms that could one day be exploited to slow or potentially even reset the clock for key determinants of the aging process.

In 2014, for example, researchers led by Stanford University neuroscientist Tony Wyss-Coray performed experiments in which they infused old mice with blood from young animals—a technique known as parabiosis. The effects were striking, with the treated mice showing increased growth and connectivity among brain cells, and clear improvements in memory and cognition. Between 2017 and 2022, the Switzerland-based NOMIS Foundation would fund efforts by Wyss-Coray's group to identify specific blood-borne molecules that can facilitate similar "rejuvenation" of the brain in humans.

This notion is still not universally embraced. "I think the general concept that age is malleable is something that has only recently sort of become acknowledged by most scientists," says Wyss-Coray. "I think there's still many who have not heard about it or are surprised about it." But the implications could be tremendous and far-reaching.

Nir Barzilai, director of the Institute for Aging Research at the Albert Einstein College of Medicine in New York City, notes that the benefits of increasing the number of healthy years that humans can experience go well beyond public health and individual quality of life. "If they're not in the hospital, where are they? They are traveling, they're shopping, they have a contribution to the economy," says Barzilai. As an example, he cites research from Andrew Scott at the London Business School suggesting that extending the healthy lifespan of humans could generate tens of trillions of dollars of economic activity.

But before humanity can achieve that potential, researchers must crack the many remaining mysteries underlying the aging process—and learn how to translate those into treatments. Fortunately, the field is rapidly making headway, thanks to innovative research powered by growing investment from government agencies, industry, and organizations like the NOMIS Foundation.

Hints of heterogeneity

Not so long ago, there was considerably less enthusiasm. Early in his career, Barzilai says that "nobody wanted to talk about aging." Even studies into age-related diseases

tended to underemphasize the physiological effects of aging.

For example, Wyss-Coray's early research centered on Alzheimer's disease, where he was especially interested in uncovering the influence of aging-related inflammation and immune dysfunction in disease pathology. "But in the Alzheimer's field at the time, there were only neuroscientists, and neuroscientists 30 years ago only studied neurons," he says. "Even glial cells were some fringe field."

And for those relatively few who were studying aging as a broader physiological process, much of the effort revolved around trying to find specific biological pathways that transform "young" cells into "old" cells. Harvard University aging researcher Vadim Gladyshev recalls attending talks earlier in his career where presenters would focus on individual genes as "key drivers" of age-related decline. "This is complete nonsense, because there are no key drivers of aging," he says. "It's the whole system that transitions." This is not to say that the genes identified in those studies were irrelevant, but rather that it is impossible to boil a complex and heterogeneous process like human aging down to a handful of genetic dials and toggle-switches.

Barzilai's own interest in aging was fueled by his own encounters with this heterogeneity as a young medical resident, where he saw dramatically different outcomes for elderly patients depending on the extent to which they "looked their age." "A person who's 75 years old that looks like he's 90, you know that you're dealing with end of life," says Barzilai, whereas one who looks 50 or 60 might have far greater resilience against injury or disease. These patients offered a clear object lesson: the rate and consequences of biological aging can vary widely between individuals, and differ considerably from what one might expect based purely on the passage of time.

Rejuvenating the field

More recently, rapid advances in technology have given researchers the capacity to conduct extensive "omic" studies that deeply survey the DNA, RNA, proteins, and metabolic activity of both tissues and individual cells. "Before, we had to pick something that we said 'represents aging,' and then we measured that," says Wyss-Coray. "Now we can measure thousands of [molecules] and show how the whole thing looks more like a younger organism." These capabilities have advanced the field's understanding of how and why biological systems break down over time.

Stanford University neuroscientist Tony Wyss-Coray received NOMIS funding to identify blood-borne molecules that can facilitate brain "rejuvenation."

In 2013, for example, University of California–Los Angeles geneticist Steven Horvath published a highly influential article in which he described the identification of a "molecular clock" for the aging process. His approach focused specifically on how patterns of genomic methylation—chemical modifications to chromosomal DNA—have a profound impact on the expression of nearby genes. "It was really, really impressive," says Gladyshev, noting that the clock offered remarkable accuracy and predictive power in terms of tracking chronological age in mammals.

Subsequent studies have built on Horvath's foundational work to assemble a more detailed dossier of age-related biomarkers. In December 2023, for example, a multi-national team led by Wyss-Coray published an extensive survey of blood-borne proteins secreted by 11 different organs in the human body. Remarkably, this analysis revealed that different organs age at different rates in different people, where the existence of an "age gap" between an organ and its owner can be a red flag for future disease risk. "If we measure that your kidney is 2 years older than you actually are, a person with that profile is more likely to have kidney failure in the future," says Wyss-Coray.

Finding such genomic, proteomic, and metabolomic signatures is extremely useful, but not the end-all-be-all of aging research. "Biomarkers are just a tool—our focus is actually to understand aging," says Gladyshev. Accordingly, his group has put considerable effort into deciphering the biological role of those biomarkers. He notes that some signatures of aging are purely chronological—marks on the biological calendar, rather than products of ongoing physiological decline. By separating out the biomarkers that specifically herald or arise from such decline, Gladyshev aims to develop even better clocks that distinguish "bad" aging from healthy aging.

Growing opportunities

What a difference a few decades make. In the early days of Barzilai's career in aging research—some 30 years ago—there was limited funding to support the development of anti-aging interventions. "The problem is the concept was not accepted widely, and therefore the funding was rate limiting," says Barzilai.

The money problem has not disappeared, but investment has been steadily growing in recent years. For example, in 2023, the U.S. National Institutes of Health (NIH) spent 6.4 billion USD on aging-related research—more than doubling its budget from 2016. Numerous other funding opportunities have emerged as well. "There's large foundations that just give money to people that studied basic mechanisms of aging," says Wyss-Coray, who has received support for multiple such projects through the NOMIS Foundation.

Barzilai says the funding situation is "still rate-limiting," but things are moving in the right direction. He and Gladyshev both point to the proliferation of startups dedicated to advancing aging research and turning their findings into anti-aging therapies, such as Altos Labs—a Jeff Bezos-backed venture that launched in 2022 with 3 billion USD in funding. This excitement is also drawing talent from other fields, and Gladyshev is optimistic about the future. "I feel things are improving," he says. "Smart people who used to go to study cancer or some other field—now they go to aging."

Sorting this out makes it possible to get a better handle on the different ways biology fails with aging. "To me, aging is the accumulation of negative consequences of life," says Gladyshev. The human genome naturally acquires mutations over a lifetime that can gradually increase the risk of developing cancer, and there are numerous other mechanisms by which the mere act of living can contribute to the accumulation of toxic biological byproducts. For example, cellular metabolism naturally produces chemical compounds that inflict damage on DNA, proteins, and other biomolecules. If the body cannot clear these cellular pollutants in a timely fashion, they have the potential to accelerate the aging process.

From mice to medicines

Turning these insights into interventions has not been easy, however. And to date, there are no supplements or clinically approved drugs with a proven capacity to mitigate or reverse aging-related damage in humans.

Much of the field's work to date has been built on a foundation of experiments in animal models. This typically means mice and rats, but other studies have employed more specialized species such as the African killifish, which undergoes the aging process in a matter of months, or the long-lived naked mole rat. "I think all of the major findings that we have in the field come from animal models," says Gladyshev. In 2023, his team published a study in which they systematically evaluated signatures of longevity in 41 different mammalian species to identify mechanisms that might offer good targets for anti-aging interventions.

This makes basic research with cell and animal models an essential component of contemporary aging research (see **Sidebar: Growing opportunities**)—but that does not make it sufficient. "If you're a mouse, I can take care of you—I can make you live longer by 24 percent," says Barzilai. "But it hasn't been translated." Accordingly, much of his group's work focuses on discovery in humans—for example, generating genetic and molecular profiles of centenarians and their families to understand the processes that contribute to their remarkably healthy aging. "They're sick for a very little time at the end of their lives," says Barzilai. "Thirty percent of our centenarians don't take drugs and don't have any disease." The pathways uncovered here can then be tested more systematically in animal models.

This combination of basic and clinical research is finally beginning to bear fruit, and the experts are finding cause for enthusiasm about a few particular drug candidates. Importantly, some of these already have an established clinical track record. Wyss-Coray highlights the various GLP-1 agonists like semaglutide—the diabetes drug that is now upending the weight loss world. "Approaches like that, which we know have clear effects on the molecular processes of aging, I think will probably be the first to be applied," he says.

This drug family was also among the most promising anti-aging candidates identified in a recent publication from Barzilai, in which he evaluated a wide range of U.S. Food and

Drug Administration (FDA)-approved drugs for their impact on various physiological "hallmarks of aging." Barzilai's list was topped by another category of drugs known as SGLT-2 inhibitors, such as empagliflozin. These drugs already have an excellent track record as approved diabetes therapies, with extensive testing data from both mice and people, but these and other studies also suggest more far-reaching health benefits as well. "In humans, when you look over 32 months versus placebo, they decrease incidence of diseases—renal-specific, cardiac-specific, and deaths from all causes," says Barzilai. He and his colleagues are also in the process of trying to launch the Targeting Aging with Metformin (TAME) program, a series of clinical trials designed to evaluate yet another FDA-approved diabetes drug with putative anti-aging benefits.

Towards tailor-made treatments

Success with a repurposed drug could give clinical aging research some much-needed momentum, but Wyss-Coray cautions that we are unlikely to see a single blockbuster drug that delivers broad benefits to the aging population. For example, the anti-aging effects produced by certain blood-borne factors in his team's parabiosis experiments were not evenly distributed across cell and tissue types. Furthermore, the parabiosis effect works both ways—donor serum from older animals can accelerate aging in young recipients, revealing the presence of other bloodborne molecules that promote decline and degeneration.

"You may need dozens of different factors that have beneficial effects, and you may at the same time have to block detrimental factors in an old organism to really mimic what we see with parabiosis," says Wyss-Coray. This scenario is further complicated by his team's recent findings on organ aging, which highlight how different organ systems in the same individual can decline at radically different rates relative to each other.

These findings suggest that effective aging therapy will require the same level of personalization currently applied in domains like cancer care. In some cases, this might involve simple lifestyle changes—for example, dietary interventions tailored to stall accelerated aging in the digestive tract or cardiovascular system. The evidence also suggests that many of the FDA-approved drugs now under consideration for repurposing may only benefit a subset of patients and organ systems. "We have clinical studies that have endpoints related to obesity, pre-diabetes, renal, cardiovascular, frailty, flu," says Barzilai, "and we're trying to

basically help clinicians to know when do you choose one drug over the other."

There is also considerable evidence for a firm upper bound to the human lifespan. To date, only one person—Jeanne Calment, who survived to 122 years—has been verified as living beyond the age of 120. Gladyshev says that scientists currently estimate a biological limit in the range of 138 years, after which the body will inevitably fail even with access to treatments and lifestyle adjustments that slow the aging process. "We expect a relatively small effect, maybe 5 or 10 percent," he says.

To get a bigger benefit, more radical technological solutions may be needed. For example, Wyss-Coray is particularly excited about ongoing research into so-called reprogramming methods that can literally roll back the clock. This concept is well-validated in cell culture, where certain cocktails of chemicals or proteins can cause fully mature cells to transform into "young" stem cells with many, if not all, of the properties of embryonic tissue. If that process could be tightly controlled—to avoid, for example, the risk of creating unwanted reservoirs of tumorigenic stem cells—a similar "partial reprogramming" strategy could significantly extend the range of human longevity or delay the onset of age-related disease.

Wyss-Coray sees real opportunities to help ensure that people can stay healthy and active at the upper limits of human aging, but believes that greater leaps in rejuvenation are unlikely to occur in the near future. And this may not be a bad thing, he adds. "What would we do with our lives if we live twice as long?" he asks. "It has so many implications in terms of the resources on this planet—everything becomes philosophical very quickly."

Sponsored by

NOMIS
FOUNDATION
Creating the Spark

NEWS

29% Share of 3387 biological research articles in 15 high-impact journals with bar graphs that do not accurately reflect the findings. The study authors recommend data visualization guidelines for scientists, and checklists and training for journal editors. (bioRxiv)

IN BRIEF

Edited by **Jeffrey Brainard**

A worker at Los Alamos's Plutonium Facility uses a glovebox to safely adjust a piece of equipment.

NUCLEAR WEAPONS

U.S. resumes making nuke triggers

os Alamos National Laboratory (LANL) last week announced it had manufactured its first plutonium "pit"—used in the core of a thermonuclear warhead to initiate an explosion—since the United States largely halted such production in 1989. Under a nuclear arsenal modernization plan launched in 2008, the lab will scale up production to 30 pits per year, with an additional 50 to be produced annually starting in the mid-2030s at the Savannah River Site in South Carolina. But the work faces opposition from critics who say it could help fuel a new international nuclear arms race and also risks the health of workers and the environment. Last week opponents won a ruling in federal district court in South Carolina when a judge ruled the U.S. Department of Energy failed to adequately consider other options for locating the production facilities. The judge is considering a request to pause production at LANL. The United States halted pit manufacturing in 1989 at the Rocky Flats Plant, near Denver, after an inspection revealed hazardous waste contamination.

French COVID-19 doctor barred

SCIENTIFIC INTEGRITY | The French Medical Association has banned controversial microbiologist and physician Didier Raoult from practicing medicine for 2 years beginning in February 2025, according to news reports. Last week the association announced it had found that Raoult breached the public health code by promoting the antimalarial drug hydroxychloroquine as a COVID-19 treatment despite a lack of evidence, and by conducting a hydroxychloroquine study involving more than 30,000 patients without approval from French authorities. The decision supersedes a regional chapter's move in 2021 to reprimand Raoult but allow him to continue practicing. A separate criminal investigation of his actions continues. In September 2022, Raoult, now 72, retired as director of the Hospital Institute of Marseille Mediterranean Infection and currently has no official role there. Raoult has grounds to appeal, his lawyer Fabrice Di Vizio told *Libération*.

UK Biobank adds health data

GENOMICS | Researchers mining the UK Biobank, one of the world's largest human genetics databases, will soon have access to a vastly expanded set of health data about its 500,000 volunteers. Launched in 2006, the widely used resource for linking genes to disease risks and traits currently share only hospital records for most participants. Although each person who signed up consented to share their National Health Service primary care records on diagnoses, prescriptions, and referrals, physicians with England's NHS worried the data would not be secure. But last week, the government announced NHS England will share deidentified data with the UK Biobank and other research studies via a secure platform, OpenSAFELY, set up during the COVID-19 pandemic to enable studies of the disease. Access to the additional data "will transform the research potential of our database overnight," said UK Biobank head Rory Collins.

A London police investigator dusts a confiscated elephant tusk for smugglers' fingerprints.

FORENSICS

Magnetic clues draw police to ivory smuggling suspects

Police have notched progress in the fight against ivory smugglers and poachers using an emerging forensic tool: a magnetic powder that reveals fingerprints on tusks taken from elephants. The technology, reported in 2016 by researchers at King's College London working with the city's Metropolitan Police Service, allows investigators to obtain fingerprints from confiscated ivory up to 28 days after it was touched, compared with 2 to 3 days using conventional powder—increasing the chances of identifying those involved in its illegal trade. The International Fund for Animal Welfare has distributed more than 200 kits containing the magnetic powder to authorities in 40 countries, enabling at least 15 arrests, according to the U.K. Natural History Museum. This week the museum, which produces the annual Wildlife Photographer of the Year awards competition, gave one of the prizes to photojournalist Britta Jaschinski, who took this photo of the procedure at London's Heathrow Airport. Poachers annually kill an estimated 20,000 African elephants, which are classified as endangered.

EU to delay deforestation law

ENVIRONMENT | The European Commission has moved to delay, by 1 year, the implementation of controversial agricultural import rules aimed at curbing deforestation. The rules, approved by the 27-nation European Union in 2023 and set to take effect late this year, require firms importing seven major commodities—including palm oil, soy, coffee, cocoa, and rubber—to certify that their production did not contribute to deforestation or human rights violations. Numerous nations and some scientists have criticized the rules, arguing traders and farmers didn't get enough time to prepare and lacked the high-quality forest mapping data needed to comply. Those critics welcomed the delay, which the European Parliament is expected to approve. But green groups were disappointed, with the World Wildlife Fund calling it a "green light for deforestation to continue."

Brazil hit by fire, drought

CLIMATE | A satellite monitoring system has detected more than 47,000 forest fires this year in Brazil—the most in at least a decade. The numerous blazes are the result of continued burning of Amazon forest to clear land for agriculture, coupled with persistent dry conditions caused by climate change, says the nonprofit World Resource Institute's Global Forest Watch project. The satellite data, reported through mid-September, also indicate that the area affected by fires has increased by 85%, compared with the annual average since 2019. At the same time, Brazil is currently enduring its longest drought in 70 years, impacting more than 80% of its territory, the institute noted last month.

India research prizes questioned

PROFESSIONAL REWARDS | Researchers in India are asking the government of Prime Minister Narendra Modi to clarify how it awards top science prizes after reports that officials vetoed prizes for scientists who had criticized Modi's policies. Nearly 180 researchers, including the former directors of several prominent science institutes, signed a letter last month to the government's principal science adviser, Ajay Sood, presenting the request. The controversy has its roots in the Modi government's move, begun in 2022, to replace hundreds of prizes and cash bonuses for scientific accomplishments with a set of four awards called the Rashtriya Vigyan Puraskar. The goal, the government said, was to ensure "transparency and fairness in the entire selection process" for the prizes. But some researchers say the ministry scaled back the list of awardees for political reasons. Others are willing to give the government the benefit of the doubt, saying other factors may have been involved. When *Science* went to press, Sood had not responded to a request for comment.

IN DEPTH

ACADEMIC WORKFORCE

How racial bias can taint tenure decisions

Black and Hispanic professors fare worse when voters are less familiar with their work

By **Katie Langin**

Going up for tenure and promotion can be nerve-wracking for any academic. It's supposedly an unbiased evaluation of a scholar's work, but other dynamics can come into play. New research highlights the impact of one of those factors: race. In more than 1500 tenure and promotion decisions at five U.S. research-intensive universities, Black and Hispanic faculty members received more negative votes than white and Asian colleagues who were no more productive.

It's "some of the most robust evidence of racial bias in promotion and tenure," says Damani White-Lewis, an education expert at the University of Pennsylvania who has studied racial disparities in academia but wasn't involved in the new research. "Now that we have actual data across multiple institutions, it makes for a more compelling case that we need to laser in on this."

Data on tenure and promotion committee votes are notoriously difficult to come by, as institutions treat them as highly confidential. So, the researchers behind the new study, published last week in *Nature Human Behaviour*, partnered with administrators at the five universities, who anonymized and compiled records of tenure and promotion decisions in all disciplines be-

tween 2015 and 2022. "That gave us enough of a sample size where we can look at racial disparities," says psychologist Juan Madera of the University of Houston.

His team compared outcomes for faculty members from groups underrepresented in academia—Black and Hispanic—with those who are not: white and Asian. Too few faculty self-identified as Native American, Native Hawaiian, or "other" to allow analysis of these groups.

Race did not appear to influence the typical first step of the review process, when colleagues in the applicant's department vote on whether they deserve promotion to associate professor with tenure, or from associate to full professor. But bias began to emerge at the next step, a vote by the college—a broader group of colleagues from various departments who are likely less familiar with the candidate and their work.

At that stage, Black and Hispanic faculty with high h-indices—a measure of scholarly productivity and impact—actually fared better than similarly productive white and Asian faculty. But when the college committees were evaluating faculty with lower h-indices, evidence of bias against Black and Hispanic scholars surfaced—especially for women. In those situations, "underrepresented minorities, particularly women of color, are held to a different standard,"

Madera says. Overall, Black and Hispanic faculty received 7% more negative votes from college committees and were 44% less likely to receive unanimous "yes" votes than white and Asian colleagues.

"We can only speculate why that is," says lead author Theodore Masters-Waage, a University of Houston postdoc, but it could be related to the lack of disciplinary expertise on the college committees. "Where there's less detailed information and people have less to go on ... they start relying on the stereotypes that they might hold about someone."

The committee votes alone don't determine whether a faculty member receives tenure and promotion; typically the provost makes the final recommendation to the university's president and board of regents. Still, they're critically important because the vote percentages can sway the provost. Even one negative vote can serve as a "poison pill," causing others to look at a case more closely, Masters-Waage says. As Madera notes, "That's why looking at unanimous votes is so important, because if all the committee members support it, it's a really strong case."

Even faculty members who go on to receive tenure and promotion can feel the impact of negative votes, Madera adds. The votes are supposed to be confidential, but committee members sometimes gossip. "It is an awful feeling to know that not all of

your colleagues value your research or support your career trajectory."

Cinzia Cervato of Iowa State University says the study addresses a common misconception by faculty who are not part of a historically underrepresented group: that the process is the same for everyone. By showing that's not the case, the findings could spur change, says Cervato, a geoscientist who has studied the tenure process.

Some, however, wish the research team hadn't lumped Asian faculty members together with their white colleagues. "By folding into one group, this study ignores and denies ongoing disparities and racial biases that face Asian scholars in academia," Catherine Tan, a sociologist at Vassar College, wrote on the social media platform X. Madera appreciates the concern, adding that in a separate analysis not reported in the paper, the team saw no difference in voting patterns for white and Asian candidates. "Our paper only reports what the data showed." But that doesn't mean discrimination and bias against Asian faculty doesn't exist, he says.

The research team identified some possible remedies for the observed racial disparities, including adjustments to how universities solicit external review letters from disciplinary experts when a faculty member goes up for tenure or promotion. After analyzing review letters, Madera's team found that when they emphasized a candidate's scholarly productivity, voting outcomes for Black and Hispanic women faculty improved.

Most research-intensive universities in the United States ask external reviewers about a faculty member's scholarly productivity, Cervato and her colleagues have found. But many also instruct them to discuss a scholar's contribution to teaching and service. Cervato thinks that practice should stop—especially for teaching, because reviewers can only assess teaching skills based on student evaluations. "Student evaluations are very well known that they're biased against women, faculty who were not educated in United States, faculty of color, faculty with accents," she says. "It's a flawed instrument."

Ultimately, Masters-Waage acknowledges, there isn't a "perfect pile of solutions for universities to implement" to fix the problem. But he and Madera are hopeful the study will point to problem areas—such as voting at the college level and the language in external review letters. "There's definitely areas you could focus your effort on," Masters-Waage says.

Simply having the study out there is a good first step, Cervato says. "The first step of actually countering bias is awareness of bias." ∎

PARTICLE PHYSICS

Fermilab gets a not-quite-new boss to tackle its troubles

University of Chicago will again lead a group to manage the United States's only dedicated particle physics lab

By **Adrian Cho**

The United States's celebrated but troubled particle physics lab, Fermi National Accelerator Laboratory (Fermilab), will have a new manager that looks much like the old one. Starting on 1 January 2025, a consortium led by the University of Chicago will succeed a smaller one—also led by UChicago—that had failed multiple performance reviews. The lab's owner, the U.S. Department of Energy (DOE), put the contract for managing Fermilab up for bids 22 months ago and announced the new contractor on 30 September.

Although widely expected, the choice is sure to inflame Fermilab's harshest critics, such as those who in July anonymously posted online a savage critique cataloging many supposed instances of mismanagement and malfeasance at the lab. "We asked for change and there is zero change," says a Fermilab physicist who co-authored the critique and requested anonymity to avoid retribution. But other observers say the university has brought in new partners with capabilities that should help right the ship.

Paul Alivisatos, president of UChicago and a chemist who was once director of DOE's Lawrence Berkeley National Laboratory, pledges the new contractor will respond to physicists' complaints. "[W]e have listened closely to the heartfelt concerns that some colleagues and friends in the Fermilab community have raised," he says in a statement.

"The new … team will prioritize addressing these challenges with a fresh perspective."

The situation underscores a tricky truth for DOE's Office of Science: The fee for running each of its 10 national labs is typically so small, and the knowledge required so specialized, that often the office has few choices for a contractor.

Founded in 1967, Fermilab has a staff of 2160 and an annual budget of $739 million. For decades it boasted the world's highest energy atom smasher, the Tevatron Collider, until it was eclipsed by Europe's Large Hadron Collider in 2010. Fermilab physicists discovered the massive bottom and top quarks and the wispy tau neutrino. More recently, they confirmed that the muon may be slightly more massive than predicted, a potential sign of new physics.

In recent years the lab has faced internal troubles. Many researchers say the future of U.S. particle physics depends on a huge experiment, under construction, to shoot neutrinos from Fermilab in Illinois to a huge subterranean detector 1300 kilometers away in South Dakota. But its cost has ballooned from $1.5 billion to $3.3 billion. And according to DOE reviews, the lab has struggled to write proper contracts and pay its bills. In May 2023, a worker was badly injured in a fall, and in August the lab furloughed most staff for 5 days because of a budget problem.

DOE declined to renew Fermilab's current contractor, Fermi Research Alliance (FRA) LLC, which has run the lab since 2007 and

Fermi National Accelerator Laboratory's management has faced criticism from outsiders and staff.

comprises UChicago and the Universities Research Association (URA), a consortium of more than 90 universities. The new contractor will be Fermi Forward Discovery Group LLC, which includes UChicago, URA, and the companies Longenecker & Associates and Amentum. The group "brings together the best parts of Fermilab with two new industry partners … who bring broad expertise and a deep bench of personnel from across the DOE complex," Alivisatos says in his statement.

Founded by a former DOE administrator, Longenecker specializes in helping federal agencies improve performance and project management. Amentum is an advanced engineering firm with more than 53,000 employees that works in the energy and military sectors. Both firms have taken on projects in DOE's Environmental Management program and National Nuclear Security Administration, which maintains the United States's arsenal of nuclear weapons. "I'm optimistic because I know Amentum and I know Longenecker," says William Madia, former director of two of the Office of Science's national labs. "They bring better operational discipline and better project management discipline to a lab that needs it."

DOE declined to say how many bids the Fermilab contract attracted, but outside observers say there were at most two others. One reason UChicago faced little competition might be the meager fee the Office of Science pays lab operators. FRA's performance-dependent fee was $5 million for this fiscal year—far less, as a fraction of the lab budget, than DOE pays for operating a weapons lab. "The fees are small," notes James Decker, who served as deputy director of the Office of Science from 1985 to 2007, "and the department has expectations that the contractor is going to plow some of that fee back into the laboratory."

Given the paper-thin margin, most private companies won't bid to operate a DOE science lab, Madia notes. That leaves DOE to choose among nonprofits, and multiple Office of Science labs are deeply entwined with universities. Some even sit on land owned by the universities that manage them. A university will typically run a lab for the prestige, but may lack the resources to manage big projects, multiple observers say.

Even if the new Fermilab contract does not clean house, some of the lab's critics say they're hopeful things will improve. Bill Barletta, a physicist at the Massachusetts Institute of Technology who wrote a forward to the so-called whistleblower report, thinks that with its new partners, UChicago can cure Fermilab's ills. "I have to be optimistic," he says. "It's my field, I want it to survive." ■

NOBEL PRIZES

Duo honored for tiny RNAs key to development and disease

Probes of abnormal worm development led to discovery of microRNAs and their role in gene expression

By **Catherine Offord**

Another year, another Nobel for RNA research, this time recognizing the discovery of tiny molecules called microRNAs that help regulate gene expression inside cells. Victor Ambros from the University of Massachusetts Chan Medical School and Gary Ruvkun from Harvard Medical School and Massachusetts General Hospital (MGH) share the Nobel Prize in Physiology or Medicine for work they started together in the 1980s, studying gene regulation in worms.

Unlike messenger RNAs, the focus of research that led to several COVID-19 vaccines and won last year's prize, microRNAs are very short—about 22 nucleotides—and bind to mRNAs, thereby determining which mRNAs get translated into proteins. In the past couple of decades, scientists have discovered tens of thousands of microRNAs across the tree of life and are uncovering roles for them in development and disease.

MicroRNA-guided gene regulation was a "new physiological mechanism that no one expected, completely out of the blue," Olle Kämpe, a member of the Nobel Committee, said after the announcement. "It shows that curiosity research is very important."

In the 1980s, Ambros and Ruvkun were postdoctoral fellows at the Massachusetts Institute of Technology in the lab of Robert Horvitz, who himself would win a Nobel Prize in 2002. Studying the roundworm *Caenorhabditis elegans*, they found that a gene called *lin-4* seemed, somehow, to block the expression of another gene, *lin-14*. Later, Ambros's group at Harvard University and Ruvkun's group at MGH and Harvard Medical School revealed how: In back-to-back papers in 1993, they reported that *lin-4* produced a short RNA molecule that didn't code for any proteins and had a sequence that was complementary to *lin-14* mRNA, meaning it could bind that mRNA and block it from making protein. "It was

> ### "It's a completely new physiological mechanism that no one expected, completely out of the blue."
>
> **Olle Kämpe**
> Nobel Committee

surprising that there could be enough information contained in only 22 bases in the genome for this microRNA to regulate another gene so precisely," Ambros says.

Yet microRNAs didn't receive wide attention until 7 years later, when Ruvkun discovered another called let-7. Unlike lin-4 microRNA, let-7 was found widely across the animal kingdom. The discovery launched a new field to understand how these molecules function, not just in normal development in plants and animals, but in disease. "Components of microRNA pathways are mutated in dozens of diseases," says Witold Filipowicz, a biochemist at the Friedrich Miescher Institute for Biomedical Research.

Attempts to harness these findings in treatments have yet to pay off, and toxic side effects appear to have halted some trials, but Filipowicz and others say they remain optimistic about microRNA's prospects. Furthest along are efforts to adjust gene expression within diseased cells by giving drugs that block microRNAs or by using microRNAs themselves.

One challenge—that each microRNA can potentially bind to hundreds of different mRNAs—could end up being an advantage in treating complex diseases such as cancer, says Purdue University biologist Andrea Kasinski. "These microRNAs have the power to target many genes. We just have to make sure they're the right ones."

This year's Nobel is yet another win for a celebrated model organism. "It's striking that the work on the simple worm *C. elegans* is still resulting in Nobel Prizes," says Dutch molecular biologist Ronald Plasterk, who worked on microRNA before becoming a politician. "So much basic biology has been discovered in this simple animal of only 959 cells." Speaking after the award news broke, Ruvkun also emphasized the role of curiosity in the research. "We weren't thinking, 'This is going to win a Nobel Prize,' we were thinking, 'This is really interesting.'" ■

In Sweden, John Hopfield and Geoffrey Hinton were announced as winners of the Nobel Prize in Physics.

NOBEL PRIZES

Physics award honors pioneers in artificial intelligence

Models based in statistical physics led to today's neural networks that power chatbots and self-driving cars

By **Adrian Cho**

Artificial intelligence (AI) is everywhere these days, even in the recondite realm of the Nobel Prizes. This year's prize in physics honors John Hopfield of Princeton University and Geoffrey Hinton of the University of Toronto for their work on artificial neural networks, the models of brainlike circuitry that can learn to recognize patterns and that have powered the ongoing AI revolution. The $1 million prize cites the origins of that work in statistical physics.

The award appears to have taken one of the new laureates by surprise. "I'm flabbergasted," Hinton said during the announcement this week. But statistical physicists are pleased to see their abstract discipline grab a share of the limelight. "For us, it's super-great," says Aurélien Decelle of the Complutense University of Madrid. "It's recognition at the broader level that what we're doing matters a lot."

Yet the award seems likely to elicit sniping that the cited work is not physics. "Not more than 10 minutes ago, I heard the same comments in the corridor," Decelle says. The prize also does not recognize more recent advances in neural networks that have enabled AI to power everything from chatbots to self-driving cars—and even perform research.

In the brain, a neuron fires electrical pulses that can either stimulate other neurons to fire or inhibit them from doing so, depending on the nature of the connections. So, an artificial neural network is just a collection of abstract nodes connected by lines with "weights" that determine whether a connection excites or inhibits. A magnetic material can be modeled the same way. In the simplest case, each node represents an atomic "spin" that can point up or down, while each connecting line, or "edge," represents a magnetic interaction that nudges the linked spins to align in the same direction or in opposite ways, mimicking excitation and inhibition.

In 1982, Hopfield, a polymath who hopped from theoretical physics to genetics to neuroscience, explored the capacity of such a system to store patterns in its lowest energy arrangements. Hopfield showed that under the right conditions, many patterns could be stored. Were the up-and-down spins black-and-white pixels, these patterns might be rudimentary images. What's more, Hopfield showed how such "memories" might be landed on and recalled if the system started in some other configuration.

Just a few years later, Hinton, a computer scientist and cognitive psychologist, expanded on Hopfield's network. Hopfield had considered a deterministic model with no random noise. In contrast, Hinton and colleagues included such noise by imagining spins with a nonzero temperature, jiggling with thermal energy that could cause them to flip. The probability of such flipping is determined by the statistical theory discovered by Ludwig Boltzmann, so Hinton's network was known as a Boltzmann machine. Such noise is integral to the function of neural networks. Hinton also studied how sample patterns could tune the weights of the edges in the network. And, crucially, he included "hidden" nodes that greatly sped up the training of the network, Decelle says.

Neither the Hopfield network nor the Boltzmann machine closely resembles the neural networks that underpin AI algorithms today, notes Austen Lamacraft, a statistical physicist at the University of Cambridge. In current networks, the down-or-up spins are replaced with continuous variables that can have values between –1 and 1. Today's deep neural networks also have a more complex, directional structure, with an input layer of nodes, an output layer of nodes, and many layers of hidden nodes in between. Perhaps most important, modern networks are trained using a technique called backpropagation, during which weights from an output layer are passed back to an input layer.

Whether all of this counts as physics may depend on whom you ask. Particle physicists and cosmologists may "be frothing a bit," Lamacraft says, but he notes that neural networks "have long been considered a part of statistical physics."

The prize could also raise howls from computer scientists who might want recognition for later, more technologically relevant work. However, Danica Kragic Jensfelt, a computer scientist at Sweden's KTH Royal Institute of Technology, notes that in 2018, the $1 million A. M. Turing Award in computer science recognized Hinton along with computer scientists Yoshua Bengio and Yann LeCun for their work in AI. "The Turing prize has been given for the computer science part," Kragic Jensfelt says. "This prize is given to people who did the physics part."

Hinton predicts the AI revolution will do for intellectual work what the Industrial Revolution did for physical labor. But he also worries about the downsides for society. He left a job at Google in 2023 because of fears of unintended consequences, concerns he raised again this week. "We also have to worry about a number of possible bad consequences, particularly the threat of these things getting out of control." ∎

CONSERVATION BIOLOGY

Seed disperser declines threaten Europe's plants

First broad look at conservation status of animals that transport seeds raises alarms

By **Erik Stokstad**

When hunters wiped out most fruit-eating birds in the tropical forest of Lambir Hills National Park on western Borneo by the 1990s, the skies grew duller—and in a few years so did the forest. Without birds to spread their seeds, the diversity of fruit-producing plants declined, illuminating the critical importance of seed dispersal for ecosystem health. The movement of seeds across the landscape in animals' guts is a "glue that keeps ecological communities together," says Jordi Bascompte, an ecologist at the University of Zurich. Now,

Lynn Dicks, a conservation scientist at the University of Cambridge. "You just think, 'Why has no one done that before?'"

Figuring out which animals disperse which plant seeds requires analyzing hundreds, if not thousands, of species interactions. Sara Mendes, a Ph.D. student in the lab of Ruben Heleno, a community ecologist at the University of Coimbra, took on the gargantuan task. She scoured thousands of studies in 26 languages that mentioned terms such as seed dispersal or were focused on one of the more than 900 European animals likely to consume seeds. "The project required a certain amount of craziness to be accomplished," she says.

ture (IUCN) or are dwindling in numbers. For example, the garden warbler (*Sylvia borin*), a common migratory bird that spreads the seeds of some 60 plant species, is declining across Europe. The same is true for the redwing (*Turdus iliacus*), some populations of which migrate thousands of kilometers and could move seeds along part of their journey. "We should not be afraid of using the word crisis," Heleno says, given the number of at-risk species.

The study does not detail how the crisis is affecting ecosystems. For one thing, IUCN assessments haven't yet been conducted for 67% of the plant species in the data set. But Mendes and Heleno did find that more than 60% of the plants had five or fewer animals that eat and distribute their seeds, which could make them particularly susceptible to the decline or disappearance of any of these critical seed dispersers.

And some plants do seem to be in trouble. The authors came up with a list of nearly 80 "very high concern" interactions, in which both the plant and the animal are threatened or declining. The list includes the European fan palm (*Chamaerops humilis*), a stumpy tree that can take decades to reproduce. Across its Mediterranean range, the plant relies on the dispersal services of 10 species, including the European rabbit (*Oryctolagus cuniculus*), a species listed by IUCN as "endangered" in Spain and Portugal. When assessing how to help plant species that are in trouble, "We need to make sure we're looking after the dispersers that are doing a lot of the work," Dicks says.

Despite the massive data set Mendes was able to compile, much remains unknown about the extent of the problem. It's a mystery which animals move the seeds of some rare plants, for instance—in spite of centuries of natural history observations in Europe. Even for common, well-studied species, it's not always apparent which dispersers are most important and whether other species might take over if they were to decline or disappear.

Similar relationships are probably unraveling on other continents, too, including North America. But the picture there is even less clear, says ecologist Haldre Rogers of the Virginia Polytechnic Institute and State University. "We really have no idea which plants are missing their dispersers or at risk of that." ∎

The redwing, shown here eating berries from a European holly (*Ilex aquifolium*), is listed as "near threatened."

ecosystems in temperate climates seem to be coming unglued as well.

On p. 206, a team reports that at least one-third of European plant species could be in trouble because most of the animals that move their seeds are threatened or declining. The study is "brilliant and compelling" but also "alarming," says Pedro Jordano, an ecologist at the University of Seville. The decline in seed dispersers—not just birds, but also mammals, reptiles, and ants—could jeopardize the ability of plants to expand their range to cope with climate change or recover after wildfire, he adds, especially in Europe's highly fragmented landscape. "It's a fantastic analysis," says

Mendes compiled a list of 592 native plant species that have adaptations—mainly fleshy fruit—to encourage animals to spread their seeds, as well as 398 animals known to transport those seeds. Many of the dispersers eat the seeds of multiple plants, so she ended up with a data set that included more than 5000 pairings of plants and their animal dispersers.

The next step was to look at how the species are faring. The team found that in all major biogeographical areas in Europe, from the Mediterranean to the Arctic, more than one-third of seed-dispersing animal species are ranked as threatened by the International Union for Conservation of Na-

A Pacific Ocean sunset in 2023, when record heat washed over the planet.

CLIMATE SCIENCE

El Niño fingered as likely culprit in record 2023 temperatures

Research suggests swings in Pacific Ocean can account for planet's sudden and perplexing temperature jump

By **Paul Voosen**

For the past year, alarm bells have been going off in climate science: Last year's average global temperature was so high, shooting up nearly 0.3°C above the previous year to set a new record, that human-driven global warming and natural short-term climate swings seemingly couldn't explain it. Some, like famed climate scientist James Hansen, suggested Earth is entering an ominous new phase of accelerated warming, driven by a rapid decline in sunlight-dimming air pollution. Others, like Gavin Schmidt, director of NASA's Goddard Institute for Space Studies, said the rise might represent a "knowledge gap," some new climate feedback that might tip the planet toward a future even warmer than models predict.

Now, a new series of studies suggests most of the 2023 jump can be explained instead by a familiar climate driver: the shifting waters of the tropical Pacific Ocean. The combination of a 3-year-long La Niña, which suppressed global temperatures from 2020 to 2022, followed by a strong El Niño could account for the unexpected temperature jump, the work suggests. "Earth can do this," says Shiv Priyam Raghuraman, a climate scientist at the University of Illinois Urbana-Champaign, who led one study.

During La Niña, strong trade winds push warm surface water west along the equator toward Indonesia and pull up a fountain of deep, cold water in the eastern Pacific that helps cool the planet. During El Niño, the winds collapse, allowing warm water to slosh east and shut off the ocean air conditioner.

Last year, analyses suggested the combination of global warming and El Niño fell far short of explaining 2023's heat, leading to the worries that something else might be at play. But Raghuraman and his co-authors weren't convinced those studies captured El Niño's full potential. Indeed, looking back, they found that 1977 was an awful lot like 2023, when temperatures rose by more than 0.25°C after a multiyear La Niña tipped into an El Niño.

But that's just 2 years out of the 70-some for which El Niño records exist. To generate better statistics, Raghuraman and his co-authors compiled every climate model run they could find that simulated the planet in a steady state, unperturbed by humanity, totaling 58,021 years of simulations. Then they looked to see how often temperature spikes higher than 0.25°C occurred.

Their study, accepted for publication in *Atmospheric Chemistry and Physics*, showed such spikes were rare, happening only 1.6% of the time, nearly always during an El Niño. But when a long La Niña set the stage, the probability of a spike jumped to 10.3%. And during those model years, the geographic pattern of warming often matched what occurred last year, such as a large increase in the tropical Atlantic Ocean. The models show big El Niño jumps are rare but possible, Raghuraman says. "We're not missing something."

The result lines up with another study, published in August in *Communications Earth & Environment*, which compared sea surface temperatures in 2023 and the recent past. If global warming was accelerating, that trend would also be seen in the oceans. And although the oceans were anomalously hot in 2023, they were only a little warmer than during an El Niño in 2015 and 2016, says study co-author Marianne Tronstad Lund, research director at Norway's Center for International Climate and Environmental Research. "We don't find signs of any rapid acceleration," she says.

Recent runs of a "climate pacemaker" experiment at the Scripps Institution of Oceanography also implicate El Niño as the main culprit in the globe's extra heat. Unpublished results from the experiment, which feeds real-world Pacific temperatures into a climate model, re-created temperature patterns similar to those observed last year, says Scripps climate scientist Shang-Ping Xie, with some exceptions over the North Atlantic.

Meanwhile, several studies have found that cleaner, clearer air due to falling pollution from China and lower sulfur marine fuels made only a small contribution to last year's temperatures. One study, submitted to *Atmospheric Chemistry and Physics*, found that declining pollution could raise global temperatures by 0.03°C over the next 20 years, with the strongest effect not occurring until later this decade. It's not nothing, says study co-author Duncan Watson-Parris, an atmospheric physicist at Scripps, but far too little to explain last year. Taken together, says Mika Rantanen, a climate scientist at the Finnish Meteorological Institute, the results are "a good reminder that it was indeed El Niño that was the major player."

Yet the timing of 2023's heat remains odd, Schmidt says. It came on faster and stronger than in normal El Niño years, and it lingered far longer, even as Earth tips again toward a La Niña. Perhaps the biggest wild card is the increasing amount of sunlight satellites have detected hitting the planet's surface over the past decade (*Science*, 12 April, p. 147). Falling pollution can only explain some of the increase; the rest might be due to reduced cloudiness or changing surface reflectivity. How much the extra sunlight might have heated the planet in 2023 is unclear.

The new studies aren't the final word on the problem, Schmidt says. Even as the eastern Pacific cools off again, the debate within climate science continues to simmer. ∎

FEATURES

COMING OF AGE

Twenty years after the ballyhooed discovery of graphene, the atom-thin carbon sheets are finding their footing

By **Mark Peplow**, *in Manchester, England*

On a rare sunny day in northern England, the National Graphene Institute (NGI) here gleams like a five-story block of obsidian. Squeezed into the University of Manchester's sprawling downtown campus, the research center is clad in almost 2000 lustrous black panels with small hexagonal perforations—an architectural nod to the structure of the atom-thin sheet of carbon that gives the building its name.

NGI exists because graphene was first isolated a short walk away in a University of Manchester lab. Andre Geim and Konstantin Novoselov presented it to the world 20 years ago this month and later won a Nobel Prize for the work. Since its unveiling, billions of dollars of R&D funding have flowed to graphene, in a global race to exploit its peerless properties. It is better at carrying electricity than any metal, a superb heat conductor, and hundreds of times stronger than steel—selling points trumpeted in the marketing materials of universities and companies alike.

Early on, researchers were not shy about promising graphene breakthroughs, with predictions that it would enable superthin rollable TVs and space elevators, and even supplant silicon in computer chips. "Expectations were very, very high," Geim says. "The companies I was involved in were mostly based on hype."

For all its allure, graphene had drawbacks, not least that it is difficult to incorporate into mass-produced devices without sacrificing its much-vaunted properties. Many companies came and went, taking their futuristic graphene dreams with them.

Graphene supply has long outstripped demand, and major manufacturing companies still haven't leapt into graphene production, says Conor O'Brien, a technology analyst at IDTechEx, a U.K. market research consultancy that tracks the graphene industry.

Yet today, some graphene businesses seem to finally be finding their footing. That's partly because the name "graphene" is now applied to a plethora of other substances—cheaper forms of carbon stacked in multiple atomic layers or decorated with various chemical appendages. These heterodox forms of graphene can now be found not only in consumer electronics, but also in concrete, pickup trucks, and brain devices. "Graphene is now hundreds, maybe even a thousand different things," says Peter Bøggild, a graphene researcher at the Technical University of Denmark.
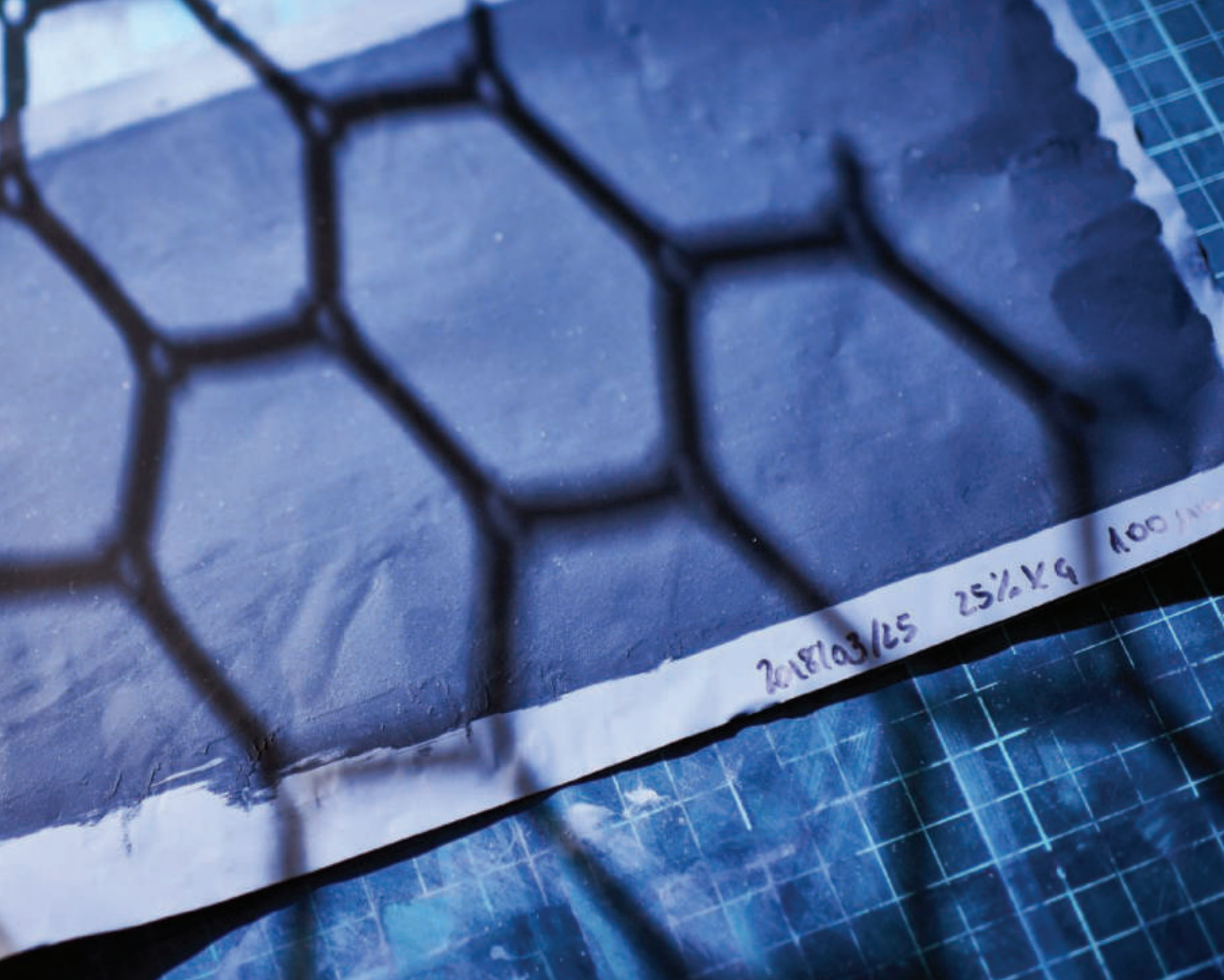
But what of the iconic graphene monolayer itself? Even it may be emerging from its slough of despond. Companies are getting better at making it cheaply and consistently, and they are using atom-thin graphene to build magnetic sensors, optical communication chips, and medical diagnostic devices, for example. "It's gone through the hype cycle," says Natasha Conway, research director at U.K. graphene device manufacturer Paragraf. "It's gone up the hill, down into the big valley—and now it's coming out as people see that it really does have value."

GRAPHENE'S ORIGIN was famously humble: Geim and Novoselov's team used ordinary Scotch tape to peel thin fragments from a chunk of graphite. They then sifted through the flakes to find those that were just 1 atom thick. They soon realized that graphene's 2D honeycomb structure makes it a seriously weird material.

For instance, it is neither a metal, nor an insulator, nor even a true semiconductor. In conventional semiconductors, electrons are confined to a range of energies (known as the valence band) that prevents them from conducting electricity—until a jolt of energy boosts them into the conduction band, enabling current to flow. Instead, graphene is a semimetal, somewhat like a semiconductor but with zero band gap, so the tiniest of nudges makes it conduct electricity. Indeed, the unusual arrangement of graphene's energy states makes its electrons behave as if they have no mass, and helps them flow through graphene like greased lightning, experiencing very little electrical resistance.

Other researchers showed how graphene's flawless array of strong chemical

The National Graphene Institute (NGI, left) is near the Manchester labs where graphene was discovered. In an NGI lab, graphene ink coats a metal foil (right).

bonds gives it great strength, and how its honeycomb structure allows atomic vibrations to ripple freely through the material, rapidly transporting heat as they go. It also turned out to be keenly sensitive to small magnetic fields, which sweep its mobile electrons to one side in a phenomenon known as the Hall effect.

Within a few years of its isolation, researchers had largely explored the fundamental science of graphene. "Essentially, graphene was done and dusted by 2007," says Geim, who co-authored a review that year. "I wanted to call it 'Graphene is dead,' but I was not allowed," he recalls with a wry smile.

Hand-peeled microscopic flakes of graphene were perfect for laboratory studies, but hardly suitable for commercial products. The "black gold" rush really accelerated after 2009, when researchers in Texas found that a common industrial process called chemical vapor deposition (CVD) could be used to grow strong yet flexible graphene monolayers on copper foil. The copper acts as a catalyst, liberating carbon atoms from a feedstock of methane at about 1000°C to build up the graphene.

Companies quickly scaled up this production method. In 2013, for example, Bluestone Global Tech in New York state claimed it could make at least 20 square meters of CVD graphene per day, and announced it would open a new production plant at the University of Manchester. Bluestone and other companies promised that graphene's conductivity, flexibility, and trans-

clad in bunny suits operating multimillion-dollar instruments.

As graphene reached peak hoopla, it also hit some snags. CVD graphene is actually a jigsaw of graphene monolayer crystals, and companies soon found that the seams in between dramatically affect the material's properties, usually for the worse. "I think they all expected they were going to get the same performance as the tiny little Scotch tape flakes of graphene that people published in research articles," Conway says. It's also notoriously difficult to displace incumbent materials in multibillion-dollar industries. (Spoiler alert: Most touchscreens are still made from that reliable workhorse, indium tin oxide.)
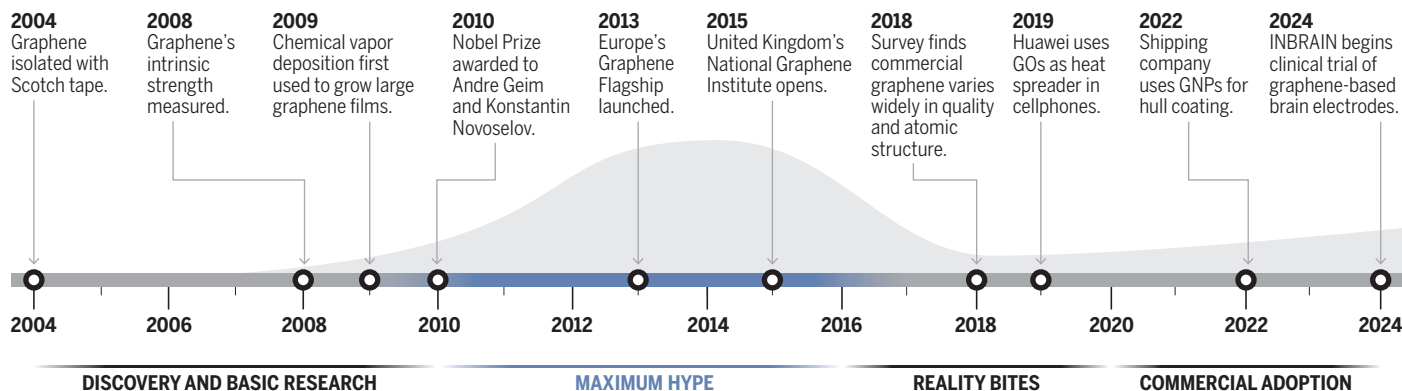
And growing graphene is only the first step—getting it off the copper foil and into a device can be a huge headache. One common method uses a polymer film to transfer

known as liquid exfoliation, involves shredding it into tiny pieces. (A kitchen blender will do.) Another uses strong oxidants to chew graphite into thinner fragments.

But this graphene is very different from Geim's gossamer flakes. It often contains multiple graphene layers stuck together, a form known as graphene nanoplatelets (GNPs). Other forms, called graphene oxides (GOs), are peppered with oxygen atoms introduced by the chemicals used to break down graphite. Further chemical treatment to remove the oxygen atoms produces reduced graphene oxides (rGOs), which are similar to graphene but have more defects in their hexagonal structure. "All these things are basically different materials," Bøggild says. As graphene layers stack up, their superpowers decline—and some properties of 10-layer stacks become indistinguishable from those of graphite.

A roller coaster ride

Graphene's discovery 20 years ago led to a wave of hype that saw billions of dollars in R&D funding go to universities and startups attempting to commercialize the atom-thin layers of carbon. More recently, markets for the so-called wonder material seem to be growing, not just for graphene monolayers, but also for graphene oxides (GOs) and graphene nanoplatelets (GNPs).

2004 Graphene isolated with Scotch tape.

2008 Graphene's intrinsic strength measured.

2009 Chemical vapor deposition first used to grow large graphene films.

2010 Nobel Prize awarded to Andre Geim and Konstantin Novoselov.

2013 Europe's Graphene Flagship launched.

2015 United Kingdom's National Graphene Institute opens.

2018 Survey finds commercial graphene varies widely in quality and atomic structure.

2019 Huawei uses GOs as heat spreader in cellphones.

2022 Shipping company uses GNPs for hull coating.

2024 INBRAIN begins clinical trial of graphene-based brain electrodes.

2004 | 2006 | 2008 | 2010 | 2012 | 2014 | 2016 | 2018 | 2020 | 2022 | 2024

DISCOVERY AND BASIC RESEARCH | **MAXIMUM HYPE** | **REALITY BITES** | **COMMERCIAL ADOPTION**

parency would make it an ideal material for the touchscreens of electronic devices, for example.

In the same year, the European Union launched the Graphene Flagship, a 10-year, €1.4 billion research program to put the bloc at the forefront of graphene commercialization.

The United Kingdom doubled down on graphene when it opened the £61 million NGI in 2015. The institute was such a source of national pride that when Chinese President Xi Jinping visited the U.K. that year, his itinerary included a tour of NGI, hosted by Geim, Novoselov, and other dignitaries. A photograph of Xi's visit still hangs in the exact spot where the image was snapped, right outside the facility's enormous clean room. Windows allow passersby on the street to peer inside at scientists

it, but this can cause tears and wrinkles, and it leaves residues that act like kryptonite on graphene's superpowers. In those early days, the quality of CVD graphene varied wildly from batch to batch, making it impossible to get consistent results.

Bøggild says surprisingly few of the reported procedures for handling, transferring, and applying CVD graphene can be reliably repeated. He attributes this "reproducibility gap" to the lack of detail about these procedures in many academic papers, and is now working with a consortium of researchers, companies, funders, and publishers to develop more rigorous standards for verifying such results.

The challenges posed by graphene built up atom by atom by CVD prompted some researchers to take a top-down approach, starting with bulk graphite. One method,

Yet these materials are still useful. Whereas CVD-grown graphene is prized as an electrical conductor, GNPs can lend strength to other materials, and GO films can draw heat from sensitive electronics (see graphic, p. 142). The challenge is to find the best kind of graphene for a particular application.

THE UNIVERSITY OF MANCHESTER'S answer to that question lies a short walk from NGI: the £60 million Graphene Engineering Innovation Centre (GEIC), an industrial incubator that opened in 2018. Its stern grayscale décor and large pilot-scale production hall, complete with overhead crane to move heavy equipment around, signal that this place means business. Sitting at a heavy boardroom table in a meeting room with a glossy, corporate sheen, James Baker—CEO

of Graphene@Manchester, which oversees the university's business-related graphene activities—explains how GEIC works.

Companies pay GEIC to rent laboratory space and work with the center's in-house experts. "Our business model is quite simple: You pay to come here, and if you succeed in exploiting the project developed at the GEIC, we take a royalty based on net sales," Baker says. That income is reinvested in the facility, and further research. Sixty startups have worked at NGI and GEIC over the past decade, and GEIC currently hosts roughly 230 people from about 30 partner companies.

Canadian company HydroGraph Clean Power opened a lab here in January, in part to develop applications for GNPs that it makes by exploding acetylene and oxygen in a steel chamber. Chief Science Officer Ranjith Divigalpitiya says the United States has been relatively slow to commercialize graphene applications, compared with Europe and Asia. "The biggest challenge is connecting the technology with the customer," he says. "And that's where the GEIC comes in, because they have a lot of credibility, and better connectivity to customers. I think it has just opened so many doors for us."

GEIC also helps companies reliably produce graphene and blend it with other substances, giving them tools and expertise to precisely characterize their materials. "What industry really wants is reproducibility and confidence in the data," Baker says. "Getting that consistency is key."

That's a big improvement on the Wild West environment of the early graphene market, when companies often sold "graphene" without being clear about the material's true identity. In 2017, the International Organization for Standardization decided the word graphene should only refer to monolayers, whereas "few-layer graphene" could contain three to 10 layers. The thinnest GNPs overlap with this second category, although many GNPs are thicker and contain dozens of layers. But many companies did not take the hint. In 2018, a survey of 60 companies found a large variation in the size of flakes and number of layers in materials sold simply as "graphene," which was holding back the development of applications.

Graphene's split personality also sowed confusion. Baker says some of the companies working at GEIC initially expected to get monolayer properties from GNPs, only to be disappointed when they were not 200 times stronger than steel (see sidebar, p. 143). Even today, it's all too common for companies or researchers working with GNPs to proclaim the amazing properties of monolayer graphene in their papers or websites. "That bugs me, it's overselling," Bøggild says. "I think it's bad because it's taken seri-

Slurries of graphene and polymer binders could help boost the performance of batteries (top). The Graphene Engineering Innovation Centre hosts industrial partners commercializing graphene products (bottom).

ously by people who don't understand it."

Despite these tribulations, some of these graphene-ish materials are showing promise. For example, Spain's INBRAIN Neuroelectronics makes biocompatible rGO electrode arrays to measure brain activity, which could help clinicians distinguish healthy and diseased tissue when removing a tumor, for example. The rGO films have tiny pores and a high conductivity, which makes them very responsive to tiny electrical changes. That means these flexible implants, thinner than a human hair, can be smaller than conventional devices and offer a much sharper map of neural activity, says company co-founder Jose Garrido of the Catalan Institute of Nanoscience and Nanotechnology. INBRAIN is now running a clinical trial of the device in Manchester, with the first patient receiving an implant in late September.

The company has also won "breakthrough device" designation from the U.S. Food & Drug Administration for similar rGO electrode arrays that can stimulate the brain as well as recording it, to alleviate symptoms of Parkinson's disease. Garrido hopes the designation will speed the device into clinical trials in the next 4 years.

One of the biggest producers of GOs and rGOs is the China-based Sixth Element (Changzhou) Materials Technology Co., which has a capacity of about 1000 tons per year. In 2019, electronics company Huawei started to use Sixth Element's GO to draw heat from the circuits in its smartphones, enabling them to be thinner. Other brands followed suit, and Bernhard Münzing, the company's European sales director, estimates that its GO is now inside millions of electronic devices. "I think it's the largest ap-

plication [of a graphene material] you can find on the market at the moment," he says.

Other bulk applications are on the rise. Canada's GIT Coatings uses GNPs to make a tough coating that keeps barnacles and other residue off ship hulls. It avoids some of the toxic compounds in conventional coatings and reduces friction to save on fuel. Finnish shipping company Finnlines started using the hull coating in 2022, and several other ship operators have joined in since then.

One early GEIC project with the automaker Ford led to an injection-molded foam engine component that included GNPs to improve heat protection. The GNPs also act as nucleating agents to create little air bubbles in the foam that reduce engine noise. "That's already gone into production in over 5 million vehicles in the U.S.," says John Whittaker, engineering director at GEIC. Another product, which an undisclosed GEIC partner company is due to launch soon, blends an elastic polymer with 0.1% GNPs to produce a strong, lightweight, and stretchy material suitable for personal protective equipment.

Meanwhile, some startups are exploiting graphene's inert and impermeable honeycomb in anticorrosion coatings and air-tight packaging. Others are puncturing their graphene, or stacking flakes to leave gaps between layers, to create porous membranes for separating gases or desalinating seawater.

There are also high hopes that mixing graphene materials into concrete could re-

INBRAIN Neuroelectronics's implants contain reduced graphene oxides (top). Graphenea uses monolayer graphene in its field effect transistors (bottom).

duce the amount of cement needed to make it. Cement production accounts for up to 8% of global carbon dioxide emissions. "I think the largest market, long term, will be the concrete market," Münzing says. For example, GEIC partner company Concretene can reduce cement content by 30% with a mixture that includes just 0.1% GNPs and GO nanoplatelets. "It allows us to take cement out of the concrete with-

out any impact on strength," says Craig Dawson, Concretene's chief scientific officer. Three trial construction projects already used the additive, and the company hopes precast concrete parts with it will be commercially available next year. "We believe we are the first to do this, and we are well ahead of the game," Dawson says.
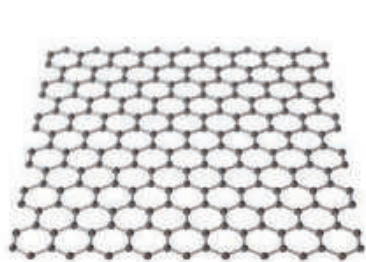
IDTECHEX ESTIMATES the market for all types of graphene is currently worth about $150 million per year, mostly from bulk graphene applications, and could rise to $1.6 billion by 2034. But the company reckons monolayer graphene is poised to grow its share of that market, because some companies are finally overcoming the problems of transferring CVD-grown graphene.

Spain's Graphenea, for example, can now use polymer films to transfer CVD graphene from its copper foil to 200-millimeter-diameter wafers of silicon or other substrates, large enough to use in full-scale semiconductor fabrication lines. "I think transfer is not the main bottleneck anymore," says Graphenea's Chief Scientific Officer Amaia Zurutuza. Engineers are also learning exactly how good CVD graphene needs to be for specific applications.
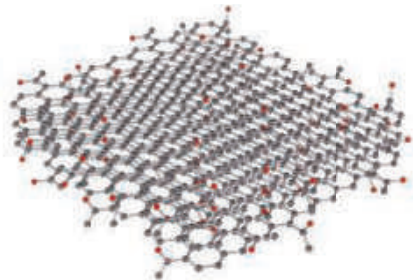
Graphene biosensors, for instance, are often based on field effect transistors (FETs), which can sense tiny changes in electric fields. In a typical FET biosensor, the graphene is decorated with receptor molecules that bind to a target biomolecule, and this
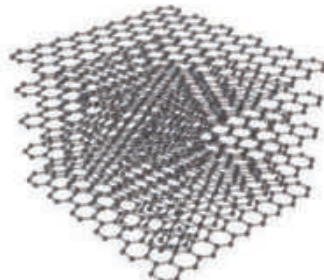
How graphene stacks up

Graphene products may incorporate monolayers, but they might also use graphene oxides or graphene nanoplatelets, which have radically different properties.

	Monolayer graphene	Graphene oxides	Graphene nanoplatelets
Layers	1	1–10	5–10 (or more)
Applications	Magnetic/biochemical sensors; electronics; photonics	Heat management; water treatment; filtration membranes; composites	Strong clothing, composites, and concrete; anticorrosion coatings
Production	Scotch tape; chemical vapor deposition	Chemical oxidation of graphite; controlled gas combustion	Mechanical or chemical breakdown of graphite; controlled gas combustion
Electrical conductivity	High (2x copper)	Low—effectively an insulator	Good, depends on number of layers
Other properties	High thermal conductivity (10x copper); high strength (>100x atom-thin steel)	Mixes well with water; easy to chemically modify	High strength and toughness

interaction changes the electrical current flowing through the graphene. For this application, the graphene's electron mobility needs to be impressive—but nowhere near as high as that of a Scotch taped flake.

INBRAIN is working on graphene FETs that would be even more sensitive than its rGO electrode arrays, capable of detecting the electric activity of individual neurons firing inside the brain. Experiments in mice show that, unlike electrode arrays, the graphene FETs can record low frequency changes in electrical activity that are related to epilepsy.

Garrido says these developments would not have been possible without the long-term investment provided by the EU's Graphene Flagship program, not least because the company needed to spend years testing and optimizing hundreds of the devices in animal models. "This continuity in the funding has been key for us, definitely," Garrido says. "I think that's why Europe, at least from the point of view of graphene medical devices, has done so much more compared to the U.S."

Monolayer graphene is also being deployed in optoelectronic devices, which rapidly convert light into electrical current and vice versa. Germany's Black Semiconductor, a Graphene Flagship partner, is adding graphene to silicon microchips to speed up optical communication between chips. Ultimately that means faster data processing, and in June the company secured more than €250 million in public and private funding to develop its devices.

But CVD growth and transfer will likely need to improve further before major chipmakers adopt the material. "For the semiconductor industry, you really need to go to the next level," says Inge Asselberghs, director of the Graphene Flagship's €20 million 2D Experimental Pilot Line. The project is teaming with Graphenea and other partners to transfer high-quality CVD graphene using rigid glass substrates. That causes less damage than flexible polymers, although researchers still inspect the material under a microscope to weed out any bad bits.

Meanwhile, Bøggild co-founded a Danish Graphene Flagship spinoff called 2D that is taking a different approach to the transfer problem. His research team realized it could cleanly transfer monolayer graphene from copper foil to a polymer sheet using nothing more than an office laminator. This polymer can then be applied to another surface, and peeled off to leave the graphene behind, with less contamination than other methods. 2D has now commercialized that method in a roll-to-roll process that produces meter-long stretches of polymer-backed single-layer graphene.

Strong stuff

Press releases rarely miss an opportunity to trumpet graphene's strength. But where does the claim that it is "200 times stronger than steel" come from? Back in 2008, a team led by James Hone at Columbia University came up with a clever way to measure graphene's strength. Researchers draped monolayer graphene flakes extracted with Scotch tape over tiny holes in silicon, and then pressed down with the tip of an atomic force microscope until the graphene ruptured.

This was the first time anyone had directly measured the "intrinsic strength" of a material—the strength that comes from a flawless array of chemical bonds, unmarred by defects or cracks. The measurement took advantage of graphene's atomic perfection. But what did the strength of a 2D sheet mean for a 3D object made of graphene? "In any bulk material, the actual practical breaking strength is much lower than the intrinsic strength," Hone says.

To get an idea, the researchers imagined stacking billions of these perfect graphene sheets together to make a beam 1 meter high, and multiplied their intrinsic strength of 42 newtons per meter accordingly to give a 3D answer of roughly 130 gigapascals. That's more than 200 times as strong as structural steel.

Hone acknowledges that creating this theoretical block of graphite from perfectly aligned graphene sheets would be impossible. "It's really just a way to compare what we see in this two-dimensional world to the three-dimensional world," he says. –*M.P.*

ONE U.K. COMPANY has found a way to circumvent the transfer step entirely. Paragraf instead uses a semiconductor manufacturing method called metal-organic chemical vapor deposition (MOCVD) to grow graphene directly where it's needed on devices. "That's what sets us apart," Conway says. "We can grow directly on the substrates. That makes a huge difference, because you just get much better reproducibility."

The company claims it is the first to use standard semiconductor processes to mass-produce graphene-based electronic devices, including magnetic sensors and biosensor FETs, and is producing thousands of devices per year for customers. Its magnetic sensor exploits graphene's high sensitivity to the Hall effect, "so you can measure very small magnetic fields much more accurately," Conway says. Automakers are now testing the sensors to monitor the health of electric vehicle batteries.

Over the next decade, Paragraf plans to use MOCVD to deposit other 2D materials, such as molybdenum disulfide, into devices alongside graphene. Creating stacks of different 2D materials is now a major basic research frontier because each material can bring specific properties and modify those of its neighbors, allowing researchers to fine-tune how the whole stack behaves. For example, sandwiching graphene between two layers of an insulating 2D material called hexagonal boron nitride can help preserve graphene's superlative electronic properties within devices.

Solid state physicist Roman Gorbachev is already building these kinds of graphene club sandwiches at NGI, using a homemade production line. It consists of a 5-meter-long gleaming steel pipe, bristling with joints, side arms, and analytical instruments, that encloses an ultraclean, high-vacuum environment. Inside one part of the system, a robot assembles stacks of 2D materials; a small train then ferries these stacks into other chambers where they are connected to electrical contacts and studied. "Building this machine absorbed roughly 5 years of my life," Gorbachev says. "At the moment, this is one of a kind."

Recently, Gorbachev has used his machine to study how twisting the sandwich layers can dramatically alter their electronic properties. Other researchers in the burgeoning field of "twistronics" have found that in two-layer graphene, twisting one layer by 1.1° transforms the material into a superconductor, for example.

Gorbachev says his modest assembly line is an important step along the long road to commercializing the 2D layered structures. "If we're talking about these really high-end applications, we will need more time," he says. "But they are happening, they are coming."

Geim agrees that this wider world of 2D materials might eventually have a greater impact in electronic devices than graphene itself. "In terms of applications, I would spread my bets on many different materials," he says. "But graphene was certainly an adventure, and I didn't expect that it would last 20 years, for sure." ∎

Mark Peplow is a science journalist in Penrith, England.

LANDMARK: GRAPHENE

Graphene, beyond lab benches

Twenty years since its discovery, the journey to reach graphene's true potential is still underway

By **Yixuan Zhao**[1,2,3] and **Li Lin**[1,2]

Twenty years after the discovery of graphene—a one-atom-thick layer of carbon atoms in a honeycomb structure (*1*)—research on this remarkable material has evolved from benchtops to large-scale productions and commercial applications. Graphene is known for its breadth of exotic properties, including electrical conductivity up to 70% higher than that of copper, thermal conductivity at room temperature that is more than twice that of pyrolytic graphite, and strength that is 200 times as great as that of steel (at one-sixth the weight). These attributes make graphene and its derivatives ideal for a myriad of uses, including electronic, structural, and biomedical uses. However, even broader industrial applications of graphene are yet to be realized.

As described in the 2004 pioneering paper by Novoselov *et al*. (*1*), a highly oriented pyrolytic graphite is mechanically exfoliated to produce monolayer graphene flake using an adhesive tape. The new material exhibits intriguing properties. The authors observed a strong ambipolar electric field effect—an ability to switch between positive and negative charges upon an applied gate voltage. They also noted high room temperature carrier mobilities, which are associated with how fast an electron or a hole moves through a material. These properties make graphene promising for electronic applications. Although the mechanical peeling process produced pristine graphene, it only made small flakes with sizes ranging from a few tens to hundreds of micrometers. Thus, early graphene research focused on synthesizing high-quality monolayer or multilayers of graphene in a continuous form, primarily by chemical vapor deposition on metal substrates using carbon-containing precursors. However, limited scale of production, a large variation in the quality across different batches, and a lack of reliable methods to transfer graphene from one substrate to another have slowed down the application of high-quality

[1]School of Materials Science and Engineering, Peking University, Beijing, P. R. China. [2]Beijing Graphene Institute, Beijing, P. R. China. [3]Center for Nanochemistry, Beijing National Laboratory for Molecular Science, College of Chemistry and Molecular Engineering, Peking University, Beijing, P. R. China. Email: linli-cnc@pku.edu.cn

graphene films for electronic devices.

The initial focus of graphene research and commercialization centered on its derivatives, such as graphene nanoplatelets (two-dimensional particles made of stacked sheets), graphene oxides with functional oxygen groups, and reduced graphene oxides in which the oxygen groups are removed (see the figure). These are relatively easier to produce in large quantities than the pristine counterpart, which is produced primarily by exfoliating graphite mechanically by an adhesive tape, but produce a large amount of defects within the product (2). Thus, potential applications of graphene have switched toward those that can tolerate structural imperfections but still benefit from the strength and light weight of graphene derivatives, such as composites, coatings, and reinforcements. For example, graphene nanoplatelets have been used in anticorrosion coating, fire retardants, and electromagnetic shielding materials, which are all close to commercialization (3). Additionally, the sufficiently high conductivity, electrochemical stability, and porous structure give graphene derivatives the potential to be used for energy storage and conversion devices.

By contrast, industry application of high-quality, continuous graphene is still in the rudimentary stage. Electronic and optoelectronic devices require the synthesis of large-scale uniform monolayer or multilayer graphene films with sufficiently high carrier mobilities (usually above 10,000 cm^2/Vs). Laboratory- and industry-scale efforts have been devoted to optimizing the chemical vapor deposition–based growth process of graphene to reduce the overall defect density as well as the number of grain boundaries and wrinkles, which impede electrical transport. In addition, side reactions occur on the graphene surface during a large-scale production, such as the formation of amorphous carbons, which alter the intrinsic properties of graphene. Metal substrates, such as copper foils, have been widely used for growing large-area monolayer graphene films. A surface-mediated growth mechanism, in which a further growth of additional layers is hindered by the reduced catalytic activity of the substrates with graphene coverage (4), enables fine control over the film thickness. Currently, wafer-scale growth of single-crystalline graphene film with reduced density of grain boundaries has only been achieved on metals (5). Additionally, depositing graphene films on nonmetal substrates, such as semiconductors and insulators, is desired for a broader scope of electronic and photonic devices. Chemical vapor deposition growth of graphene on insulating sapphire substrates has been demonstrated recently (6), moving graphene a step closer to commercial electronic and optoelectronic device application. The ca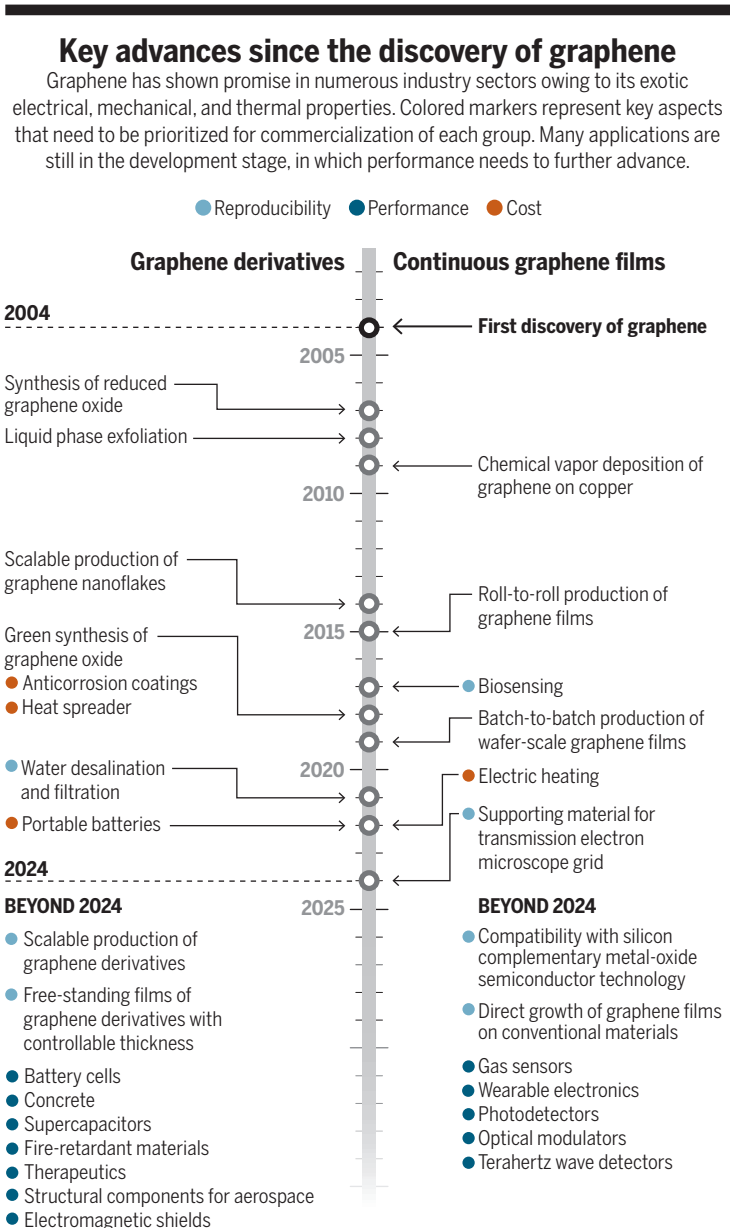rrier mobility of chemical vapor deposition–produced graphene of more than 1 million cm^2/Vs suggests that the gap between laboratory-grown and pristine graphene no longer exists (7). However, the reported laboratory-scale devices are based on a carefully selected area within a large graphene-coated substrate and thus suffer from considerable device-to-device variations.

From the perspective of electronic applications, the compatibilities of chemical vapor deposition graphene with the existing microfabrication facilities and techniques for silicon-based complementary metal-oxide semiconductor technologies is critical. Unfortunately, conformal growth of graphene on silicon/silicon dioxide substrates is difficult owing to the formation of impurities, such as carbides, during the growth process (8). In addition, amorphous silicon substrates cannot decompose carbon-containing precursors, resulting in high defect density of graphene films. Catalytic capacity has been improved by introducing gas-phase metal catalysts (8), but optimizing the growth parameter for large-scale production of graphene on silicon substrates is still challenging.

One solution to circumvent the challenges with silicon substrates is to grow graphene on a separate substrate and then transfer it onto a silicon surface. However, this approach often produces cracks, wrinkles, folding of graphene sheets, and other defects, which create large device-to-device variations and affect the production yield. These imperfections also alter integration of dielectrics onto graphene. Because dielectrics are electrical charge–storing materials with high electrical resistivity, they dictate the electronic performances and working stabilities of graphene-based electronics. Poor graphene-dielectric interfaces that are contaminated by transfer-related residues and pinholes (a type of defect in electronics) deteriorate the carrier mobility of graphene films. With device miniaturization, even a small defect can severely deteriorate performance.

Many strategies have been developed to suppress the transfer-induced imperfections and to improve dielec-

Key advances since the discovery of graphene

Graphene has shown promise in numerous industry sectors owing to its exotic electrical, mechanical, and thermal properties. Colored markers represent key aspects that need to be prioritized for commercialization of each group. Many applications are still in the development stage, in which performance needs to further advance.

● Reproducibility ● Performance ● Cost

Graphene derivatives | **Continuous graphene films**

2004 — First discovery of graphene

2005
- Synthesis of reduced graphene oxide
- Liquid phase exfoliation
- Chemical vapor deposition of graphene on copper

2010
- Scalable production of graphene nanoflakes
- Roll-to-roll production of graphene films

2015
- Green synthesis of graphene oxide
- ● Anticorrosion coatings
- ● Heat spreader
- ● Biosensing
- Batch-to-batch production of wafer-scale graphene films

2020
- ● Water desalination and filtration
- ● Electric heating
- ● Portable batteries
- ● Supporting material for transmission electron microscope grid

2024

BEYOND 2024
2025
- ● Scalable production of graphene derivatives
- ● Free-standing films of graphene derivatives with controllable thickness
- ● Battery cells
- ● Concrete
- ● Supercapacitors
- ● Fire-retardant materials
- ● Therapeutics
- ● Structural components for aerospace
- ● Electromagnetic shields

BEYOND 2024
- ● Compatibility with silicon complementary metal-oxide semiconductor technology
- ● Direct growth of graphene films on conventional materials
- ● Gas sensors
- ● Wearable electronics
- ● Photodetectors
- ● Optical modulators
- ● Terahertz wave detectors

GRAPHIC: N. CARY/SCIENCE

tric integration. Controlling delamination and lamination forces and avoiding implanting trace impurities on the surface have helped in mitigating the defects (9). Current silicon technologies also give hints as to how to create these processes. For example, a conventional wafer bonding process has been adopted to directly bond graphene grown on copper substrates onto a desired substrate with an adhesive layer between them followed by the removal of the copper layer by chemical etching (10). Another approach facilitates the transfer of graphene without cracks by controlling the adhesion between tape and graphene by ultraviolet light illumination (11). Moreover, van der Waals layered dielectric materials, such as hexagonal boron nitride (hBN) and bismuth selenate (Bi_2SeO_5) (12), can be used as supporting layers to assist with transferring and to avoid interfacial contaminations.

Nevertheless, these techniques have only been tested at laboratory scale, and careful testing at industry scale is needed. To avoid heavy investments in building new processes and facilities to mass-produce graphene products, it is essential to develop procedures that are compatible with the current manufacturing processes. Further, transfer of large-area graphene films requires repeatability at production capacity. Integrating automatic transfer processes, using robotics and machine designs, into the existing semiconductor facilities can improve consistency. At the same time, developing methods to directly grow graphene films on conventional substrate materials bypasses transfer-related problems and can help accelerate commercialization (13). Thus, commercialization of graphene-based products primarily relies on collaboration between academia and industry.

A major challenge in commercializing graphene and its derivatives is the high production cost for the limited production volume. Raw materials, such as methane gas or raw graphite powders, only account for a small proportion of the overall cost. Rather, equipment designs for large-scale batch production and processing of synthesized graphene account for most of the initial investment. Thus, the industry is driven toward high-margin applications that require less investment for substantial performance improvement over existing products. This has led to faster commercialization of graphene derivatives for various structural components compared with high-quality graphene films for devices. For example, graphene in transparent conductive films has been hindered by its high production cost compared with the conventional indium tin oxide. Even the graphene derivatives that have already been commercialized or are close to commercial

applications need to be further developed to reduce cost. The current mainstream methods of producing graphene nanoplatelet films are filtration and layer-by-layer assembly using electrostatic interaction, which are slow and not ideal for mass production. Reducing costs without deteriorating its properties is crucial for graphene products to break into the market.

Variations in production methods across industry result in a substantial difference in the properties and performance of graphene products. This highlights the need for developing standards for graphene and its derivatives. With this, the development of high-throughput characterization is crucial for measuring the properties to sort out those that do not meet the standards. Optical characterization methods, such as ellipsometry contrast microscopy and terahertz time-domain spectroscopy, can inspect large-area graphene rapidly (14). However, some detrimental structural defects can only be examined using advanced microscopy, which are generally time-consuming. Automatic characterization combined with machine learning and artificial intelligence can help overcome the limitation. For instance, automated imaging and data analytics can find features that represent defects with minimal human contribution (15).

There is still a substantial gap between academia and demands from the industry. Closer collaboration to reduce this gap is required. Particularly, a comprehensive system of standards should include application-related grades for guiding the products and their entry into specific markets, such as specific properties for specific functions and corresponding measurement methodologies. These standards should be drafted through collaborations between researchers, manufacturing companies, and application enterprises. ∎

REFERENCES AND NOTES

1. K. S. Novoselov *et al.*, *Science* **306**, 666 (2004).
2. V. Nicolosi, *et al.*, *Science* **340**, 1226419 (2013).
3. B. Wicklein *et al.*, *Nat. Nanotechnol.* **10**, 277 (2015).
4. X. Li *et al.*, *Science* **324**, 1312 (2009).
5. G. Yuan *et al.*, *Nature* **577**, 204 (2020).
6. J. Li *et al.*, *Nat. Mater.* **21**, 740 (2022).
7. Y. Zhao *et al.*, *Nat. Commun.* **13**, 4409 (2022).
8. P.-Y. Teng *et al.*, *Nano Lett.* **12**, 1379 (2012).
9. N. Hong *et al.*, *Adv. Mater.* **34**, 2106615 (2022).
10. A. Quellmalz *et al.*, *Nat. Commun.* **12**, 917 (2021).
11. M. Nakatani *et al.*, *Nat. Electron.* **7**, 119 (2024).
12. C. Zhang *et al.*, *Nat. Mater.* **22**, 832 (2023).
13. K. Chen *et al.*, *Nat. Photonics* **13**, 754 (2019).
14. P. Bøggild *et al.*, *2D Mater.* **4**, 042003 (2017).
15. M. Ziatdinov *et al.*, *Sci. Adv.* **5**, eaaw8989 (2019).

Genomic patterns in the schizophrenia brain

Somatic noncoding mutations likely contribute to schizophrenia development

By **Joon-Yong An**[1,2,3] *and* **Yujin Kim**[2,3]

Human germline mutations—genetic changes in an egg or sperm—are inherited, present in every cell, and play roles in early development (1), whereas somatic mutations arise postzygotically and may or may not affect developmental trajectories. Neurons accumulate hundreds to thousands of somatic mutations throughout development, with distinct mutational processes and rates occurring at various developmental stages (2). Whole-genome sequencing studies characterizing somatic mosaicism in early human brain development indicate that the mutation rate is relatively low during early pregastrulation (the first 2 weeks after fertilization) but increases substantially during late neurogenesis in the prenatal period (starting at 22 weeks of gestation), primarily owing to oxidative damage (3). On page 217 of this issue, Maury *et al.* (4) report that somatic mutations in the brains of individuals with schizophrenia occurred during neogenesis. This suggests that intrauterine factors might influence mutational mechanisms and brain development in utero.

Somatic mutations in the human brain are not evenly distributed across the genome but accumulate in regions implicated in neurodevelopmental and neuropsychiatric disorders (5–7). Protein-altering somatic mutations in the brains of individuals with schizophrenia or autism often occur in the same genes that are affected by

[1]School of Biosystem and Biomedical Science, College of Health Science, Korea University, Seoul, Republic of Korea. [2]Department of Integrated Biomedical and Life Science, Korea University, Seoul, Republic of Korea. [3]L-HOPE Program for Community-Based Total Learning Health Systems, Korea University, Seoul, Republic of Korea. Email: joonan30@korea.ac.kr

germline mutations associated with these conditions. This overlap suggests that both types of mutations may contribute to similar neurological processes in these disorders. In addition, mutations in non-coding regions of the genome accumulate in neurons over time, especially in transcriptionally active regions, and contribute to disrupting neuronal function. For instance, brains of those with autism were enriched with noncoding mutations in regions that bind to the transcription factor MEIS (myeloid ectopic viral integration site 1) (5). Thus, the uneven distribution and concentration of somatic mutations suggest that certain genomic regions and developmental periods may be affected by these mutations, potentially leading to developmental and functional abnormalities.

Maury et al. investigated somatic mosaicism in 61 postmortem brains from individuals diagnosed with schizophrenia and in 25 control brains from healthy donors, focusing on the role of noncoding somatic mutations in schizophrenia. In contrast to germline variants associated with schizophrenia or other neurodevelopmental disorders, somatic mutations in schizophrenia brains did not show substantial enrichment for protein-truncating coding variants or copy-number variations. Instead, somatic single-nucleotide variants were enriched in gene promoter regions, particularly within transcription factor binding sites. These regions overlapped to a great extent with deoxyribonuclease (DNase) hypersensitivity sites in normal fetal brains, implicating the regulatory landscape of the developing brain as potentially contributing to neurodevelopmental abnormalities observed in schizophrenia.

Gene promoters and transcription factor binding sites have emerged as critical noncoding regions associated with risk for developing neurodevelopmental conditions. Large-scale whole-genome sequencing studies have consistently indicated an association between de novo germline mutations in promoter regions, particularly in actively transcribed genomic regions during early neurodevelopment, and autism risk (8, 9). This recognition extends to schizophrenia, as recently reported in a study that used massively parallel reporter assays to analyze loci in the genome that are associated with schizophrenia and other neuropsychiatric disorders. Risk loci within regions containing transcription factor binding sites that are important for neurodevelopment exhibited the largest alterations in functional activity (10). These observations and the findings of Maury et al. collectively point to the need to prioritize promoters and transcription factor binding sites when

investigating the association between noncoding regions of the genome in neurodevelopmental conditions.

Maury et al. propose that distinct mutational signatures underlie schizophrenia brain somatic mosaicism, including CpG (C, cytosine; G, guanine) transversions and T-to-G (T, thymine) substitutions at active transcription factor binding sites (see the figure). CpG are dinucleotides that mutate at a high rate because cytosine is vulnerable to deamination. Indeed, CpG transversions to GpG increased 24.0-fold at active transcription factor binding sites located near promoters in schizophrenia brains compared with the expected C-to-G genome-wide rate of 0.083. In addition, T-to-G substitutions increased 82.6-fold in transcription factor binding sites near promoter regions in these brains compared with the expected rate. CpG transversions at tran-

scription factor binding sites might disrupt gene regulation during crucial stages of brain development. The effects of these mutations were highly localized, occurring less frequently beyond 100 base pairs from the midpoint of the transcription factor binding site. Maury et al. also found recurrent T-to-G mutations at the same locus from three unrelated schizophrenia brains. Given that this pattern was not observed in autism brains, the strong enrichment of T-to-G substitutions at transcription factor binding sites might be a common signature in schizophrenia risk development. Functional characterization using massively parallel reporter assays showed marginal differences in the effects of somatic single-nucleotide variants observed in schizophrenia brains compared with those in control brains. Despite this, some mutations led to down-regulated gene expression in neurodevelopment-related genes, such as *NFIX* (*nuclear factor I X*) or *ELAVL3* (*ELAV-like RNA binding protein 3*), or disrupted regulatory elements that target schizophrenia risk genes identified in genome-wide association studies. This highlights not only the nonrandom characteristics of somatic single-nucleotide variants in schizophrenia brains regarding mutation signatures but also their functional convergence in schizophrenia etiology.

Genomic patterns that underlie brain mosaicism hold potential to dissect the complex genetic architecture of schizophrenia. Such mutational signatures may further help to distinguish between subtypes of schizophrenia, thereby providing insights into the phenotypic heterogeneity observed in patients. Just as mutational sig-

Mutational hotspots in the schizophrenia brain

The frequency of somatic mutations in neurons of schizophrenia brains indicates a preference for specific genomic regions characterized by specific patterns. Hotspots are depicted as including transcription factor binding sites, particularly near promoters, and a distinct pattern of base substitutions.

A, adenine; bp, base pairs; C, cytosine; G, guanine; T, thymine.

natures have led to the identification of a new subtype and therapeutic intervention in various cancers, similar approaches could be applied to schizophrenia. Recognizing distinct mutational patterns within somatic mosaicism in the brain could allow for precise stratification of schizophrenia subtypes, each potentially linked to different developmental pathways or environmental interactions. Incorporating brain somatic mosaicism with polygenic risk scores (which reflect the likelihood of a certain genetic trait manifesting as a disorder) may enhance individual risk profile prediction and contribute to identifying patients who are likely to benefit from specific therapeutic interventions. This, in turn, could explain the wide range of symptoms and severities observed among individuals

with schizophrenia and offer a personalized approach to diagnosis and treatment.

As understanding of somatic mutation signatures advances, the necessity for sophisticated tools to decode the complex genomic features that underlie these mutations becomes increasingly apparent. Deep-learning models present powerful tools for analyzing genomic features associated with disorders (11) or the sequence lexicon that underlies gene promoters and transcription initiation (12). These models can identify patterns and features in DNA sequences that may be associated with somatic mutations, enabling researchers to unravel subtle genomic and regulatory signatures that are not easily detected using traditional methods. Large language models can classify disease conditions, such as tumor types, because the models recognize key mutations that frequently appear in tumor samples (13). Similarly, by leveraging mutational hotspots identified in Maury *et al.*, these models could classify a genomic subtype of schizophrenia and prioritize meaningful features that underlie schizophrenia etiology.

Fine-tuning large language models should enable the detection of sequence patterns in regulatory regions of the genome where somatic mutations associated with schizophrenia frequently occur. Considering the cis-regulatory patterns observed in schizophrenia somatic single-nucleotide variants, integrating diverse data, such as epigenetic marks and chromatin accessibility, may provide a comprehensive mechanistic understanding of how somatic mutations arise during early neurogenesis. Thus, this knowledge may enhance the ability to predict the functional consequences of somatic mutations, refine understanding of the genetic architecture of schizophrenia, and potentially identify new therapeutic targets. ■

REFERENCES AND NOTES

1. D. M. Werling *et al.*, *Cell Rep.* **31**, 107489 (2020).
2. S. Bizzotto, C. A. Walsh, *Nat. Rev. Neurosci.* **23**, 275 (2022).
3. T. Bae *et al.*, *Science* **359**, 550 (2018).
4. E. A. Maury *et al.*, *Science* **386**, 217 (2024).
5. T. Bae *et al.*, *Science* **377**, 511 (2022).
6. C. Chung *et al.*, *Nat. Genet.* **55**, 209 (2023).
7. H. Y. Koh *et al.*, *Ann. Neurol.* **90**, 285 (2021).
8. J.-Y. An *et al.*, *Science* **362**, eaat6576 (2018).
9. T. N. Turner *et al.*, *Cell* **171**, 710 (2017).
10. C. Deng *et al.*, *Science* **384**, eadh0559 (2024).
11. J. Zhou *et al.*, *Nat. Genet.* **51**, 973 (2019).
12. K. Dudnyk, D. Cai, C. Shi, J. Xu, J. Zhou, *Science* **384**, eadj0116 (2024).
13. J. Anaya, J. W. Sidhom, F. Mahmood, A. S. Baras, *Nat. Biomed. Eng.* **8**, 57 (2024).

Unwrapping stemness to revive T cells

Targeting epigenetic regulators prevents T cell exhaustion

By **Carlson Tsui** and **Axel Kallies**

T cell exhaustion, characterized by the loss of cytokine production and sustained expression of inhibitory receptors or immune checkpoints such as programmed cell death 1 (PD-1), poses a major hurdle in cancer immunotherapies, including immune checkpoint blockade (ICB) and adoptive cell therapies (1, 2). Prolonged CD8$^+$ T cell responses, such as those found in cancer and chronic infection, are maintained by so-called precursors of exhausted T (T$_{PEX}$) cells, which exhibit stemlike features, allowing them to undergo self-renewal while also giving rise to effector-like exhausted T (T$_{EX}$) cell progeny (3, 4). The balance of stemness and effector differentiation determines the quality of the T cell response and makes T$_{PEX}$ cells prime targets for therapeutic manipulation (2). On page 165 of this issue, Kang *et al.* (5) report that deleting key epigenetic regulators can preserve ICB-responsive T$_{PEX}$ cells during chronic antigen exposure, revealing potential targets to overcome T cell exhaustion and achieve durable therapeutic response to ICB.

Recent studies have provided insights into the role of epigenetic regulators in enforcing the exhaustion program, which limits T cell function in chronic diseases and their capacity to be reinvigorated by ICB (6). Genome-wide analyses revealed extensive epigenetic reprogramming of chronically stimulated T cells compared with acutely stimulated counterparts, with thousands of differentially accessible regions (7) and context-specific methylation programs (8). Kang *et al.* investigated three epigenetic regulators, including DNA methyltransferase 3 alpha (*Dnmt3a*), a known driver of exhaustion (8); ten-eleven translocation 2 (*Tet2*), known to limit the durability of chimeric antigen receptor (CAR) T cells (9); and additional sex comb-like 1 (*Asxl1*), linked to the survival of a subset of myelodysplastic syndrome patients after ICB (10). Notably, all three genes are commonly mutated in clonal hematopoiesis (CH), a phenomenon driven by somatic mutations that confers inappropriate proliferative fitness to hematopoietic stem cells, resulting in a clonally biased expansion of its progeny (11, 12).

Using CRISPR-Cas9 technology, Kang *et al.* disrupted these CH factors in T cell receptor transgenic CD8$^+$ T cells specific for lymphocytic choriomeningitis virus (LCMV) and investigated their role in population expansion and maintenance during chronic LCMV infection. Disruption of each factor improved the long-term maintenance of CD8$^+$ T cells and in all cases was linked to enhanced T$_{PEX}$ cell maintenance and increased capacity for cytokine production. Notably, CD8$^+$ T cells deficient in these factors responded more robustly to ICB compared with controls. Disruption of CH factors altered the transcriptional landscape of T$_{EX}$ cell subsets, which was marked by increased expression of genes associated with T cell stemness. These included *Tcf7*, *Lef1*, and *Myb* [which encode the transcription factors T cell factor 1 (TCF1), lymphoid enhancer-binding factor 1 (LEF1), and myeloblastosis proto-oncogene protein (MYB), respectively] and genes associated with effector function such as *Ifng* [which encodes interferon gamma (IFN-γ)]. CH factor–deficient CD8$^+$ T cells remained numerically stable for up to a year without any signs of malignant transformation. Because oncogenic transformation of CAR T cells remains a major concern in adoptive cell therapies (13), these observations underscore the feasibility of targeting CH factors for therapeutic purposes.

Kang *et al.* focused their attention on *Asxl1* and analyzed chromatin accessibility, DNA methylation, and histone modifications in *Asxl1*-deficient and control T cells to define the molecular mechanisms of *Asxl1*-mediated regulation of stemness. Although they did not observe any major differences in DNA methylation, they found considerable differences in chromatin accessibility between *Asxl1*-sufficient and *Asxl1*-deficient T cells. These differences were prevalent in both T$_{PEX}$ cells and their T$_{EX}$ cell progeny, indicating that *Asxl1* affects differentiation and function of both major subsets of T$_{EX}$ cells. Further mechanistic analyses revealed notable differences in histone modifications. Specifically, Kang *et al.* showed that *Asxl1* was responsible for the deubiquitination of lysine resi-

Department of Microbiology and Immunology, The Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, Australia. Email: axel.kallies@unimelb.edu.au; carlson.tsui@unimelb.edu.au

Chromatin remodeling shapes stemness in T cells

In chronic infection and cancer, CD8+ T cell responses are maintained by precursors of exhausted T (T_{PEX}) cells, which exhibit stemlike features, allowing them to undergo self-renewal while also giving rise to effector-like exhausted T (T_{EX}) cell progeny. Disruption of additional sex comb-like 1 (ASXL1) leads to ubiquitination of histone 2A, which allows opening of chromatin regions in genes that promote T cell stemness and function, thus improving maintenance and response to immune checkpoint blockade.

Lef1, lymphoid enhancer-binding factor 1; *Myb*, myeloblastosis proto-oncogene; *Tcf7*, transcription factor 7 gene, encoding T cell factor 1; Ub, ubiquitin.
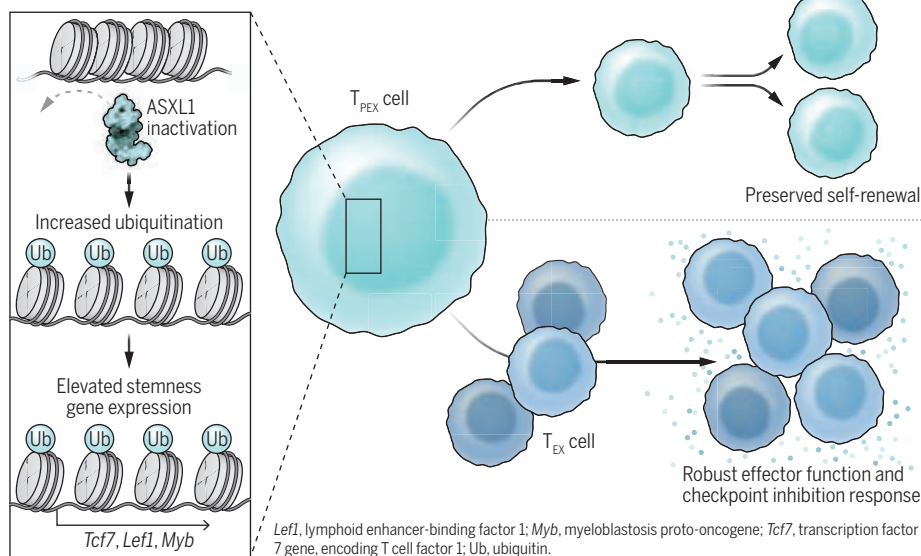
due 119 on histone 2A (H2AK119) through the polycomb group–repressive deubiquitinase. Accordingly, loss of *Asxl1* in T cells resulted in a region-specific increase in ubiquitinated H2A (H2AK119Ub) across the whole genome. H2AK119Ub-enriched regions, but not those co-enriched for trimethylated histone 3 (H3K27me3), were associated with accessible chromatin regions in genes such as *Tcf7* and *Lef1*. These epigenetic patterns also mirrored those found in human CD8+ T cells, where H2AK119Ub, controlled by *ASXL1*, was enriched in naïve cells compared with more differentiated effector memory T cells.

Using mouse models of cancer, Kang *et al*. showed that *Asxl1* deficiency in T cells enhanced T_{PEX} cell maintenance, TCF1 expression, and IFN-γ production and conferred better tumor control, both on its own and in synergy with ICB. Tumor-draining lymph nodes have recently been shown to play a role in the antitumor T cell response (*14*). The effects of ASXL1 deletion were partly independent from the input from the tumor-draining lymph nodes, which suggests that ASXL1-dependent epigenetic remodeling of stemness features in CD8+ T cells can counteract immunosuppressive cues within the tumor microenvironment. Consistent with the notion that *ASXL1* limits ICB responsiveness in cancers, the authors found that high *ASXL1* expression correlated with poorer survival for myelodysplastic syndrome cancer patients treated with ICB. Furthermore, ICB therapy resulted in a selective survival advantage to T cells bearing a disruptive *ASXL1* mutation.

Thus, Kang *et al*. not only identified *ASXL1* to be a potential biomarker and target to improve T cell stemness but also provide an explanation for some of the differences observed in cancer patients' responses to ICB.

Kang *et al*. define a previously unknown epigenetic regulatory pathway that controls T cell stemness, paving the way to new strategies that may improve antitumor therapy (see the figure). By drawing parallels between chronic T cell responses and aberrant hematopoiesis, the findings, together with a previous study (*15*), demonstrate the parsimonious nature of transcriptional and epigenetic regulators that underpin stemness and self-renewal in the hematopoietic system. This thus once more highlights the importance of cross-disciplinary biological research. ∎

REFERENCES AND NOTES

1. A. Chow *et al*., *Nat. Rev. Clin. Oncol*. **19**, 775 (2022).
2. M. Philip, A. Schietinger, *Nat. Rev. Immunol*. **22**, 209 (2022).
3. A. Kallies *et al*., *Nat. Rev. Immunol*. **20**, 128 (2020).
4. D. Zehn *et al*., *Nat. Immunol*. **23**, 836 (2022).
5. T. G. Kang *et al*., *Science* **386**, eadl4492 (2024).
6. J. A. Belk *et al*., *Nat. Immunol*. **23**, 848 (2022).
7. D. R. Sen *et al*., *Science* **354**, 1165 (2016).
8. H. E. Ghoneim *et al*., *Cell* **170**, 142 (2017).
9. J. A. Fraietta *et al*., *Nature* **558**, 307 (2018).
10. C. L. O'Connell *et al*., *Clin. Cancer Res*. **28**, 5306 (2022).
11. S. Jaiswal, B. L. Ebert, *Science* **366**, eaan4673 (2019).
12. R. Belizaire *et al*., *Nat. Rev. Immunol*. **23**, 595 (2023).
13. K. M. Cappell, J. N. Kochenderfer, *Nat. Rev. Clin. Oncol*. **20**, 359 (2023).
14. N. Prokhnevska *et al*., *Immunity* **56**, 107 (2023).
15. C. Tsui *et al*., *Nature* **609**, 354 (2022).

Tiny comets under the sea

A mucus drag on sinking marine snow may have big biogeochemical implications

By **B. B. Cael**[1] and **Lionel Guidi**[2,3]

Phytoplankton produce organic matter in the sunlit upper ocean, which forms particles of "marine snow" that sink and transfer carbon from the surface to the deep ocean. These particles sustain a vast carbon reservoir and are consumed to provide energy to ecosystems in the deep sea (*1*). Deciphering what controls carbon transfer by marine snow has been a central question in oceanography since the 1980s. However, the formation, sinking, and consumption of this particulate matter involve many physical, chemical, and biological processes. Thus, the overall impact of sinking particles on the ocean's carbon cycle remains highly uncertain and poorly understood (*2*). On page 166 of this issue, Chajwa *et al*. (*3*) report mucus "comet tails" of marine snow that substantially affect the sinking speed and transfer of carbon in the upper ocean. These observations provide new insights into how the composition of marine particles alters ocean ecology and biogeochemistry.

Decades of research into marine snow's physical and ecological properties has helped disentangle their intricate nature. Physically, the particles often exhibit complex structures with fractal-like geometry (*4*). This architecture substantially affects the sinking speed, formation, fragmentation, and flow dynamics of marine snow (*5*, *6*). Ecologically, the particles emit a long tail of chemical plumes that aid motile organisms in locating them (*7*). They also act as islands of bacterial activity and interactions (*8*). However, mechanistic understanding of how marine snow falls into the deep ocean, and how this affects the global carbon cycle, has been challenging because it is difficult to directly observe their sedimentation.

To disentangle the complex processes of sinking particles, Chajwa *et al*. combined

[1]National Oceanography Centre, Southampton, UK. [2]Sorbonne Université, CNRS, Laboratoire d'Océanographie de Villefranche, LOV, Villefranche-sur-mer, France. [3]Research Federation for the Study of Global Ocean Systems Ecology and Evolution, FR2022/Tara GOSEE, Paris, France. Email: cael@noc.ac.uk; lionel.guidi@imev-mer.fr

sediment trap sampling, tracking microscopy (9), and particle-flow interactions (10) to resolve the hydrodynamics of naturally occurring marine snow. The authors collected particle aggregates in the Gulf of Maine and then suspended them in a controlled wheel of rotating fluid to directly image their sedimentation dynamics. A detailed observation of flow fields within and surrounding marine snow revealed that an invisible mucus coated the particles, with comet-like mucus tails trailing behind them. Chajwa et al. developed a fluid mechanical theory to derive the drag force applied to the particles by these comet tails. The model estimated nearly double the residence time of the sinking aggregates in the upper ocean at the collection site. Previous studies reported the presence of transparent exopolymers (organism-secreted biopolymers) on marine snow and their importance on the particles' hydrodynamics (11). However, Chajwa et al. have now resolved the fluid flow impacts and modeled how they alter the dynamics of particle sinking.

The comet tails slow down the sinking of marine snow and reduce the amount of carbon transferred and sequestered by these aggregates. Consequently, this increases Earth's atmospheric carbon dioxide concentration (12). If such comet tails are ubiquitous, as suggested by the authors' preliminary findings from other measurements made during other research cruises, then the drag that these mucus tails apply to particles could substantially reduce both the amount of carbon sequestered from the atmosphere and exported out of the upper ocean by sinking particles. Chajwa et al.'s model suggests that the difference between the atmospheric carbon dioxide calculated with and without comet tails is on the order of 40 parts per million. Equivalently, the sinking particles with mucus tails sequester a couple hundred gigatonnes less carbon from the atmosphere (see the figure). Although such estimates are very approximate at this stage, the large magnitude change in carbon sequestration highlights the need to study this phenomenon further.

The comet tails may also modify other

Hidden comet tails affect carbon transfer and storage

A mucus comet tail on a marine snow particle has a potential effect on the amount of carbon sequestration in the ocean. Comet tails slow down the sinking particle, which reduces the amount of carbon transfer and storage in the deep ocean.
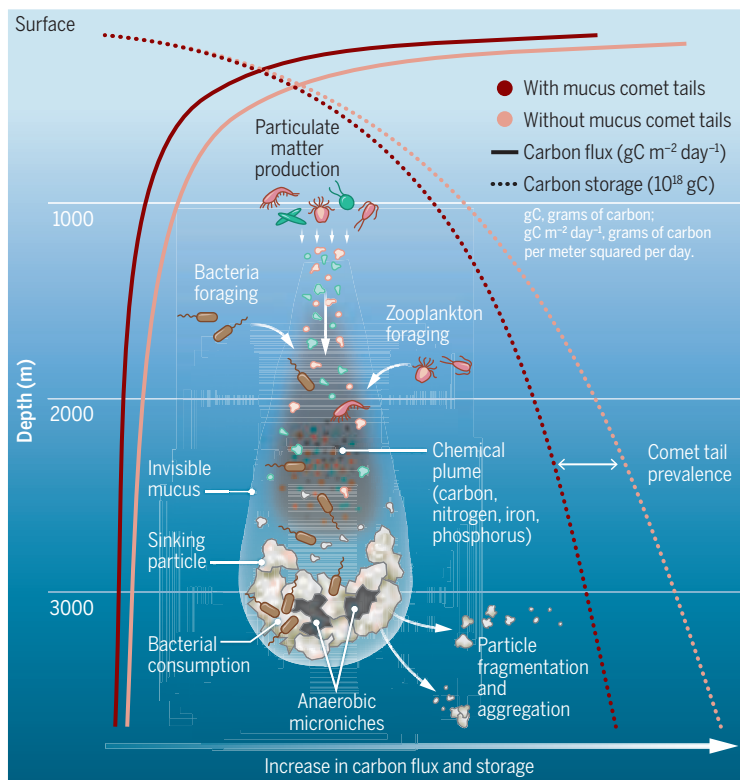
important aspects of marine snow dynamics and related ocean biogeochemical processes. Mucus-bound particles are more likely to aggregate and less likely to fragment because of the strong binding force provided by the mucus. The existence of an "invisible" volume filled by mucus comet tails also makes particles far less porous to fluid flow than their sparse shapes suggest. Estimating this invisible volume should help resolve the variable relationship between aggregates' size and sinking speed. Nonetheless, the biogeochemical consequences of these comet tails are ultimately more important than their physical properties. Impeded flow of material through particles may facilitate microenvironments that are hotspots for key anaerobic reactions such as denitrification (13). This may also influence how aggregates are consumed by bacteria and zooplankton, both by making particles easier to detect for foragers and by influencing how material is consumed by bacteria within the particulate materials (14). If these mucus tails are also themselves a source of carbon and energy, it may help explain the missing gap between food demand by deep ocean ecosystems and the energy supply to the deep ocean (15).

Nevertheless, the study of Chajwa et al. focuses on a single geographic location dur-

ing a phytoplankton bloom, when the particle populations are more likely to be homogeneous in their physical and biological properties. To scale up this work to the global ocean accurately, it is essential to perform similar measurements across different locations, depths, and times. Fortunately, the method of Chajwa et al. can be performed at sea on research vessels, thereby allowing their measurements to be replicated at different locations with different ecosystem conditions. A collection of observations is necessary to translate detailed process-based information into broader biogeochemical estimates. Additionally, these measurements can be paired with in situ analyses of sinking marine snow, such as fractal dimension assessments, to account for any potential effects that may come from handling the samples. A comparison of the properties as measured in the present study to other in situ measurements, such as the structure of the plankton community out of which these particles are formed, or to the magnitude of fluid turbulence to which particles are exposed may also yield important insights. ∎

REFERENCES AND NOTES

1. P. Boyd, H. Claustre, M. Levy, D. Siegel, T. Weber, *Nature* **568**, 327 (2019).
2. M. Iversen, *Annu. Rev. Mar. Sci.* **15**, 357 (2023).
3. R. Chajwa, E. Flaum, K. D. Bidle, B. Van Mooy, M. Prakash, *Science* **386**, eadl5767 (2024).
4. B. Logan, D. Wilkinson, *Limnol. Oceanogr.* **35**, 130 (1990).
5. A. Burd, G. Jackson, *Annu. Rev. Mar. Sci.* **1**, 65 (2009).
6. N. Briggs, G. Dall'Olmo, H. Claustre, *Science* **367**, 791 (2020).
7. T. Kiørboe, G. Jackson, *Limnol. Oceanogr.* **46**, 1309 (2001).
8. M. Simon, H. Gsrossart, B. Schweitzer, H. Ploug, *Aquat. Microb. Ecol.* **28**, 175 (2002).
9. D. Krishnamurthy et al., *Nat. Methods* **17**, 1040 (2020).
10. U. Alcolombri et al., *Nat. Geosci.* **14**, 775 (2021).
11. U. Passow, *Prog. Oceanogr.* **55**, 287 (2002).
12. E. Kwon, F. Primeau, J. Sarmiento, *Nat. Geosci.* **2**, 630 (2009).
13. D. Bianchi, T. Weber, R. Kiko, C. Deutsch, *Nat. Geosci.* **11**, 263 (2018).
14. D. H. Rothman, D. C. Forney, *Science* **316**, 1325 (2007).
15. S. Giering et al., *Nature* **507**, 480 (2014).

GRAPHIC: K. HOLOSKI/SCIENCE

ARTIFICIAL INTELLIGENCE

Considerations for governing open foundation models

Different policy proposals may disproportionately affect the innovation ecosystem

By **Rishi Bommasani**[1,2]**, Sayash Kapoor**[3,4]**, Kevin Klyman**[1]**, Shayne Longpre**[5]**, Ashwin Ramaswami**[6]**, Daniel Zhang**[7]**, Marietje Schaake**[7,8]**, Daniel E. Ho**[1,7,9,10,11,12]**, Arvind Narayanan**[3,4]**, Percy Liang**[1,2]

Foundation models (e.g., GPT-4 and Llama 3.1) are at the epicenter of artificial intelligence (AI), driving technological innovation and billions of dollars in investment. This has sparked widespread demands for regulation. Central to the debate about how to regulate foundation models is the process by which foundation models are released (1)—whether they are made available only to the model developers, fully open to the public, or somewhere in between. Open foundation models can benefit society by promoting competition, accelerating innovation, and distributing power. However, an emerging concern is whether open foundation models pose distinct risks to society (2). In general, although most policy proposals and regulations do not mention open foundation models by name, they may have an uneven impact on open and closed foundation models. We illustrate tensions that surface—and that policy-makers should consider—regarding different policy proposals that may disproportionately damage the innovation ecosystem around open foundation models.

The landscape of foundation model release (3) is multidimensional: Different assets (e.g., training data, code, and model weights) can be released to select entities or to the public at large. Developers have many options along a gradient of model release [see the figure; modified with permission from (3)]. We use the term open foundation models to identify foundation models with widely available weights, which today indicates that the weights are available for free. This aligns with the distinction drawn in the 2023 US Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, which tasked the National Telecommunications and Information Administration with preparing a report on open foundation models for the president. This is but one example of attention being given to foundation models, including open models, by governments around the world. Under the European Union (EU) AI Act, open foundation models trained with fewer than 10^{25} floating point operations (a measure of compute expenditure) are exempt from many requirements. The UK AI Safety Institute considers "open-source systems as well as those deployed with various forms of access controls" as a key priority.

Many of the concerns surrounding open foundation models arise from the fact that once model weights are released, developers relinquish control over their downstream use. Even if developers attempt to restrict downstream use, such restrictions can be ignored by malicious users. By contrast, in the face of malicious use, developers of closed foundation models can restrict access to them. It should be stressed, however, that this categorical distinction may oversimplify the gradient of model release: Closed models are also susceptible to malicious use, given that current safeguards are circumventable.

Open foundation models are reminiscent of, but different from, open-source software. Machine learning models depend on datasets as well as code, making them fundamentally different from most software. The standard definition from the Open Source Initiative for open-source software prohibits restrictions on specific users or use cases, whereas open foundation models often include these restrictions; Meta restricts the use of its Llama 3.1 model by entities with more than 700 million monthly active users, and other organizations use open and responsible AI licenses with use restrictions. These differences have led to claims that leading AI companies are "openwashing"—providing model weights while not following the principles of open-source software (4).

Nevertheless, the history of open-source software provides insight on how to govern open foundation models. There is no empirical evidence that open-source software is more vulnerable or insecure than closed-source software (5). At the same time, open-source software validates the tremendous societal benefits of open technologies, for example by stimulating economic benefit; supporting critical infrastructure; promoting reusability, robustness, transparency, and collaboration; and improving reliability and security through continuous and broad peer review.

BENEFITS OF OPEN MODELS

We highlight three fundamental societal objectives where open foundation models provide clear benefits.

Distributing power

Given the growing influence of foundation models, they create new forms of socioeconomic power, which demands an assessment of how that power is distributed. Closed model developers exert greater power in defining and restricting use cases that they deem unacceptable, whereas downstream consumers of foundation models can better make these decisions for themselves with open models. Further, closed model developers may more directly shape downstream markets through vertical integration, potentially leading to problematic monocultures where many downstream products and/or services depend on the same foundation model. Overall, closed foundation models may contribute to an increase in concentrated power in the hands of developers, which should be scrutinized given the well-established risks of market concentration for digital technologies.

Catalyzing innovation

Foundation models are general-purpose technologies that can produce sharp increases in innovation. Notably, foundation models bolster economic and scientific productivity, with Bloomberg Intelligence projecting that generative AI will become a $1.3 trillion

market by 2032. Open foundation models are necessary for research on several topics, such as interpretability, watermarking, security, and efficiency. Overall, open foundation models are more customizable and provide deeper access, which are key ingredients for greater innovation.

Ensuring transparency

Digital technologies such as foundation models are plagued by opacity. Adequate transparency from foundation model developers is instrumental for many objectives: Civil society, governments, industry, and academia have all called for greater transparency. Transparency is important not just from the perspective of model training and release, but also in terms of downstream details, such as reporting model usage. The Foundation Model Transparency Index demonstrates that major open foundation model developers are more transparent on average compared with their closed counterparts (6). Such transparency may help to avoid reproducing the harms caused in part by opaque digital technologies in the past. In some instances, transparency hinges not only on the release of model weights but also other artifacts. For instance, the disclosure of training data and code aids reproducibility (4).

RISKS OF OPEN MODELS

Much of the policy focus on the risks of open foundation models has been motivated by their potential for malicious use (2). Here, we consider a range of misuse threat vectors to better characterize the state of evidence for each. Correctly characterizing the distinct risks of open foundation models requires centering marginal risk: To what extent do open foundation models increase risk relative to (i) closed foundation models or (ii) preexisting technologies, such as search engines? Kapoor *et al.* (7) provide a framework for analyzing the marginal risk of open foundation models. For many threat vectors, existing evidence of marginal risk is limited. This does not mean that open foundation models pose no risk along these vectors but, instead, that more rigorous analysis is required to substantiate policy interventions. Although some might propose regulation on the basis of the precautionary principle even in the absence of such evidence, the lack of evidence of marginal risk indicates that caution should be taken in imposing policies that substantially burden open foundation model developers.

Disinformation

Foundation models may reduce the cost of generating persuasive disinformation (8). Although closed foundation model providers are better positioned to encourage their models to reject requests to generate disin-

A gradient of model release

Developers have a range of options (spanning the gradient of shaded blue dots below) for whether and how to release particular assets (e.g., model weights, code, and training data) to particular parties or to the public at large. Modified from (3).

● **Fully closed**
Not available to anyone outside the developer organization, like Google DeepMind's Flamingo

● **Hosted access**
Available through a web interface, like Inflection's Pi

● **Cloud-based access to model**
Available through an application programming interface (API), like Anthropic's Claude 3

● **Cloud-based access to model and fine-tuning**
Both the model and the ability to fine-tune it are available through an API, like OpenAI's GPT-3.5

OPEN FOUNDATION MODELS WITH WIDELY AVAILABLE WEIGHTS

● **Widely available weights**
Like Meta's Llama 3.1

● **Fully open, with use restrictions**
Weights, code, and data available with use restrictions, like BigScience's BLOOM

● **Fully open, without use restrictions**
Weights, code, and data available without use restrictions, like EleutherAI's GPT-NeoX

formation, the ambiguity of what constitutes disinformation calls into question the technical feasibility of such refusals. More fundamentally, the key bottleneck for effective influence operations is not disinformation generation but disinformation dissemination: "Cheap fakes," such as videos from past events, Photoshop-edited images, events that occurred in another context, or even video game footage, have been used to spread disinformation (9). Online platforms that control the reach of content are better targets for policy intervention than foundation model developers. To date, we are unaware of empirical evidence that open foundation models increase societal susceptibility to disinformation campaigns.

Biorisk

Several studies have claimed that open foundation models can instruct users on how to construct bioweapons (10). But the evidence remains weak. Research indicating that today's language models provide "dangerous" information related to bioweapons does not acknowledge that the same information is available from other sources. When studies have compared the use of language models to internet access, they have found no substantial benefit of using language models for finding information relevant to bioweapon creation (11, 12). Concerns about open language models may be misplaced, as special-

ized biological design tools may instead provide greater leverage for discovering dangerous pathogens. Beyond sensitive information, the biorisk pipeline requires pathogen synthesis and dissemination in the real world. Each of these steps requires considerable expertise, equipment, and laboratory experience. As with other threat vectors, the best policy choke points may therefore lie downstream. For example, the US AI Executive Order aims to strengthen customer screening for purchasers of biological sequences.

Cybersecurity

Although open code models could improve the speed and quality of offensive cyberattacks, it appears that cyber defenses will also improve (13). For example, Google recently demonstrated that code models vastly improve the detection of vulnerabilities in open-source software. As with previous automated vulnerability-detection tools, widespread access to open models for defenders, supplemented by investment in tools for finding security vulnerabilities by companies and governments, could strengthen cybersecurity. Open models can be used and customized (e.g., through fine-tuning) locally, allowing organizations to use them in privacy-sensitive contexts.

Spear-phishing scams

Foundation models can generate high-quality spear-phishing emails seeking to persuade a victim to provide sensitive information, send money, or download malicious software (14). Both open and closed models could be used for this because the key factor that makes spear-phishing emails dangerous is the malware that accompanies the email; the text itself is usually benign. As with disinformation, the key bottleneck for spear phishing is not generally the text of emails but downstream safeguards: Modern operating systems, browsers, and email services implement several layers of protection against such malware. Phishing emails might not reach the intended recipient in the first place owing to these existing protections.

Voice-cloning scams

Voice-cloning scams, where malicious users impersonate a person's friends or family and persuade them to transfer money, may rely on foundation models that can clone someone's voice based on a few seconds of audio—for instance, from their social media account. As of now, it is unclear whether voice-cloning scams are more effective or scalable compared with traditional scams, especially because tens of thousands of traditional scams are already being reported each year. Although it is yet to be determined whether closed model developers can successfully pre-

vent such scams, they do offer a measure of deterrence by, for instance, requiring users to sign up using credit cards and having the ability to trace any audio back to the specific user who created it.

Nonconsensual intimate imagery (NCII) and child sexual abuse material (CSAM)

Open text-to-image models appear to present distinct risks related to NCII and CSAM because they lower the barrier to generating such content. Safeguards for closed models are more effective in this area, and monitoring closed models can deter users from generating such imagery, especially of real people. Open text-to-image models have already been used to create NCII and CSAM. Thiel found that a prominent dataset for training open text-to-image models contains vast amounts of CSAM, pointing to upstream interventions, such as training data filtering, for mitigating this risk (*15*). There remains an open question about whether policy interventions to combat the spread of AI-generated NCII and CSAM are more effective if targeted at downstream platforms, such as Civit AI and social media companies. Organizations that are tasked with combating NCII and CSAM, such as the National Center for Missing & Exploited Children, may benefit from additional resources and support to address AI-generated CSAM.

POTENTIAL ADVERSE IMPACTS

As policy efforts across the United States, China, the EU, the United Kingdom, and the G7 nations focus on foundation models, we consider how policy initiatives in these jurisdictions may affect open foundation models. Specifically, they may impose greater compliance burdens on open foundation model developers compared with their closed counterparts, even when open developers are more likely to be resource poor compared with the largest AI companies, which are disproportionately closed developers.

Liability for downstream use

Because the distinction between open and closed foundation models is predicated on release, policies that impose penalties for certain uses of foundation models are likely to have differential impacts. Some proposals, such as SB 1047 introduced in the California State Senate and the framework for a US AI Act proposed in the US Senate, impose liability for the downstream use of a foundation model, including for derivatives of the foundation model that are the result of fine-tuning. These proposals aim to introduce penalties for the release of unsafe models that, potentially after modification, catalyze misuse. Yet, liability for such downstream harms could chill the open foundation model

ecosystem by exposing open foundation model developers to severe liability risk. By contrast, because closed foundation model developers exercise greater control over downstream use, some developers already provide liability protections to downstream users of their models (e.g., Google offers users of its generative AI products indemnification for copyright claims). Although clarifying or increasing liability for downstream use may have benefits, these legislative proposals expose a broad and hard-to-control liability surface for open foundation model developers.

Content provenance for downstream use

Given that the most salient applications of foundation models are generative AI systems, there is demand for content provenance techniques, such as watermarking, to detect machine-generated content. Content provenance could help with the tracking or moderation of AI-generated content, such as deepfakes, CSAM, and NCII. But akin to liability, if foundation model developers must ensure content provenance for downstream use, then these requirements may be technically infeasible for open foundation model developers.

The US Executive Order, White House voluntary commitments, the Canadian Voluntary Code of Conduct, Chinese generative AI regulations, and the G7 international code of conduct all highlight content provenance. However, today's watermarking methods for language models do not persist if models are modified (e.g., fine-tuned) and require that users of a model follow certain protocols for the watermarking guarantee to hold. Fundamentally, open foundation model developers do not control how their models are modified or used to generate content.

Liability for open data

Although foundation models do not require the release of the underlying data used to build the model, some developers choose to release both the model weights and the training data. Of the 10 major foundation model developers assessed by the 2023 Foundation Model Transparency Index (*5*), the two that released data openly also released their foundation models openly. A number of other open foundation model developers tend to release data openly. However, open release of data exposes these entities to greater liability risk, as exemplified by lawsuits against Stability AI based on its use of datasets from the nonprofit Large-scale Artificial Intelligence Open Network (LAION) that allegedly included plaintiffs' work. Although the legality of training foundation models on copyrighted data remains unclear across many jurisdictions, the status quo presents perverse incentives. Namely, model develop-

ers that transparently disclose and openly provide data are subject to greater risk than developers that conceal the data that they use, even if the underlying facts are identical. Considering this perverse incentive, government-mandated disclosure of training data may be beneficial in some cases.

CONCLUSIONS

The design and implementation of policies being crafted by governments around the world should consider both open and closed foundation model developers. When regulations directly address open foundation models, the precise definition used to identify these models and developers should be duly considered. Hinging regulation exclusively on open weights may not be appropriate given the gradient of release. And even when regulations do not directly address open foundation models, they may have an adverse impact. Consequently, if policy-makers are to implement such policies, they should directly consult the open foundation model community, with due consideration given to their interests ■

REFERENCES AND NOTES

1. P. Liang, R. Bommasani, K. Creel, R. Reich, "The Time Is Now to Develop Community Norms for the Release of Foundation Models" (Stanford Institute for Human-Centered Artificial Intelligence, 2022).
2. E. Seger *et al.*, "Open-Sourcing Highly Capable Foundation Models: An evaluation of risks, benefits, and alternative methods for pursuing open-source objectives" (Centre for the Governance of AI, 2023).
3. I. Solaiman, "The Gradient of Generative AI Release: Methods and Considerations" in *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (ACM, 2023), pp. 111–122.
4. M. White *et al.*, https://arxiv.org/abs/2403.13784 (2024).
5. G. Schryen, *Commun. ACM* **54**, 130 (2011).
6. R. Bommasani *et al.*, https://arxiv.org/abs/2310.12941 (2023).
7. S. Kapoor *et al.*, "On the Societal Impact of Open Foundation Models" (International Conference on Machine Learning, 2024).
8. J. A. Goldstein *et al.*, https://arxiv.org/abs/2301.04246 (2023).
9. B. Paris, J. Donovan, "Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence" (Data and Society, 2019).
10. E. H. Soice *et al.*, https://arxiv.org/abs/2306.03809 (2023).
11. T. Patwardhan *et al.*, "Building an early warning system for LLM-aided biological threat creation" (OpenAI, 2024).
12. C. A. Mouton, C. Lucas, E. Guest, "The Operational Risks of AI in Large-Scale Biological Attacks: A Red-Team Approach" (RAND Corporation, 2024).
13. M. A. Ferrag *et al.*, https://arxiv.org/abs/2307.06616v1 (2023).
14. J. Hazell, https://arxiv.org/abs/2305.06972 (2023).
15. D. Thiel, "Identifying and Eliminating CSAM in Generative ML Training Data and Models" (Stanford Digital Repository, 2023).

BIOLOGY

Denizens of the dark

An exploration of night-dwelling organisms encourages an appreciation of all things nocturnal

By **Sarah Boon**

n *Night Magic*, author Leigh Ann Henion explores Appalachia to find flora and fauna that thrive and survive at night. Her goal is to share the marvelous nocturnal organisms in her backyard, dividing her narrative into seasons—spring, summer, and fall—with different species discussed in each one.

Henion connects with a variety of experts and participates in a range of community science events to see the denizens of the night. She starts in spring with Grandfather Glows in Linville, North Carolina—an event for which tickets sold out in 60 seconds. Here, visitors are treated to the sight of fireflies, including one species known as "blue ghosts," as well as "glowworms," which are maggots that live on the ground and glow blue. The encounter leaves Henion scouring her own rural neighborhood for more luminescent creatures, wondering whether neighbors will be suspicious of what she is doing looking in ditches and under leaf litter. "I'm beginning to know darkness, for the first time in my life, not as a realm of unknowns best avoided, but rather as a potential state of enchantment," she writes.

Henion goes on a nighttime salamander tour near Appalachian State University, where she learns more about those strange creatures that can regenerate parts of themselves. They only live in vernal pools, and thus have a very short life span, and artificial light compromises their regenerative abilities. In Alabama's

Night Magic: Adventures Among Glowworms, Moon Gardens, and Other Marvels of the Dark
Leigh Ann Henion
Algonquin Books, 2024.
336 pp.

Bankhead National Forest, she goes on a "bat blitz" in hazmat garb (because bats can transfer pathogens to humans and vice versa) and later attends Mothapalooza, an event for amateur naturalists in Bainbridge, Ohio, that often has a waiting list. There she discovers the captivating giant leopard moth. She joins a group walk to find fox fire, the glowing bioluminescence of tree fungus, and watches owls from a colleague's porch. She and a friend organize a nighttime flower viewing to observe tobacco plants open as the sun goes down—the first time either of them has seen this phenomenon.

Historically, the night has been portrayed as sinister and scary—something to stay away from or to keep at bay with artificial light. Henion notes that we are afraid of not just the dark itself but also who or what might be out in it. "The things we fear most aren't usually the things that wind up hurting us," a friend reminds her. "If you're frightened by something, it probably means you have something to learn from it," advises Vicky, a bat caretaker.

Henion argues that we need more dark preserves, free of artificial light. Birds die by the thousands each year because they crash into artificially lit windows or lose their sense of direction owing to unexpected illumination. Artificial light can also make birds migrate up to a month early, meaning that they may arrive at their destination well before the species they eat and so could starve in the meantime. Dark preserves could also help trees. When they are under streetlights, trees do not close their evapotranspiration pores, making them more sensitive to pollution and drought.

A common theme in *Night Magic* is the destruction of once-dark habitats and the

Fireflies light up the night in synchrony in Tennessee's Great Smoky Mountains National Park.

ubiquity of artificial light. Henion takes her own steps to reduce artificial light, closing her curtains at night so insects are not drawn to the indoor light and turning off her porch lights. She also spends more time outside at night. She is more comfortable now that she knows what she is looking for. She even learns to make fire using a bowstring and a plank to create her own light in the darkness.

There are many nocturnal creatures that people do not know about because we do not often explore the landscape at night. There are also many nocturnal creatures unknown to scientists, who likewise do not study night-dwellers as often as they do those that are active during the day. Henion shares ideas for maintaining dark habitats to ensure the continued survival of these mysterious organisms, including replacing blue LED streetlights with something softer. As she writes, "If darkness is not re-centered as a force that nurtures and harbors life...it's doubtful that we will ever be able to shift the cultural conditioning that has led to a broad-spectrum lauding of light and the vilification of darkness. No one nurtures what they view as worthless."

Overall, *Night Magic* was an enjoyable, if formulaic, read, following the same structure as many nature books. Some of Henion's asides into her own life had tenuous connections, at best, with the subject under consideration, and—in my opinion—the book would have been stronger without much of the last section, as Henion does not clearly explain how making fire fits with the focus on nocturnal creatures. However, the book ultimately succeeds in making readers more aware of the nocturnal organisms with whom we share the planet, which is a huge positive. ∎

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

10.1126/science.adr8873

COMMUNICATION

The life of "ideas worth spreading"

TED turns 40 in a rapidly evolving media landscape

By **Alex Gomez-Marin**

"Ideas worth spreading" has long been the visionary slogan of TED, the US-Canadian media organization famous for its punchy lectures known as TED Talks. Founded 40 years ago by architect and graphic designer Richard Saul Wurman and broadcast designer Harry Marks, TED—an acronym meant to convey the group's interest in technology, entertainment, and design—launched its first conference in 1984, and, after some years of financial trouble, the events resumed in 1990 with annual conferences that attracted assorted speakers and enthusiast audiences with a transdisciplinary vibe filled with curiosity and excitement.

In 2001, entrepreneur Chris Anderson acquired the brand and became its curator. In 2006, recordings of the high-priced exclusive events started to be posted online for free, some reaching tens of millions of views. Then, in 2009, the TEDx program was created, allowing for independently organized events following TED's recipe. These moves catapulted the company globally as a major platform promoting "powerful ideas, powerfully presented" and settling it as a fascinating case study for the democratization of elitism.

Today, TED has become a very mixed bag, covering all sorts of topics—from creativity and poverty to procrastination and orgasms—and featuring speakers ranging from Naomi Klein and Monica Lewinsky to Bono and yours truly. The growth of such a global media empire has been staggering, with talks translated into 100 languages and watched billions of times. However, as the numbers have exploded, has the quality imploded? The hyperbolic insistence on inspirational ideas seems to have banalized them. To be quite honest, I have not watched a TED Talk in a long time.

Are humans capable of absorbing a life-changing idea every 20 min, applauding, and then moving on to the next one? What is the point of a talk on the seven rules for making successful seven-rule talks? Or about the significance of water when you are thirsty? Making what is trivial sound profound is hardly a new art form. Moreover, the "thought leader" fragrance may seduce, but it is often cloying, even risible. TED's clichéd style is both loved and hated.

Regardless of one's personal feelings about the platform, TED's impact on the media landscape, and particularly on science outreach, is noteworthy. The talks have popularized an idiosyncratic way of communicating ideas, elevating the aesthetics of staging and putting format on equal footing with content. TED Talks have also helped academics reach wildly wider audiences, raising the bar of our communication skills by spurring conciseness, expressive body language, and stunning slides. Science is storytelling too, after all. Teaching and lecturing have thus become an educated way of entertaining and networking.

TED reimagined traditional lecture formats, emphasizing aesthetics and performance as key factors in communicating big ideas to broad audiences.

What, if anything, makes TED's content special today? Its current challenges mirror the opportunities of a vibrant, evolving, and uncertain media landscape. We are saturated daily with an ever-growing, overwhelming bulk of new content, mercilessly sinking in the bottomless ocean that is the internet. Content creators must also deal with the hard problem of attention and its puzzling timescales: from the 18-s torpid TikTok clip, to the TED Talk's 18-min golden rule, to the booming 180-min long-conversation format. Furthermore, podcasting has completely opened the gates to self-managed curation, directly matching audiences and content that would be mutually inaccessible otherwise.

However, the diversity of views available to us ossifies and suffocates as our algorithmic echo chambers turn the virtual "public square" into a pit of fear, anger, and division. In turn, the globalization of ideas has incited their pasteurization, resulting in shadow banning and gaslighting when internet arbiters deem material infectious to society's noosphere. TED has not been immune to unfortunate controversies, including the cancellation of some provocative presentations. As it turns out, some ideas are not worth spreading.

The printed word has relentlessly dominated society for the last half millennium. But we now seem to be going through an archaic revival in the age of screens, expanding our personal Plato's caves through images and the spoken word. Nonetheless, the near future of content generated by artificial intelligence poses an unfathomable threat to thought and society. Real conversations—with real people and real voices—are the midwifery of ideas. Chatbots are the end of real talks.

TED's new tagline is "Ideas change everything," and its updated mission reads as follows: "Discover and spread ideas that spark conversation, deepen understanding, and drive meaningful change." I can only hope that, beyond personal posturing and professional branding, it can truly embody that spirit. ∎

The reviewer is at the Instituto de Neurociencias, Consejo Superior de Investigaciones Científicas–Universidad Miguel Hernández de Elche, Alicante, Spain. Email: agomezmarin@gmail.com

QUALITY CONTENT FOR THE GLOBAL SCIENTIFIC COMMUNITY

Multiple ways to stay informed on issues related to your research

Posters

Podcasts

Sponsored Collection Booklets
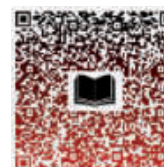
Sponsored Feature
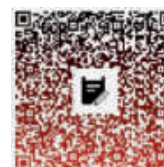
Webinars

Science
AAAS

Scan the code and start exploring the latest advances in science and technology innovation!

Science.org/custom-publishing

The illegal horn trade has contributed to the decline of white rhinocerus populations.

Edited by **Jennifer Sills**

Protect white rhinos by legalizing horn trade

The conservation of rhinos has generally been a success story: Southern white rhinos have increased from a few dozen 100 years ago to more than 20,000 in 2011 (*1*). Initially, population increases were hailed as an outcome of the international trade ban on rhino horn, established in 1977 and enforced by the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) (*2*); however, increased poaching has led numbers to fall again in recent years. This longstanding policy, in which countries without rhinos impose bans on rhino-holding states, has not worked (*3*). Legalizing the rhino horn trade would improve conservation success.

Rhino numbers had already increased before the ban was established, particularly in areas where private rhino guardianship was allowed, such as in South Africa (*4*). Roughly half of the world's white rhino population is privately managed today, providing important ecosystem services (*5*). However, since 2010, demand for white rhino horn and large-scale poaching activities have increased (*6*). Since then, white rhino numbers have decreased by roughly 1000 per year (*7*). By 2021, international poaching syndicates had surmounted every conservation countermeasure as horn prices exceeded gold prices in international illegal markets (*8*).

Facing insurmountable security costs and the loss of personal safety, rhino guardians are now giving up their custodianship (*9*). At the same time, state rhino sanctuaries, such as Kruger National Park, have decided to continuously dehorn rhinos to deter poaching, with limited success (*10*). The high number of deaths has destroyed the rhinos' social systems and changed their behavior (*11*).

Rhino horn is a renewable resource that grows like fingernails. To protect white rhino populations effectively, the international trade of rhino horn should be legalized and carefully monitored, including the tracking of each sample sold. Legal trade—the preferred option of horn consumers (*3*)—will derail international trafficking syndicates (*12*) and enable essential private rhino guardianship. Legalized trade provides the best chance for the future of white rhino populations (*3*, *12*).

Martin Wikelski[1,2]
[1]Department of Migration, Max Planck Institute of Animal Behavior, Radolfzell, Germany. [2]Department of Biology, University of Konstanz, Konstanz, Germany. Email: wikelski@ab.mpg.de

REFERENCES AND NOTES
1. N. Leader-Williams, in *The Trade in Wildlife: Regulation for Conservation*, S. Oldfield, Ed. (Routledge, 2014), pp. 111–121.
2. M. Abensperg-Traun, *Biol. Conserv.* **142**, 948 (2009).
3. H. Cheung, L. Mazerolle, H. P. Possingham, D. Biggs, *Front. Ecol. Evol.* **9**, 607660 (2021).
4. I. Player, *The White Rhino Saga* (Jonathan Ball Publishers, 2013).
5. B. Büscher, *J. Polit. Ecol.* **28**, 696 (2021).
6. L. Vigne, E. Martin, B. Okita-Ouma, *Pachyderm* **43**, 73 (2007).
7. A. Chanyandura, V. K. Muposhi, E. Gandiwa, N. Muboko, *Ecol. Evol.* **11**, 5892 (2021).
8. T. C. Haas, S. M. Ferreira, *PLOS One* **11**, e0167040 (2016).
9. R. Truscott, *New Scientist* **259**, 12 (2023).
10. S. M. Ferreira, L. Dziba, *J. Nat. Conserv.* **72**, 126359 (2023).
11. V. Duthé *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2301727120 (2023).
12. J. A. J. Eikelboom *et al.*, *Glob. Ecol. Conserv.* **23**, e01145 (2020).

10.1126/science.adq5925

Legal rhino horn trade: A thorny proposition

Between 2019 and 2023, 26 of the remaining 70 Javan rhinos were poached by two organized groups that have evaded authorities for years (*1*). In the first half of 2024, 229 South African rhinos were poached for their horns (*2*), destined for illegal markets as traditional medicine or status symbols. These incidents, driven by the illegal horn trade's entanglement with organized global crime syndicates (*3*), put rhino species at risk of extinction. Maintaining and enforcing the international ban on the rhino horn trade remain crucial for wild rhino preservation.

Advocates of legalizing international rhino horn trade suggest that a legal competing market could offset illegal trade by providing financial incentives to farm rhinos and harvest their horns painlessly at regular intervals (*4*). However, this approach carries substantial and unpredictable risks that could expedite the extinction of rhinos (*5*). Wildlife farming has had inconsistent results among other species (*6*). Farmed crocodile products, for example, have successfully replaced wild crocodile products because farmed products are cheaper to produce and are perceived as better by consumers. By contrast, when farmed stocks of bear bile became available, demand for the product grew. Cartels exploited the expanded market, and pressure on wild populations increased (*6*).

In addition to the risks posed by wildlife farming, a legal rhino horn market could increase demand, provide opportunities for money laundering, and complicate law enforcement's ability to distinguish legal sources from illegal sources (*7*). Moreover, high corruption levels increase the risk of diverting legal schemes to illegality (*8*). Current illegal rhino horn demand already far exceeds the potential legal supply and is projected to grow as wealth increases in consumer countries (*9*). Demand for rhino horn is

income led and inelastic to price changes (*10*), and there is a high risk that Asian and "wild" rhino horn (i.e., poached rather than farmed) will remain in demand even if legal trade is established (*6, 9*). On a deeper level, commodifying a threatened species—or any natural resource—calls for considering moral values and ethical implications (*11*), especially when putting species at risk of extinction.

Strengthening local conservation measures can help stem population declines in the short term. Antipoaching, antitrafficking, and anticorruption efforts can combat the activities of organized global syndicates. Improved habitat protection and local community engagement will help to protect these species on the ground.

In the long term, the priority must be to reduce the demand and reconcile traditional medicine with wildlife conservation. Traditional medicine, the preferred or only option for health care in many regions, has had devastating effects on the wildlife that serves as the source for its components. Sustainable solutions must be sought through open and culturally nuanced rather than polarized discussions (*12*). Through such outreach, consumers may become more open to accepting substitutes for wildlife parts or using alternatives to traditional medicine. Addressing demand for traditional medicine would decrease the market for products such as rhino horn, providing more sustainable protection than attempting to meet growing demand by legalizing trade.

R. J. M. Nuijten[1,2]*, T. Nguyen[3], J. A. J. Eikelboom[2], O. Drori[4]
[1]Future For Nature Foundation, Arnhem, Netherlands. [2]Wildlife Ecology and Conservation Group, Wageningen University and Research, Wageningen, Netherlands. [3]WildAct Vietnam, Khuong Trung, Thanh Xuan, Hanoi, Vietnam. [4]The Eagle Network, Nairobi, Kenya.
*Corresponding author.
Email: rascha.nuijten@gmail.com

REFERENCES AND NOTES

1. "Javan rhino poaching saga reveals serious security lapse," *Mongabay-Indonesia*, 2 August 2024; https://news.mongabay.com/2024/08/javan-rhino-poaching-horn-ujung-kulon-national-park-indonesia-wildlife-trafficking/.
2. Department of Forestry, Fisheries and the Environment, Republic of South Africa, "Minister George welcomes downward trend in rhino poaching numbers in the last two months" (2024); https://www.dffe.gov.za/mediareleases/george_rhinostats.
3. United Nations Office on Drugs and Crime, "World Wildlife Crime Report 2024: Trafficking in Protected Species" (2024).
4. D. Biggs, F. Courchamp, R. Martin, H. P. Possingham, *Science* **339**, 1038 (2013).
5. D. J. Crookes, J. N. Blignaut, *J. Nat. Conserv.* **28**, 11 (2015).
6. L. Tensen, *Glob. Ecol. Conserv.* **6**, 286 (2016).
7. J. A. J. Eikelboom *et al.*, *Glob. Ecol. Conserv.* **23**, e01145 (2020).
8. E. L. Bennett, *Conserv. Biol.* **29**, 54 (2015).
9. USAID Wildlife Asia, "Research study on consumer demand for elephant, pangolin, rhino and tiger parts and products in China" (2018).
10. K. Conrad, *Trop. Conserv. Sci.*, **5**, 245 (2012).
11. A. A. Brown, A. J. Dean, H. Possingham, D. Biggs, *Conserv. Sci. Pract.* **1**, e103 (2019).
12. H. Cheung *et al.*, *People Nat.* **3**, 115 (2021).

Dangers of aging water infrastructure

In September 2023, more than 11,000 people died after torrential rainfall triggered the collapse of two dams in Libya (*1*). In May 2024, a flash flood in Mashhad, Iran, led to loss of lives because of poorly designed and malfunctioning stormwater drainage systems (*2*). Developing nations will increasingly confront similar hazards as their aging water infrastructure encounters more frequent and intense storms, population growth, and urbanization. Financial shortfalls, mismanagement, and corruption complicate efforts to upgrade infrastructure, exacerbating the dangers.

Much of the current water infrastructure has been in place for decades and was designed based on the incorrect assumption that historical extremes would not change substantially in the future. As a result, this infrastructure is unlikely to withstand evolving climate conditions and increasing hydrologic extremes. As these structures age, inspection and maintenance costs increase, and structural defects, such as cracks, become common (*3*). Compounded by more-frequent climate extremes, even minor structural defects can potentially lead to infrastructure failure (*4*).

Climate change is expected to increase precipitation intensity and frequency, causing more floods and sediment loads in some regions and leading to water shortages in others (*5*). Rising temperatures have led to more water vapor in the atmosphere (*6*) as well as more-extensive groundwater extraction (*7*), which contributes to land subsidence, further threatening critical infrastructure. In coastal regions, more-frequent and more-severe inundation caused by rising sea levels poses a threat to levees and onshore infrastructure.

The population of sub-Saharan Africa is expected to nearly double between 2022 and 2050, and the central and southern parts of Asia are projected to become the most populous regions of the world by 2037 (*8*). This population surge will drive an increased demand for energy, agricultural products, and water. Concurrently, urbanization will lead to the expansion and densification of urban areas (*9*), heightening exposure to floods and the probability of infrastructure failure.

Given the risks, countries cannot afford to upgrade infrastructure only after it fails. Yet, the disparity between current and target investments in water infrastructure is substantial, particularly in the Global South (*10*). Moreover, mismanagement and corruption lead to inadequate maintenance, substandard construction, and inefficient allocation of resources.

To address these challenges, comprehensive reforms are necessary. Governance frameworks must be strengthened, and engineering standards should reflect the need for higher resilience. Equally important, funding allocation and project implementation should be subjected to robust anticorruption measures, including increased transparency and accountability. Government institutions should build capacity to enhance their ability to effectively manage and oversee infrastructure projects. Collaborative efforts involving government agencies, civil society organizations, and international partners are essential to combatting mismanagement and corruption and to ensuring sustainable financing and maintenance of water infrastructure.

Ahmad Ferdowsi[1,2]*, Babak Zolghadr-Asli[3,4], Amir AghaKouchak[5,6]
[1]Department of Water Engineering and Hydraulic Structures, Faculty of Civil Engineering, Semnan University, Semnan, Iran. [2]University of Applied Science and Technology, Tehran, Iran. [3]The Sustainable Minerals Institute (SMI), The University of Queensland, Brisbane, QLD, Australia. [4]Centre for Water Systems, University of Exeter, Exeter, UK. [5]Department of Civil and Environmental Engineering, University of California, Irvine, CA, USA. [6]United Nations University Institute for Water, Environment, and Health, Hamilton, ON, Canada.
*Corresponding author.
Email: ahmad.ferdowsi@semnan.ac.ir

REFERENCES AND NOTES

1. I. A. Alvi, I. S. Alvi, *Civ. Eng. Environ. Syst.* **40**, 150 (2023).
2. Crisis24, "Iran: Disruptions due to flooding ongoing in parts of Mashhad as of May 16" (2024); https://crisis24.garda.com/alerts/2024/05/iran-disruptions-due-to-flooding-ongoing-in-parts-of-mashhad-as-of-may-16.
3. R. A. Mesalie, D. Aklog, M. S. Kifelew, *Appl. Water Sci.* **11**, 138 (2021).
4. F. Vahedifard, J. D. Robinson, A. AghaKouchak, *J. Geotech. Geoenviron. Eng.* **142**, 02516001 (2016).
5. E. Ragno *et al.*, *Water Resour. Res.* **54**, 1751 (2018).
6. K. E. Trenberth, J. T. Fasullo, T. G. Shepherd, *Nat. Clim. Change* **5**, 725 (2015).
7. G. Herrera-García *et al.*, *Science* **371**, 34 (2021).
8. UN Population Division Data Portal, "Rates of population growth vary significantly across regions" (United Nations, 2024); https://population.un.org/dataportal/home?df=50edb8c3-3680-48c7-9822-b3ee33d33736.
9. N. Kang *et al.*, *Water* **8**, 268 (2016).
10. E. Borgomeo, B. Kingdom, J. Plummer-Braeckman, W. Yu, *Int. J. Water Resour. Dev.* **39**, 895 (2022).

RESEARCH

Dodecagonal array in a scanning TEM image of 30° twisted-bilayer quasicrystal molybdenum diselenide

FERROELECTRICS

Polar vortices in twisted bilayers

Observations of electric fields in twisted molybdenum disulfide bilayers have revealed in-plane chiral vortex domains that depend on the twist angle. Tsang *et al*. used four-dimensional scanning transmission electron microscopy (TEM) and first-principles calculations to determine local polar domain structures, which may arise from twist stacking-induced charge redistribution and small in-plane ionic displacements. Mosaic chiral vortex patterns were seen for large twist angles. A 12-fold quasicrystal bilayer had complex vortex patterns that could be tuned by displacing layers within the microscope. —Phil Szuromi   *Science* p. 198, 10.1126/science.adp7099

ORGANIC CHEMISTRY
Asymmetrically adding prenyl

The prenyl group is a hydrophobic five-carbon fragment common across a wide range of contexts in biochemistry and natural product chemistry. One extensive class of natural products of pharmaceutical interest, polycyclic polyprenyl-ated acylphloroglucinols (PPAPs), has been the target of multiple synthetic studies, but the introduction of the prenyl groups asymmetrically has proven to be an enduring challenge. Ng *et al*. now report a copper-catalyzed method to prenylate unsaturated ketones and thereby access PPAPs with high enantioselectivity. Key to the method is an organoborate reagent with the boron in a somewhat counterintuitive position. —Jake S. Yeston

Science p. 167, 10.1126/science.adr8612

T CELLS
Regulation of female-biased immunity

X chromosome inactivation (XCI) is the process by which X-linked gene expression is repressed in female cells to balance gene dosage between the sexes. Unlike most somatic cells, regulation of XCI is more dynamic in T cells. Using allele-specific transcriptional and epigenomic profiling, Forsyth *et al.* found that T cell receptor activation and downstream nuclear factor κB (NF-κB) signaling are required for XCI maintenance in both mouse and human T cells. Inhibition of NF-κB prevented the relocalization of Xist RNA and repressive histone modifications to the inactive X chromosome. These findings indicate that defects in NF-κB–mediated regulation of XCI could contribute to sex-biased immune responses such as autoimmunity. —Claire Olingy

Sci. Immunol. (2024)
10.1126/sciimmunol.ado0398

MICROSCOPY
Subnanometer super-resolution

Measuring distances directly at the nanometer scale is a challenge for optical techniques, even for those using subdiffraction-resolution fluorescence microscopy. Sahl *et al.* refined an optical approach called MINFLUX such that they could measure precise intramolecular distances in the 1- to 10-nanometer range and below 1 nanometer for molecules with a tilt. Using a polyproline ruler, the authors demonstrate resolution of fluorophores with known single-digit nanometer spacing. They applied this approach to

inter- and intramacromolecular measurements of proteins labeled with photoactivatable dyes, including distances too short for current indirect methods. Imaging experiments demonstrated the potential of this technique to study protein-protein interactions in cells. —Michael A. Funk

Science p. 180, 10.1126/science.adj7368

SEED DISPERSAL
Species interactions at risk

Plants rely on seed dispersal, often by animals, for individual movement and population spread. Habitat loss and environmental change are known to threaten many plant and animal species, but the effects of population declines on the potential for seed dispersal are largely unknown. To address this knowledge gap, Mendes *et al.* synthesized data on plant-animal seed dispersal interactions for the whole of Europe and characterized each species interaction as being of low, high, or very high concern based on the species' conservation statuses and population trajectories. Almost one-third of all species interactions were found to be of high or very high concern, potentially further threatening plant species' persistence. However, data gaps are substantial, and this finding requires further study and evaluation. —Bianca Lopez

Science p. 206, 10.1126/science.ado1464

EPIDEMIOLOGY
Double trouble

Multiple strains of respiratory syncytial virus co-circulate, whereas severe acute respiratory syndrome coronavirus 2 strains tend to replace each other. To try to explain such heterogeneities, Park *et al.* developed a method to quantify immunological niches and fitness differences among competing pathogens in humans. This theory formalizes the phenomenon that during an epidemic, susceptible individuals become so rare that another strain cannot immediately invade. It takes time before a population of susceptible individuals emerges, which

depends on a combination of the strength and duration of immunity engendered by that pathogen, as well as any transmission advantage an invading variant may have. Thus, weak immunity tends to allow the coexistence of circulating strains, and strong immunity leads to serial replacement. —Caroline Ash

Science p. 175, 10.1126/science.adq0072

ECOLOGICAL SPECIATION
What's that you say?

When we think of a species adapting to environmental change, we mostly think about one trait. However, changes in one trait will likely affect others. In Galápagos finches, it has been shown that drought can lead to a change in bill size and shape in response to shifts in seed resources. However, birds do not just eat with their bills, they also sing with them. Podos *et al.* predicted how bill size would change in response to a series of droughts and forecasted the songs that they would produce. They found that, after a series of simulated droughts, the songs differed enough that territorial male birds no longer recognized them. —Sacha Vignieri

Science p. 211, 10.1126/science.adj4478

NEUROSCIENCE
Innervating by mitophagy

Mitophagy supports cell survival by clearing damaged mitochondria, and defects in mitophagy contribute to neurodegeneration. Abraham *et al.* found that mitophagy also mediates developmental axon guidance and maintenance (see the Focus by Davidson and Simon). Depriving cultured sensory neurons of nerve growth factor induced the expression of *Siah3* in the soma. Siah3 prevented mitophagy in axons by binding to Parkin, thereby promoting axon pruning. Mice lacking Siah3 exhibited increased innervation of the skin, suggesting that mitophagy regulates peripheral sensory innervation. —Leslie K. Ferrarelli

Sci. Signal. (2024)
10.1126/scisignal.adn5805;
10.1126/scisignal.ads1228

IN OTHER JOURNALS

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

STEM EDUCATION
Leveraging counter-spaces in physics

Counterspaces in STEM are spaces where individuals with a minoritized identity gather to share affinity and affirm their identities. Two counterspaces have emerged for undergraduate women in "Physics: Conferences for Undergraduate Women in Physics" (CUWiP) and "Women in Physics Groups" (WiPG). Hazari and Rodriguez used structural equation modeling to test the effect of participation in CUWiP and WiPG on undergraduate women's interest and sense of belonging in physics, and their results showed a significant positive direct effect. However, participation in CUWiP and WiPG was also associated

with increased beliefs in serious gender issues within physics, suggesting that discussions occurring within counterspaces may make some women more aware of deeper cultural issues. This in turn can negatively affect their sense of belonging if their interest in physics is not positively reinforced. —Melissa McCartney

J. Res. Sci. Teach. (2024)
10.1002/tea.21962

CATALYSIS
A platinum atom digs in

Rather than simply adsorb on a surface of α-phase iron oxide, platinum atoms alter the lattice and bond to surface oxygen atoms in an almost linear geometry. Rafsanjani-Abbasi *et al.* found that adsorbed platinum atoms

SEX
Life finds a way

Coral reefs support approximately 25% of all marine species on the ocean floor. Corals use sexual reproduction to generate genetic diversity, enabling them to adapt more quickly to changing environments. Research on the stony coral *Fimbriaphyllia ancora* revealed a unique reproductive strategy involving annual sex changes. Shikina *et al.* monitored 26 coral colonies in Nanwan Bay, Taiwan, over several years, finding that approximately 70% changed sex yearly. The process of sex change occurs 4 to 5 months after spawning for female-to-male transitions, whereas male-to-female changes happen within 0 to 3 months after sperm release. This finding enhances our understanding of coral reproductive biology and highlights the adaptive advantages of sex change as a strategy to increase the odds of successful propagation, and thus genetic diversity, of offspring. —Di Jiang

Stony coral (*Fimbriaphyllia ancora*) colonies have the ability to change sex annually, increasing their chances of successful reproduction.

EVOLUTION
Forestalling collapse

Tipping points mark changes in state of an environment or a population into another state with a different function. For example, water temperatures a few degrees above average can tip intact coral reefs into crumbling beds of algae. The susceptibility of microbial populations to climate change is a known cause for instability within ecological systems. With experiments and models, Blake *et al.* explored the resilience of stable states using populations of bacteria and fungi exposed to different ecological conditions. The authors found that if resistance evolved to stress that did not favor the growth rate of either partner, then greater population resilience resulted. —Caroline Ash

EARLY UNIVERSE
Too much ultraviolet light for reionization

Ultraviolet photons emitted by the first galaxies ionized the neutral gas in the intergalactic medium during the epoch of reionization. Observations have shown that high-redshift galaxies are bluer than expected, and thus emit more ultraviolet light, so more ultraviolet photons escape galaxies than was previously thought. Muñoz *et al.* calculated the effect that these have on the reionization process. They found that too many ultraviolet photons are produced, which would reionize the Universe much earlier than what has been observed. Combining three observational constraints on the reionization process, the authors showed that there is no part of the parameter space that satisfies all three constraints simultaneously. —Keith T. Smith

formed this site on the (1$\bar{1}$02)-(1 × 1) facet, which is the face commonly exposed on nanoparticles used as supports. Density function theory showed that this site was most stable despite breaking three iron-oxygen bonds. This undercoordination of platinum helps explain its high reactivity on this support. —Phil Szuromi

IMMUNE EVASION
Preserved strategies counteract immunity

The production of cyclic nucleotides within cells is an evolutionarily conserved signal that initiates immune defenses after viral infection. The bacteriophage-encoded Acb1 protein blocks these protective mechanisms in bacteria by degrading cyclic nucleotides. Hobbs *et al.* characterized the structure and function of a putative homolog of Acb1 from penguinpox virus. The homolog, renamed cGAMP phosphodiesterase, was similar in structure to Acb1 and shared a mechanism to capture and cleave a range of cyclic nucleotides. Moreover, cGAMP phosphodiesterase could replace the function of Acb1 in bacteriophages to block immune signals in bacteria. Proteins encoded by other eukaryotic viruses were also predicted to have a similar catalytic mechanism, suggesting a conserved approach to facilitating immune evasion. —Sarah H. Ross

SIGNAL TRANSDUCTION
Mast cells gobbling neutrophils

Investigating allergic responses in mouse skin, Mihlan et al. observed that neutrophils invaded degranulating mast cells but became trapped. The neutrophils stayed alive for hours, but then died, with their contents being incorporated by the mast cells. Neutrophil trapping appeared to confer the mast cells with neutrophilic proinflammatory functions, because proinflammatory molecules from the neutrophils could be released from the mast cells in response to immunoglobulin E. The physiological role of the process remains to be clarified, but it may contribute to chronic allergic inflammation and thus provide a

REVIEW SUMMARY

CLIMATE CHANGE

Environmental and societal consequences of winter ice loss from lakes

Stephanie E. Hampton*, Stephen M. Powers, Hilary A. Dugan, Lesley B. Knoll, Bailey C. McMeans, Michael F. Meyer, Catherine M. O'Reilly, Ted Ozersky, Sapna Sharma, David C. Barrett, Sudeep Chandra, Joachim Jansen, Ryan P. McClure, Milla Rautio, Gesa A. Weyhenmeyer, Xiao Yang

BACKGROUND: Lakes are rapidly losing ice in response to climate change. Most of the world's lakes freeze, with a median ice duration of 218 days. The rate of lake ice loss has markedly accelerated over the past 25 years, with ice melt in some regions across the Northern Hemisphere arriving 45 days per century earlier and with many lakes experiencing increased intermittency of ice cover during winter in addition to ice-free winters. Lake ice loss is expected to affect a substantial proportion of the world's population, who rely on these lakes for diverse needs, including drinking water, fisheries, transportation, and more. Until recently, both logistical challenges as well as misconceptions of winter as a time of quiescence resulted in limited winter research. Conspicuous decreases in ice cover and advances in technology have spurred rapid growth in winter research on seasonally ice-covered lakes. This Review is structured around a central question that winter researchers are frequently asked: Why does it matter that lakes are losing winter ice?

ADVANCES: Changes in ice cover have critical consequences for water quality, fisheries and biodiversity, weather and climate, as well as important cultural and socioeconomic activities. Socioeconomic benefits of freshwater ice include recreation, cultural identity, ice fishing, and ice roads for winter transportation. Ice roads across lakes and rivers are extensively used in oil and gas exploration, and they offer connectedness for remote communities during winter. Although the full socioeconomic importance of lake ice has not been accounted, illustrative examples exist. For Sweden's population of about 10.5 million people, ~$880 million US dollars are spent annually on ice fishing. Less easy to quantify is the extent to which interacting with lake ice contributes to cultural identity while shaping social cohesion and cultural heritage for the millions of people who live near a lake that seasonally freezes. Global data syntheses and models reveal further change. Shifts in ice quality, toward thinner and less stable white ice, have also contributed to increased fatal drowning events in areas of the world where ice was once safer for humans to traverse. Shorter ice duration has allowed lakes to warm faster, accelerating evaporative water loss globally and contributing to summer water quality issues, such as cyanobacteria blooms. Warmer water can favor invasive species and negatively affect cold-water fish and other organisms that are well adapted to ice-covered aquatic conditions. Ice seasonally creates a distinctive ecological niche space. Clear ice allows for high transmission of light that can fuel substantial algal growth, providing highly nutritious food for zooplankton grazers and their predators and generating food web effects that extend beyond winter. Ice provides a stable structure that effectively isolates lakes from the atmosphere and surrounding land, with several important implications: Ice cover can prevent shoreline erosion from wave action during the winter and reduce greenhouse gas emissions as well as evaporative water loss to the atmosphere. For lakes prone to creating lake-effect snow, ice cover prevents these extreme events from occurring in downwind communities.

OUTLOOK: Recent model projections predict complete loss of ice on thousands of lakes that historically experienced seasonal ice cover. Synthesis of historical ice records combined with global climate models suggests that up to 230,400 of the world's 1.4 million lakes larger than 0.1 km^2 will experience some years with no ice cover by 2080. Lakes at lower latitudes are vulnerable to losing 80% of ice days that are safe for humans to traverse. Anticipating the environmental and societal consequences of freshwater ice loss requires updated theory and models that consider the role of winter conditions and that incorporate data across the full annual cycle. Greater scientific and public understanding of the importance of lake ice for ecosystem health is key to supporting breakthrough science that informs sound stewardship of freshwater resources. ∎

Baikal seal pup on the ice of Lake Baikal. The Baikal seal (*Pusa sibirica*) is the world's only exclusively freshwater pinniped, and it is well adapted to a lake that freezes for half the year. It gives birth and raises pups on lake ice and maintains open holes in the ice to fish under ice through the winter.

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adl3211

PHOTO: CC BY-NC-SA 2.0 PACIFIC ENVIRONMENT

RESEARCH ARTICLE SUMMARY

FUNCTIONAL GENOMICS

Helicase-assisted continuous editing for programmable mutagenesis of endogenous genomes

Xi Dawn Chen†, Zeyu Chen†, George Wythes, Yifan Zhang, Benno C. Orr, Gary Sun, Yu-Kai Chao, Andrea Navarro Torres, Ka Thao, Mounica Vallurupalli, Jing Sun, Mehdi Borji, Emre Tkacik, Haiqi Chen, Bradley E. Bernstein*, Fei Chen*

INTRODUCTION: A fundamental challenge of genomics is to chart the impact of the three billion bases in the human genome on protein function and gene regulation. Thus, a critical goal is to develop strategies for mutagenizing genomic sequences systematically and at high throughput. In particular, targeted mutagenesis of single genomic loci could emulate the natural evolution process to reveal sequence-structure relationships, gain- and loss-of-function phenotypes, and cooperative mutations. However, no method exists that can perform continuous mutagenesis at targeted regions in the endogenous genomes of mammalian cells.

RATIONALE: We sought to develop a tool to perform targeted mutagenesis on the endogenous mammalian genome. Looking to nature, we observed that helicases are highly processive enzymes that can traverse large genomic regions. Some helicases, including those involved in DNA damage repair, can load and start unwinding DNA at single-stranded DNA regions in the genome. We reasoned that such helicases could be used for long-range targeted mutagenesis when fused to a deaminase enzyme. The fusion construct and its interval of hypermutation could then be programmably targeted, through single-guide RNAs (sgRNAs), to specific genomic regions using a Cas9 nickase. The directional and long-range DNA-unwinding event by the recruited helicase will then generate random mutations in the region.

RESULTS: We designed a platform called helicase-assisted continuous editing (HACE), which combines long-range editing of entire loci with the sequence programmability inherent to CRISPR gene editing tools. HACE uses CRISPR-Cas9 to direct the loading of a helicase-deaminase fusion for targeted hypermutation of the downstream genomic sequence. HACE achieved locus-specific deamination across >1000 nucleotides with mutations continuously accumulating over time. We further evaluated HACE prototypes incorporating diverse helicases, Cas9 variants, and deaminases, showing that they have tunable edit rates and ranges. We also showed that HACE can be multiplexed to target multiple genomic regions with a minimal number of guide RNAs. We then applied HACE in coding and noncoding genomic contexts to functionally dissect endogenous mutations conferring drug resistance, changes in enzymatic activity, and altered cis-regulatory element function. In the coding space, we identified variants that lead to mitogen-activated protein kinase kinase 1 (MEK1)–inhibitor drug resistance and also identified variants in SF3B1, a splicing factor, that lead to alternative 3′ splice-site usage. Turning to regulatory regions, we defined functional artificial variants in the enhancer regions of CD69 and pinpointed specific bases and motifs that mediate the impact of RUNX transcription factors on CD69 regulation. HACE solves two limitations faced by conventional base editing screens: the requirement of an NGG protospacer adjacent motif in the sgRNA recognition sequence and the occurrence of bystander mutations that can create artificial linkages and confound screening results. The long editing range of HACE can also uncover combinatorial effects and interactions between multiple distant mutations across a locus.

CONCLUSION: HACE makes possible the continuous, long-range, programmable diversification of endogenous mammalian genomes. We envision that HACE will substantially expand the functional genomics toolbox and enable the building of systematic sequence-function maps of both coding and noncoding genomes. Furthermore, HACE can be developed into a directed evolution system in the endogenous genome, enabling the selection of sequences for desired functions in mammalian biology. ∎

HACE: Long-range mutagenesis on endogenous DNA

① Cas9 nickase binds and exposes target

sgRNA recognition site

② HACE editor loads and performs editing over long range

Continuous evolution

Mutation / Time (days)
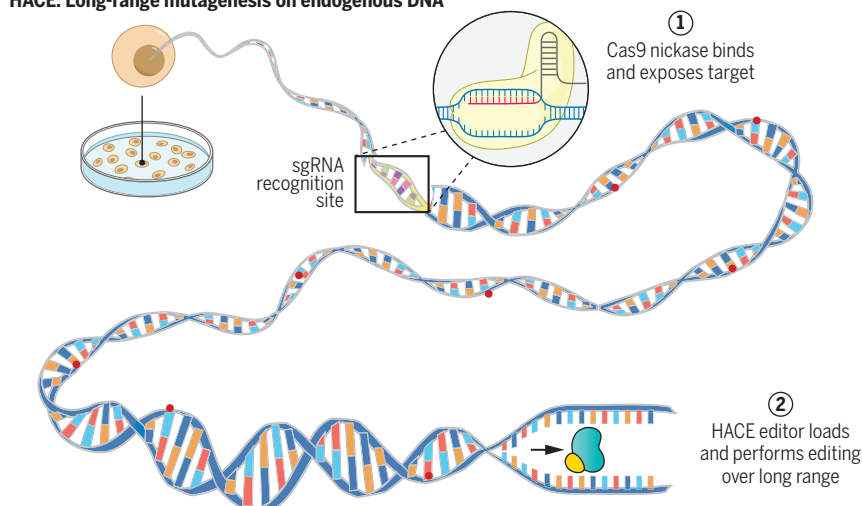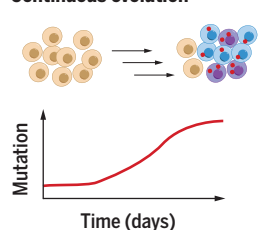
Coding variant function

Allele frequency (%)
>A mutation
>C
>G
>T
G128D
G202E
E203K
MEK1 CDS base position
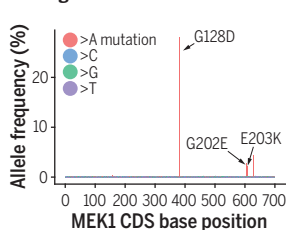
Cis-regulatory sequence function

Helicase editor
sgRNA
Gene expression
CD69 enhancer
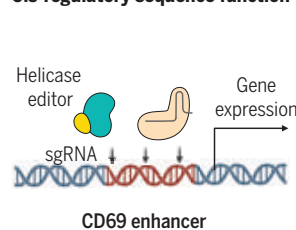
Development of a system to perform long-range targeted continuous mutagenesis of endogenous genomes. The helicase-assisted continuous editing (HACE) system allows for long-range mutagenesis on endogenous DNA and also continuous evolution over multiple cell generations. We have applied HACE to identify functional variants in both the coding and noncoding genome. CDS, coding sequence; D, Asp; E, Glu; G, Gly; K, Lys.

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adn5876

DEVELOPMENT

Temporal variability and cell mechanics control robustness in mammalian embryogenesis

Dimitri Fabrèges*, Bernat Corominas-Murtra*, Prachiti Moghe, Alison Kickuth†‡, Takafumi Ichikawa, Chizuru Iwatani, Tomoyuki Tsukiyama, Nathalie Daniel, Julie Gering, Anniek Stokkermans, Adrian Wolny, Anna Kreshuk, Véronique Duranthon, Virginie Uhlman§, Edouard Hannezo*, Takashi Hiiragi*

INTRODUCTION: Living organisms, despite their intrinsic variabilities and dynamics, establish robust forms and functions through complex interactions between molecules, cells, and tissues. Molecular and cellular variability has been considered detrimental and is actively minimized by regulatory mechanisms. However, recent studies suggest that variability plays a crucial role in cellular decision making and robustness, including in bacteria, cancer adaptation, tissue differentiation, and embryogenesis. Because early mammalian embryos exhibit variability in gene expression, cleavage timing, and various mechanical parameters, they offer an excellent model to study how variability affects developmental processes.

RATIONALE: Understanding the role of variability in development requires quantification in both space and time [three-dimensional (3D) plus time imaging], and experimental perturbation of variability. This study develops experimental and theoretical methods to quantify morphogenetic reproducibility in early mammalian embryogenesis. We characterized cleavage timing variability in mouse preimplantation embryos and compared it across different mammalian species. We also established a pipeline to segment cell membranes. Using exponential splines, we aimed to provide a rigorous and unbiased method for extracting geometrical information as well as the underlying topological structure of the embryos. This offers an exhaustive yet simple method to quantify and classify the complex dynamics of embryogenesis. Further, we developed an experimental system to examine how temporal variability influences the geometry and topology of embryos during early development. By examining the 3D geometrical and topological structures of cell packings, we sought to identify mechanisms driving the reduction of spatial variability.

RESULTS: The study found that variability in cleavage timing increases at a constant, species-specific rate during the first cleavages in mouse, rabbit, and monkey embryos. This temporal variability was not regulated and possibly contributes to the large morphological variability observed at the beginning of the 8-cell stage. Despite this, the embryos exhibited a notable spatial convergence at the end of that stage. We developed a computational pipeline based on dimensionality reduction and morphomaps to quantify in an unbiased manner the spatial dynamics of embryos, showing that their shape increases in similarity as development progresses. We show that these complex datasets can be interpreted intuitively by considering the topology of the cell packings, showing a convergence in shape mirroring a convergence toward a few rigid topologies. Our physical model indicated that surface energy minimization during compaction was sufficient to drive this convergence, promoting specific topological structures of lower energy and leading to reduced spatial variability. Additionally, we showed that although the zona pellucida encapsulating the embryo is not required for convergence, it reduces the initial set of possible topologies, thus allowing for faster convergence toward the most optimal configuration. Although the packing was maintained from one stage to the next with the natural desynchronization of cleavage timings, experimental synchronization at the 8- to 16-cell stage led to patterning defects in the blastocyst, highlighting the importance of temporal variability for proper development.

CONCLUSION: The findings reveal distinct mechanisms that lead to low spatial variability between different embryos. First, stochastic variability in cleavage timing plays a key role in achieving consistent developmental outcomes across species. Second, surface energy minimization and compaction select a few configurations with minimal energy and drive topological transitions toward these packings, leading to geometrical convergence and a reduction of spatial variability. A third mechanism may involve the zona pellucida constraining the range of potential structures of the early embryo packing. Altogether, these mechanisms ensure robustness in mammalian embryogenesis and present the potential adaptive benefits of stochastic processes in biological systems. ∎

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adh1145
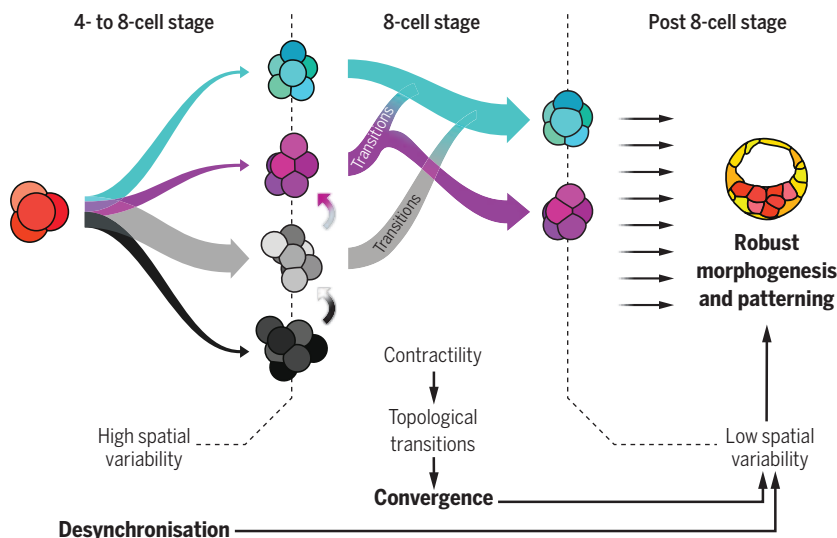
Model of robustness in embryogenesis from cell mechanics and desynchronization. 4-cell stage embryos give rise to many shapes at the beginning of the 8-cell stage, during which cell contractility triggers topological transitions. Ultimately, embryos are driven toward the most optimal packing (cyan). In parallel, the cell-autonomous desynchronization progressively increases temporal variability and helps to maintain topological optimality through generations, lowering spatial variability and promoting robustness.

RESEARCH ARTICLE SUMMARY

CANCER IMMUNOTHERAPY

Epigenetic regulators of clonal hematopoiesis control CD8 T cell stemness during immunotherapy

Tae Gun Kang, Xin Lan, Tian Mi, Hongfeng Chen, Shanta Alli, Song-Eun Lim, Sheetal Bhatara, Anoop Babu Vasandan, Grace Ward, Sofia Bentivegna, Josh Jang, Marianne L. Spatz, Jin-Hwan Han, Balthasar Clemens Schlotmann, Jakob Schmidt Jespersen, Christopher Derenzo, Peter Vogel, Jiyang Yu, Stephen Baylin, Peter Jones, Casey O'Connell, Kirsten Grønbæk, Ben Youngblood*, Caitlin C. Zebley*

INTRODUCTION: Sustained stimulation of CD8 T cells promotes the development of functional exhaustion, a terminal cell fate that has been broadly recognized to limit the durability of T cell–based immunotherapies. This state of exhaustion arises though a progressive transition of multipotent progenitor T cells (Tpex) into a terminally differentiated population (Tex) that has been defined as being nonresponsive to immune checkpoint blockade (ICB). The molecular determinants dictating the developmental transition are now being investigated as engineering targets to block the progression of T cell exhaustion and preserve the durability of immunotherapy approaches.

RATIONALE: We recently observed that the long-term survival of a small cohort of myelodysplastic syndrome (MDS) patients treated with anti–PD-L1 therapy was coupled to mutation of the ASXL1 gene among their T cells. Mutations in *ASXL1*, in addition to *DNMT3A* and *TET2*, are commonly associated with conferring a survival advantage among hematopoietic stem cells, resulting in subclonal outgrowth referred to as clonal hematopoiesis. Because of their general link to cellular stemness, we investigated the role of these regulators in the developmental transition of Tpex into Tex and the impact of deletion of these regulators on therapeutic durability during anti–PD-L1 treatment.
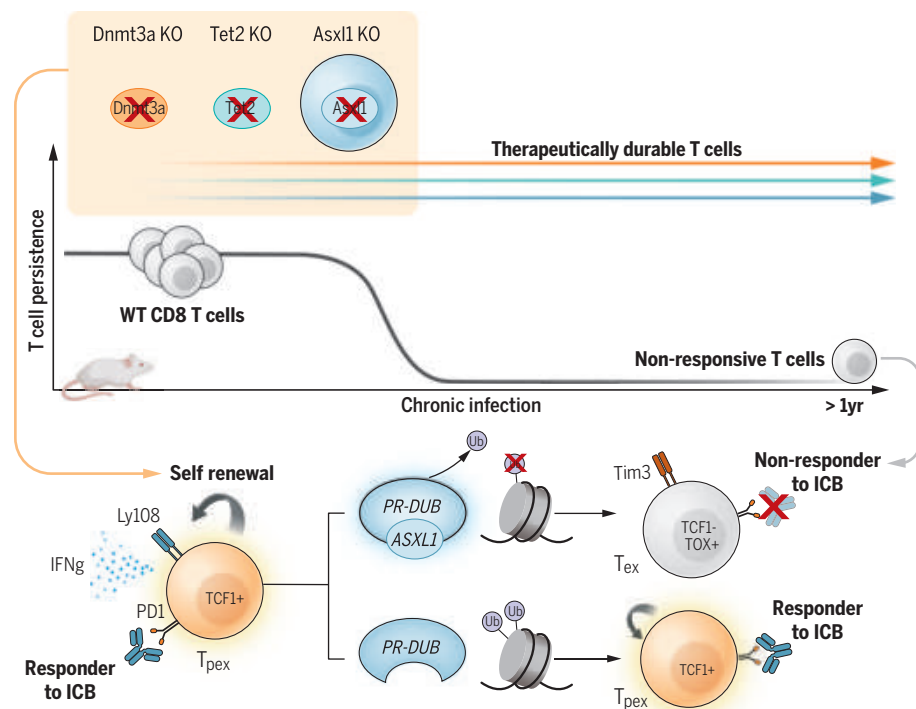
Core epigenetic regulators in maintaining therapeutic durable multipotent Tpex populations. Using the canonical murine model system of exhaustion, disruption of *Dnmt3a, Tet2*, and *Asxl1* in CD8 T cells preserved an ICB-responsive Tpex population for >1 year during chronic antigen exposure. Specific investigation into the new epigenetic regulator Asxl1 revealed that this unprecedented maintenance of Tpex is mediated by presence of monoubiquitinated histone 2A lysine 119 (H2AK119Ub) through disruption of the PR-DUB complex.

RESULTS: To investigate the role of Dnmt3a, Tet2, and Asxl1 in the development and maintenance of an ICB-responsive T cell population, we engineered T cells to contain mutations within each of these genes and then exposed them to a chronic source of antigen by adoptively transferring them into mice infected with the clone 13 strain of lymphocytic choriomeningitis virus (LCMV). Disruption of these regulators in a canonical model of T cell exhaustion allowed them to maintain their quantity and ICB-responsive state for >1 year during chronic antigen exposure. Despite being stimulated and undergoing antigen-driven proliferation for 1 year, these T cells did not exhibit any signs of malignant outgrowth. Moreover, the total pool of T cells containing each of these mutations was enriched for a large quantity of stem-like Tcf1+ T cells. Fate-tracking adoptive transfer experiments revealed that this Tcf1+ subset of ASXL1 KO T cells could both self-renew and give rise to potent effectors. Transcriptional and epigenetic profiling of the knockout (KO) T cells identified Asxl1 as a regulator of the polycomb group–repressive deubiquitinase (PR-DUB) complex, controlling the deubiquitination of histone 2A lysine 119, which serves as a molecular checkpoint for the developmental transition of Tpex into Tex. Extension of this mechanistic insight into tumor models revealed that adoptively transferred T cells with Asxl1 disruption resisted tumor microenvironment–induced exhaustion. Moreover, the adoptively transferred Asxl1 KO T cells synergized with anti–PD-L1 blockade, resulting in heightened tumor control. Finally, we show that Asxl1 disruption of tumor-specific T cells correlates with increased expansion of T cells in MDS patients after treatment with anti–PD-L1.

CONCLUSION: Results from these studies show that DNMT3A, TET2, and ASXL1 control the developmental checkpoint between Tpex and terminally exhausted T cells. Specific investigation into Asxl1 revealed its role in T cell differentiation through epigenetic modification of the PR-DUB pathway. These data articulate a reverse-translation approach for defining molecular mechanisms that restrict maintenance of the ICB-responsive stem-like T cells within a suppressive tumor microenvironment. Through our mechanistic investigation of genes associated with patient survival, we have established a scientific rationale for specific genetic modification of discrete epigenetic regulators for future engineering to enhance the durability of cancer immunotherapy. ∎

CARBON CYCLE

Hidden comet tails of marine snow impede ocean-based carbon sequestration

Rahul Chajwa, Eliott Flaum, Kay D. Bidle, Benjamin Van Mooy, Manu Prakash*

INTRODUCTION: Phytoplankton in the upper layer of the ocean agglomerates and sinks under gravity, giving rise to a natural carbon transport mechanism termed "biological pump." The perpetual shower of soft and fragile marine snow in the ocean is estimated to be annually sequestering 2 to 4.5 billion tons of carbon from the atmosphere into the abyss, regulating both the atmospheric carbon dioxideand the sustenance of marine ecosystems. A predictive underpinning of marine snow is thus crucial. However, we currently lack a quantitative microphysics-based framework for the formation, sedimentation, and remineralization of marine snow, leading to major uncertainties in the current carbon flux estimates in climate models.

RATIONALE: Because sedimentation physics is at the heart of marine snow phenomena, we took an observation-driven approach to addressing this problem. Based on the 19th-century paradigm of Stokes' law and its ad hoc generalizations, researchers had been seeking a universal trend in how size is related to sinking speed in marine snow. Because marine snow is a structurally complex soft matter that deforms while sinking under its own weight, it violates key assumptions underlying Stokes' law and presents a classic two-way fluid-structure coupling that remained unexplored. To directly investigate the sinking dynamics of this complex object, we organized an ocean expedition (Cruise ID: EN667) during an algal bloom in the Gulf of Maine (42.5°N, 69.5°W). Marine snow aggregates were collected through freely hanging sediment traps in multiple deployments at a depth of 80 m. To directly observe sedimentation dynamics of these aggregates at sea, we utilized a new scale-free vertical tracking microscope mounted on a two-axis gimbal that minimized mechanical noise from ship's motion and allowed us to track small aggregate (equivalent spherical diameter < 750 μm) sinking over long times. To visualize the flow around sinking aggregates, we used a tracer bead (diameter 700 nm to 2 μm) solution.

RESULTS: By directly measuring the sinking velocities and detailed flows around individual marine snow particles, we discovered a new morphological feature in marine snow: a physical invisible comet tail forming a halo around a visible particulate matter during sedimentation. These hitherto-unseen comet tails are made of viscoelastic transparent exopolymer, which fundamentally modifies the sinking behavior. Our observations guided a new theoretical framework based on Stokesian sedimentation in which we included this previously invisible degree of freedom and constructed a reduced-order model for these compound particles. Furthermore, the combination of field experiments and theory enabled a sedimentation-based measurement of the elastic response of the mucus. We corroborated these findings with three-dimensional volumetric imaging of marine snow particles, which illuminate the heterogeneous microstructure of marine snow.

CONCLUSION: Through detailed analysis of more than 100 marine snow aggregates studied individually, we discovered hidden comet tails that effectively act as "physical" drag lines on sinking marine snow. This mucus-induced impedance almost doubles the estimate of mean residence time of marine snow in the Euphotic zone, nearly halting some particles to a standstill. This suggests a substantial overestimate in the flux inferred by using only the visible size of a marine snow particle. The discovery of multiphase nature of marine snow and a new conceptual framework that incorporates the invisible degrees of freedom in the sedimentation dynamics lays the foundation for understanding the formation, sedimentation, and remineralization of marine snow in the purview of physics. The crucial role of viscoelasticity of marine mucus as one of the knobs of carbon flux opens rich possibilities for studying biological origin of mucus and its complex rheology in the open oceans and potential biogeoengineering remediation. ∎

S **READ THE FULL ARTICLE AT**
https://doi.org/10.1126/science.adl5767
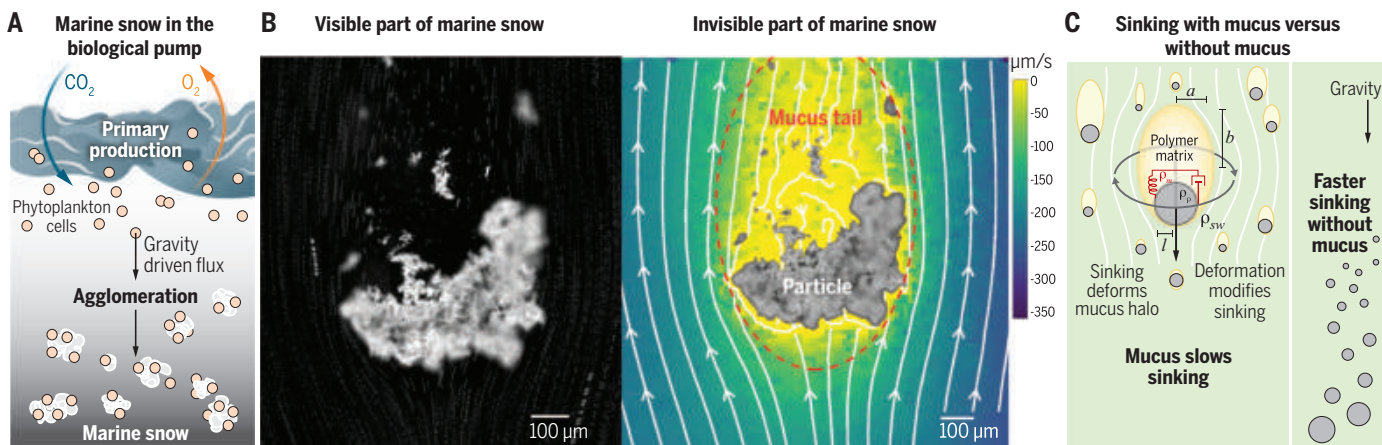
Hidden comet tails of marine snow. (**A**) A simplified depiction of carbon sequestration in the biological pump through marine snow. (**B**) Experimental data: (Left) Image of sinking marine snow visualized with tracer beads in the background and (right) fluid flow corresponding to the same particle showing the invisible mucus tail (yellow region) that falls along with the particle, greatly increasing the particle's effective size. (**C**) Impact of mucus on sedimentation: Mucus greatly increases the time marine snow can spend in the upper layers of the ocean, presenting a natural knob in this carbon flux. ρ_m, mucus density; ρ_{sw}, sea water density; ρ_p, particulate density; a, semiminor axis of the mucus comet tail; b, semimajor axis of the mucus comet tail; l, size of the visible aggregate.

RESEARCH ARTICLES

ORGANIC CHEMISTRY

Catalytic prenyl conjugate additions for synthesis of enantiomerically enriched PPAPs

Shawn Ng[1], Casey Howshall[1], Thanh Nhat Ho[1], Binh Khanh Mai[2], Yuebiao Zhou[1], Can Qin[3], Kai Ze Tee[1], Peng Liu[2]*, Filippo Romiti[1,3,4]*, Amir H. Hoveyda[1,3]*

Polycyclic polyprenylated acylphloroglucinols (PPAPs) are a class of >400 natural products with a broad spectrum of bioactivity, ranging from antidepressant and antimicrobial to anti-obesity and anticancer activity. Here, we present a scalable, regio-, site-, and enantioselective catalytic method for synthesis of cyclic β-prenyl ketones, compounds that can be used for efficient syntheses of many PPAPs in high enantiomeric purity. The transformation is prenyl conjugate addition to cyclic β-ketoesters promoted by a readily accessible chiral copper catalyst and involving an easy-to-prepare and isolable organoborate reagent. Reactions reach completion in just a few minutes at room temperature. The importance of this advance is highlighted by the enantioselective preparation of intermediates previously used to generate racemic PPAPs. We also present the enantioselective synthesis of nemorosonol (14 steps, 20% yield) and its one-step conversion to another PPAP, garcibracteatone (52% yield).

C ertain catalytic reactions can transform the way a family of bioactive natural products (NPs) is prepared in high enantiomeric purity, an attribute that has therapeutic relevance. A case in point is the impact of catalytic enantioselective aldol addition on synthesis of polyketides and related entities (*1*, *2*). Polycyclic polyprenylated acylphloroglucinols (PPAPs) (Fig. 1, A and B) are another important set of >400 NPs (*3–9*). Different members of the class express distinct therapeutic activity, ranging from anticancer, anti-HIV, antiepileptic, and anti-inflammatory activity to antimicrobial, anti-Alzheimer's, antidepressant, and anti-obesity activity. One PPAP, hyperforin, is the active component of St. John's wort (*Hypericum perforatum*), which is used for treating mild depression as well as skin injuries and neuralgia (*3*, *10*). PPAPs contain several prenyl groups, some situated two sp³-hybridized carbons away from a carbonyl unit or a secondary or tertiary alcohol. Catalytic enantioselective reactions that forge such carbon–prenyl bonds in high yield would render enantiomerically enriched PPAPs more readily accessible. And yet processes of this type remain unknown. A direct approach would entail catalytic conjugate addition (CA) of a prenyl group to a cyclic α,β-unsaturated carbonyl compound. This presents several challenges: the high reactivity, size, and dissymmetry of a metal–

prenyl complex can lead to low yields and minimal regio- and enantioselectivities.

Most PPAPs, nemorosone being one (Fig. 1A), contain a bicyclo[3.3.1]nonane framework (*3–6*). Their biosynthesis is presumed to proceed through ionic intermediates (*11*) involving the enantiotopic prenyl moiety (see **I**, shown in maroon) of a monocyclic polyprenylated acylphloroglucinol (MPAP). Reactions of **I** with prenyl pyrophosphate, leading to nemorosone and clusianone, are representative (Fig. 1A). The formation of another PPAP subgroup, a member of which is garcibracteatone (Fig. 1B), has been proposed to proceed through carbon radical intermediates. George and co-workers have shown that single-electron oxidation (*12*) of an MPAP [see (–)-**1**] containing a modified prenyl-containing *C5* substituent (in blue) affords *ent*-garcibracteatone and 5-*epi*-garcibracteatone, probably through **II** and *epi-ent*-**II**, in 14 and 8% yield, respectively. Although H-atom abstraction with **II** and *epi-ent*-**II** would deliver anti-microbial (–)-nemorosonol (*13*) and 5-*epi*-(+)-nemorosonol, respectively, the PPAP and its isomer were not formed. The above principles and findings indicate that (i) enantiomeric polycyclic scaffolds can arise from reaction at an enantio- or diastereotopic MPAP substituent (see **I** and **1**) and (ii) an enantiomerically enriched PPAP may be converted to others of the same class.

Several PPAPs occur in both enantiomeric forms (for instance, nemorosonol), further underscoring the need for enantioselective strategies. Nonetheless, most reported syntheses deliver racemic products (*6*). Among the limited number of enantioselective approaches, less than a handful involve a catalytic process (*14–16*). In one, the enantiomerically enriched intermediate was prepared after a minimally

diastereoselective sequence of six steps in 10% yield (*14*). Another route afforded an epimer of the NP [3-*epi*-hypatulin B (*16*)]. Preparation of the substrate needed for the catalytic enantioselective reaction can require up to six steps. Use of a chiral auxiliary (*17*, *18*) or an enantiomerically enriched alkylating agent (*19*, *20*) suffers from similar shortcomings. In one instance, nine steps were needed to convert an auxiliary fragment, necessary for obtaining high e.r. (enantiomeric ratio), to a prenyl group (*15*).

A key recent advance is the development of a programmable strategy outlined by Maimone and colleagues (*21*) for concise synthesis of bicyclic PPAPs that have a bicyclo[3.3.1]nonane skeleton. The pivotal early intermediate in these routes is a β-prenyl cyclopentanone. The utility of the approach was highlighted by a total synthesis of racemic hyperforin (*22*) (Fig. 1C). While the opening step was a conjugate addition, it was not of a prenyl unit but rather of a 2-methyl-1-butenyl moiety (in blue). The butenyl fragment represented a protected form of the prenyl group, probably as a countermeasure to known chemo- and regioselectivity complications (1,2- versus 1,4- and α- versus γ-addition, respectively). The CA product was isolated as an enol silane. Subsequent conversion to the metal enolate and alkylation (in maroon) were followed by unmasking the prenyl moiety through alkene isomerization under acidic conditions and at elevated temperatures. This string of transformations, which proved to be the least efficient of the total synthesis, afforded *rac*-**2** in ~30% yield [for two steps; 75:25 d.r. (diastereomeric ratio)]. A second prenyl group was installed by alkylation with prenyl bromide (in brown), generating *rac*-**3**. Further modification furnished *rac*-**4**, setting the stage for an oxidative rearrangement that unveiled the bicyclo[3.3.1]nonane framework (*rac*-**5**). Racemic hyperforin was secured after several more operations. An analogous tactic was later on applied to synthesis of garsubellin A (*21*), again as a racemate.

Maimone and co-workers' strategies (Fig. 1C) underscore the need for a catalytic enantioselective reaction that converts a β-ketoester (such as **6**) to the corresponding β-prenyl ketone [for example, (**S**)-**prod**; Fig. 1D]. The same compounds can be used for enantioselective preparation of other PPAP frameworks, such as those in nemorosonol and garcibracteatone (Fig. 1B). A β-prenyl ketoester can be modified regio- and/or diastereoselectively in several ways. With an allyl ester (R = allyl), alkylation with iodomethane followed by catalytic decarboxylative allylation, based on the Stoltz modification of the Tsuji-Trost process (*23*), would furnish **ket-1** in high d.r. and e.r. As another example, **ket-2**, a more versatile variant of a compound formerly employed in synthesis of the bicyclo[3.3.1]nonane core (racemic) (*24*),

[1]Department of Chemistry, Merkert Chemistry Center, Boston College, Chestnut Hill, MA 02467, USA. [2]Department of Chemistry, University of Pittsburgh, Pittsburgh, PA 15260, USA. [3]Supramolecular Science and Engineering Institute, University of Strasbourg, 67000 Strasbourg, France. [4]Department of Chemistry and Biochemistry, University of Texas at Dallas, Richardson, TX 75080, USA.
*Corresponding author: Email: pengliu@pitt.edu (P.L.); filippo.romiti@utdallas.edu (F.R.); amir.hoveyda@bc.edu (A.H.H.)

A *Biosynthesis pathway via ionic intermediates*

B *Synthesis of Garcibracteatone by George and co-workers*

C *Synthesis of rac-Hyperforin by Maimone and co-workers*

D *Initial plans for development of a catalytic EPCA*
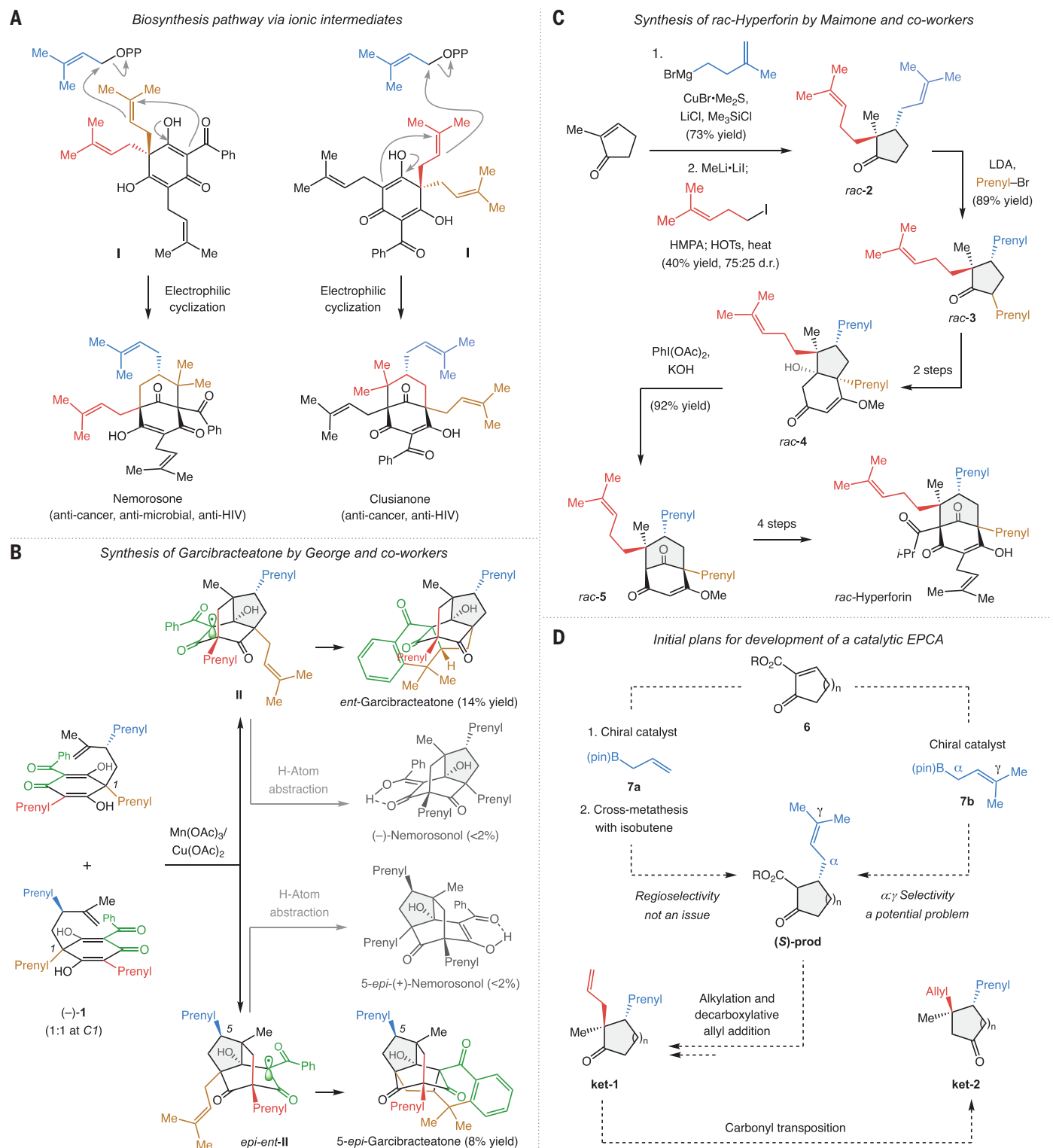
Fig. 1. PPAPs, a programmable strategy for their preparation, and a relevant catalytic enantioselective process. (**A**) PPAPs are a large family of polyprenylated natural products likely generated in nature by an intramolecular ionic cyclization through an MPAP (such as **I**). (**B**) George and co-workers (*12*) have suggested that a PPAP may be converted to another by a radical pathway. (**C**) A programmable strategy for synthesis of PPAPs is founded on the use of a cyclic β-prenyl ketone (for instance, *rac*-**4**), but a method for enantioselective preparation of the core molecule is lacking. (**D**) The envisioned catalytic enantioselective conjugate additions for accessing PPAPs in high enantiomeric purity. Me, methyl; Ph, phenyl; HIV, human immunodeficiency virus; OPP, pyrophosphate; HMPA, hexamethylphosphoramide; HOTs, *p*-toluenesulfonic acid; OMe, methoxy; Ts, *p*-tolylsulfonate; R, an organic moiety.

A *Initial studies with allyl– and prenyl–B(pin)*

6a
(in 2 steps, 85% yield,
multi-decagram scale)

7a G = H
7b G = Me
(2.0 equiv.)

6.0 mol % ligand,
5.0 mol % CuCl,

2.0 equiv. NaOEt,
1.0 equiv. EtOH,
THF, 1 h, 22 °C

8a G = H
8b G = Me

imid-1

bisphos

7a → 8a		7b → 8b	
imid-1	**bisphos**	**imid-1**	**bisphos**
>98% conv., 15% CA, 12% yield, 57:43 e.r.	>98% conv., <5% CA	>98% conv., <5% CA	>98% conv., <5% CA

B *Factors indicating **7c** is a superior prenyl–Cu precursor*

7b ---- LCuOEt

7c

more steric strain

less steric strain?

i

rev-prenyl–Cu

prenyl–Cu

ii

(pin)BOEt

(pin)BOEt

6a
+
7c
(2.0 equiv.)

6.0 mol % **imid**,
5.0 mol % CuCl,

2.0 equiv. NaOEt,
1.0 equiv. EtOH,
THF, 1 h, 22 °C

8b + **9**

imid-1
>98% conv., 15–20% CA, 75:25 **8b:9**, yield and e.r. not determined

imid 2–5

imid-2 (Ar = Mes)
>98% conv., 95:5 **8b:9**,
48% yield (**8b** + **9**), 73:27 e.r. (**8b**)

imid-3 (Ar = 2,6-Et$_2$C$_6$H$_3$)
>98% conv., 92:8 **8b:9**,
44% yield (**8b** + **9**), 79:21 e.r. (**8b**)

imid-4 (Ar = 9-Anthracenyl)
>98% conv., 92:8 **8b:9**,
63% yield (**8b** + **9**), 85:15 e.r. (**8b**)

imid-5 (Ar = 3,5-Me$_2$C$_6$H$_3$)
>98% conv., 97:3 **8b:9**,
62% yield (**8b** + **9**), 91:9 e.r. (**8b**)

C *Studies leading to development of rev-prenyl borate*

Key problem:
Complete enone decomposition in <10 min when exposed to a metal alkoxide

7d

LCuOEt

least steric strain?

iii

→ **prenyl–Cu**

(pin)BOEt
+ NaOEt

7c

1.0 equiv. NaOEt,

no EtOH or
10 mol % EtOH,
THF, 22 °C

7d
no EtOH: 2 h, 88% conv.
with EtOH:
15 min, >98% conv.

sample of **7d**
(~50 g)

6a
+
7d
(2.0 equiv.)

6.0 mol % **imid-5**,
5.0 mol % CuCl,

12 mol % NaOEt,
THF, 15 min, 22 °C

8b + **9**

>98% conv., 97:3 **8b:9**,
82% yield (**8b** + **9**),
90:10 e.r. (**8b**)

D *Catalytic EPCA with various cyclic substrates*

10
with **imid-3** as the ligand:
>98% conv., 97:3 γ:α,
74% yield, 94:6 e.r.

11
with **imid-3** as the ligand:
>98% conv., >98:2 γ:α,
74% yield, 94.5:5.5 e.r.

12
with **imid-5** as the ligand:
>98% conv., >98:2 γ:α,
78% yield, 98.5:1.5 e.r.

13
with **imid-4** as the ligand:
>98% conv., 93:7 γ:α,
72% yield, 95:5 e.r.

Fig. 2. Searching for the optimal conditions for EPCA reactions. (A) Initial explorations involved allyl– and prenyl–B(pin) reagents, but these were inefficient and minimally enantioselective. The low conversions to the desired product in the case of the more widely used **7a** was particularly surprising. **(B)** Reverse-prenyl–B(pin) (**7c**) was used, as we reasoned that reaction through **ii** would be more favored (versus **i**). Results with **7c** were encouraging, and ligand screening led us to determine that **imid-5** is the optimal ligand. **(C)** It was envisioned that formation of **prenyl–Cu** might entail a less congested noncyclic transition state (through **iii**). In light of the deleterious effect of NaOEt on CA efficiency (it likely catalyzes substrate decomposition), we used borate **7d**, which led to formation of **8b** in higher efficiency. **(D)** Electrophiles with other ring sizes are similarly suitable substrates. See the supplementary materials for details. EtO, ethoxy; (pin)BOEt, (pinacolato)ethoxyboronate; THF, tetrahydrofuran; Mes, 2,4,6-trimethylphenyl; L, ligand.

would become accessible through carbonyl transposition (*25*). With an effective catalytic enantioselective prenyl conjugate addition (EPCA) to cyclic β-ketoesters, it would become possible to synthesize many PPAPs and their skeletal analogs in high enantiomeric purity,

irrespective of biosynthetic origin or scaffold type.

Assorted catalytic enantioselective CA (or 1,4-addition) reactions have been developed (*26*), but those that involve an allylic moiety are rare. There are only two reported cases

that involve a cyclic enone. One relates to addition of an allyl moiety to a cyclic enone (*27*). It was shown that treatment of **6** (R = Me, *n* = 1 to 3) with 10 mol % of a bis(oxazoline)–Cu Lewis acid catalyst and trimethylallylsilane (the Hosomi-Sakurai process) affords cyclic β-allyl

A *Energetics of EPCA involving different organoboron compounds and rationale for enantioselectivity*

B *Origin of low e.r. in allyl CA*

C *Influence of the size of the ester moiety on enantioselectivity*

R = Me (**8c**)
>98% conv., 96.5:3.5 γ:α,
79% yield (γ + α),
72:28 e.r. (γ)

R = *i*-Pr (**8d**)
>98% conv., 96:4 γ:α,
79% yield (γ + α),
96:4 e.r. (γ)

R = *t*-Bu (**8e**)
>98% conv., 92:8 γ:α,
80% yield (γ + α),
95:5 e.r. (γ)
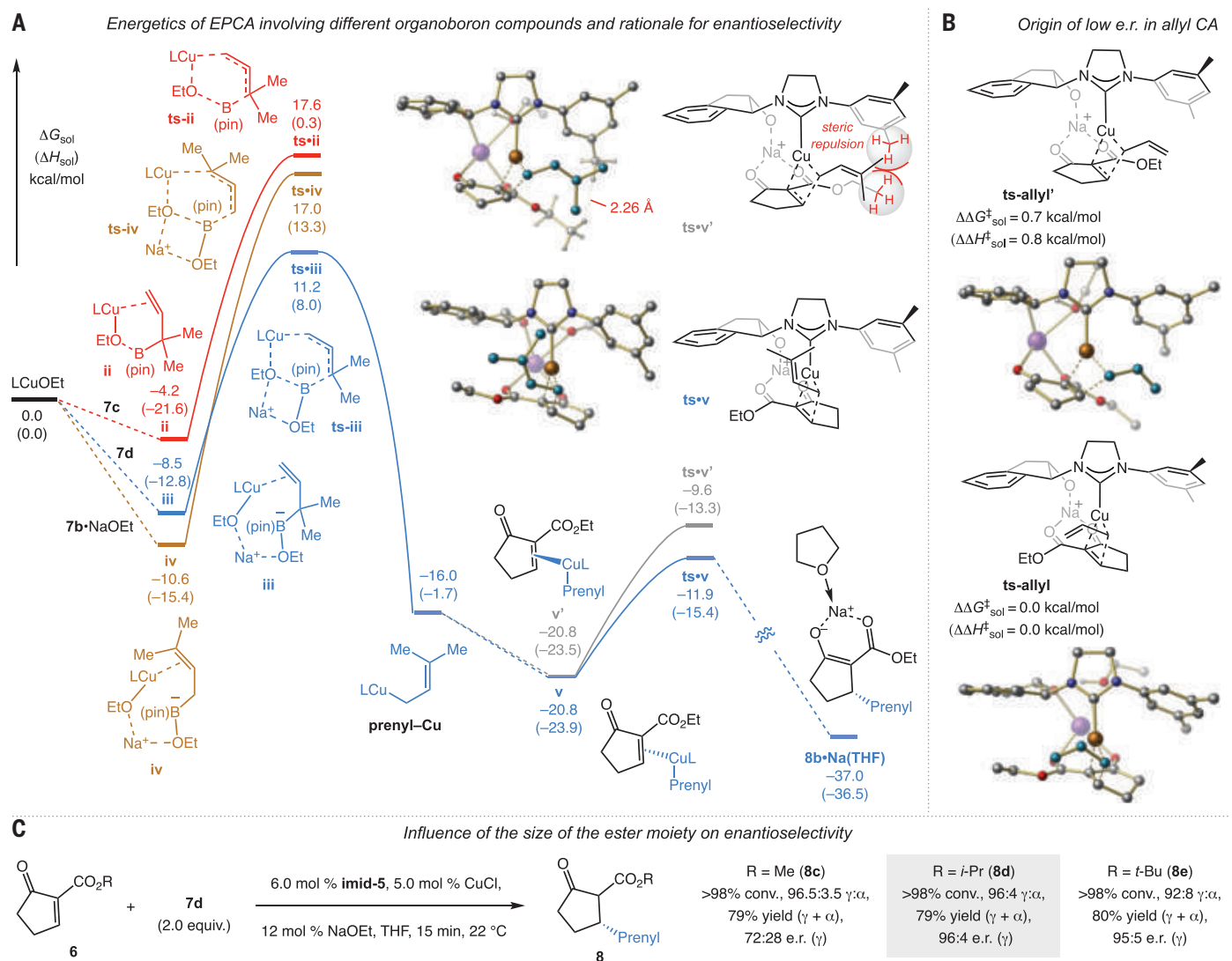
Fig. 3. Computational studies provide a rationale for obtaining higher yield and enantioselectivity. (**A**) Transfer of a prenyl ligand from an alkoxy borate is faster than a neutral one, and there is greater rate increase with the borate derived from reverse-prenyl–B(pin) (**7d**; via **iii**, in blue). The complex formed by association of the substrate with **prenyl–Cu** (**v**) is likely transformed to **8b•Na(THF)** by a lower energy route (via **ts•v** versus via **ts•v′**). (**B**) Computational studies indicate that without the methyl substituents in an allyl–Cu complex (versus prenyl–Cu), the energy gap between **ts-allyl** and **ts-allyl′** is much smaller, leading to low e.r. (**C**) Better mechanistic understanding led to higher e.r. Computational studies performed at the DLPNO-CCSD(T)/def2-TZVP/SMD(THF)//B3LYP-D3/6-31G(d)–SDD level. See the supplementary materials for details. L, ligand derived from **imid-5**; ts, transition state.

ketones. Enantioselectivities varied, depending on the ring size, with 85:15 e.r. obtained for a cyclopentenone, perhaps the most relevant system for PPAP synthesis; additionally, cryogenic conditions (−50°C or lower) and up to 48 hours were needed. A later study involved a chiral Lewis acid catalyst derived from Yb(OTf)$_3$ (OTf, trifluoromethanesulfonate) and a chiral *N,N*′-dioxide in combination with Cu(OTf)$_2$ (10 mol % of each); products were isolated in high enantiomeric purity, but the approach is confined to ester-substituted coumarins and tetraallylstannane (*28*).

A CA involving an allyl– or prenyl–copper species would probably proceed through a substrate-catalyst complex wherein the enone's

alkene is more closely associated with the metal complex (*29*) (compared to a carbonyl-bound chiral Lewis acid). It would then be easier to obtain high stereochemical induction under ambient conditions and more quickly. We entertained two options: (i) a two-step catalytic protocol entailing enantioselective CA with allyl–B(pin) (**7a**; pin, pinacolato) followed by cross-metathesis with isobutene and (ii) direct synthesis of β-prenyl ketones by α-selective and enantioselective CA with dissymmetric prenyl–B(pin) (**7b**), which is also purchasable. We initially viewed the latter as being more likely to be problematic because of the documented site- and regioselectivity issues (*30–33*).

Catalytic allyl conjugate addition

We began by evaluating the feasibility of catalytic CA of allyl–B(pin) **7a** (Fig. 2A). Various chiral N-heterocyclic carbenes (NHCs) and bisphosphines were examined (for example, **imid-1** and **bisphos**). Typically, **6a** was completely consumed, but we could rarely detect >10% **8a**, and when we did, the e.r. was negligible [for example, >98% conversion (conv.), 15% CA, 12% yield, 57:43 e.r. with **imid-1**]. Additional studies revealed that the in situ formed Cu-allyl complex is mostly consumed by side reactions. Propene formation was observed (see the supplementary materials for details), implying that the copper–allyl species reacts faster with an alcohol (needed to accelerate catalyst release)
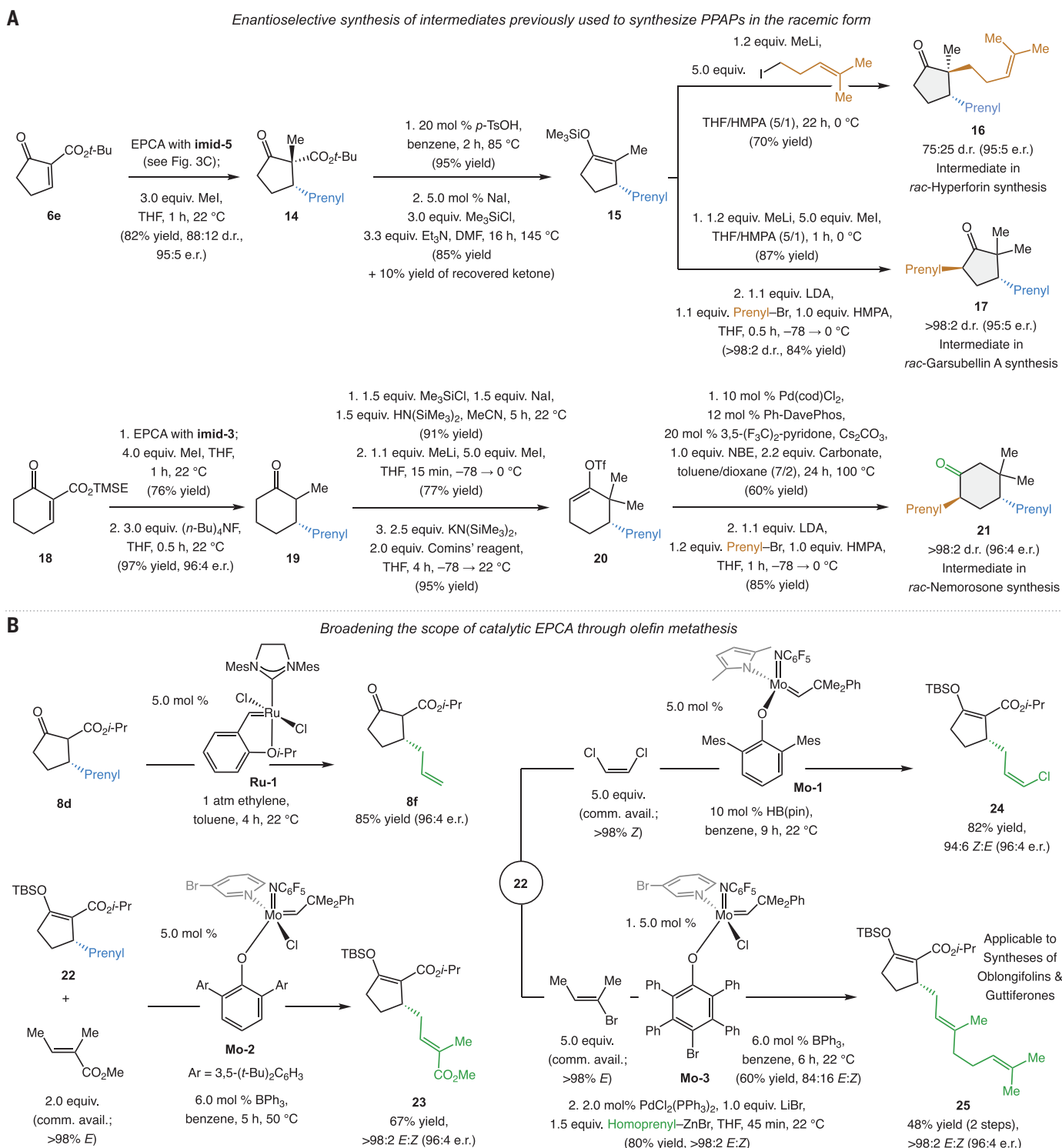
Fig. 4. Access to β-prenyl ketones that can be used for enantioselective total syntheses of PPAPs. (**A**) Several EPCA products were transformed to enantiomerically enriched intermediates used in preparation of racemic PPAPs. (**B**) Enantiomerically enriched β-prenyl ketones can be easily converted to other desirable compounds through catalytic cross-metathesis. See the supplementary materials for details. DMF, dimethylformamide; Comins'

reagent, *N*-(5-chloro-2-pyridyl)bis(trifluoromethanesulfonamide); cod, cyclooctadiene; Ph-DavePhos, 2-diphenylphosphino-2'-(*N,N'*-dimethylamino) biphenyl; NBE, azetidine-1-yl((1*S*,4*R*)-bicyclo[2.2.1]hept-2-en-2-yl)methane or azetidine-amide-substituted norbornene; Carbonate, 4,4-dimethylpentan-3-yl piperidine-1-yl carbonate; TMSE, (trimethylsilyl)ethyl; OTf, triflate; TBS, *t*-butyldimethylsilyl.

A

Enantioselective total synthesis of Nemorosonol

B

Conversion of Nemorosonol to Garcibracteatone
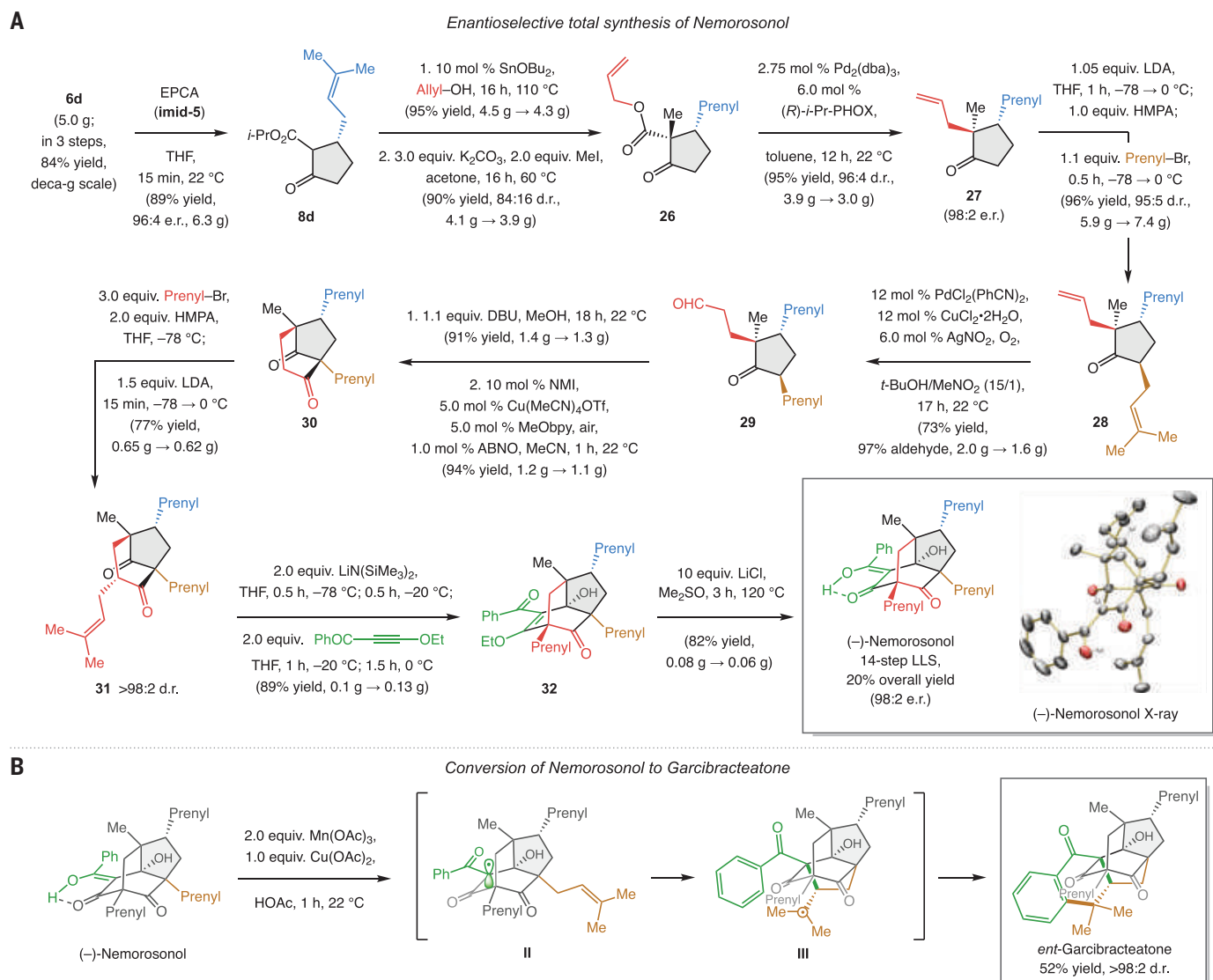
Fig. 5. Enantioselective total synthesis of nemorosonol and garcibracteatone. (**A**) Catalytic EPCA enables the enantioselective and most efficient total synthesis of (−)-nemorosonol. (**B**) The availability of enantiomerically enriched nemorosonol makes it possible to investigate its conversion to other PPAPs.

The transformation to garcibracteatone constitutes the most efficient total synthesis of this NP to date. See the supplementary materials for details. DBU, 1,8-diazabicyclo[5.4.0]undec-7-ene; NMI, *N*-methylimidazole; MeObpy, 4,4′-dimethoxybipyridine; ABNO, 9-azabicyclo[3.3.1]nonane *N*-oxyl; Ac, acyl.

---

compared to an enone. Allyl–B(pin) has been used in enantioselective additions to other electrophiles (for example, phosphinoylimines) with a copper-based catalyst and in the presence of an alcohol (*34*), indicating that reactions involving cyclic β-ketoesters are particularly challenging.

Enantioselective prenyl conjugate additions

With the first option nullified, we turned to the possibility of generating the β-prenyl ketone product directly by using prenyl–B(pin) (**7b**), remaining mindful of the documented complications vis-à-vis α:γ and 1,4:1,2 selectivity (*30–33*). Our hope was that the larger prenyl moiety would cause greater transition state energy differences and that enantioselectivity

would improve. At first, the situation appeared worse, as with **7b** there was no productive transformation regardless of the chiral ligand used (>98% conv., <5% conv. to CA product). We thought that perhaps the problem was the steric pressure associated with the formation of a prenyl–Cu complex (**i**; Fig. 2B), which led us to opt for reverse-prenyl–B(pin) **7c**. We reasoned that conformational restriction arising from the Thorpe-Ingold effect and diminished steric pressure in **ii** could facilitate Cu–C bond formation.

With the combination of **imid-1** and **7c** (Fig. 2B) there was moderate γ-selectivity (15 to 20% conv., 75:25 **8b:9**). Encouraged, we revisited the ligand screening studies, finding that with **imid-2**, a conformationally restricted rendering

of **imid-1**, EPCA is more efficient, allowing us to isolate **8b** in 48% yield, 95:5 γ:α selectivity (**8b:9**), and 73:27 e.r. (for **8b**). Further probing showed that **imid-3** is similarly effective but less γ selective (92:8 γ:α). With **imid-4**, bearing a 9-anthracenyl fragment, efficiency and e.r. improved further [92:8 **8b:9**, 63% yield (**8b** + **9**), 85:15 e.r. (**8b**)] but not γ selectivity. The complex derived from **imid-5** delivered the highest e.r. and γ:α ratio (91:9 e.r. and 97:3 **8b:9**).

Analysis of product mixtures indicated that the major issue was the rapid rate at which **6a** decomposes upon exposure to a metal alkoxide (5 to 15 min, 22°C). We attributed this to enolization and/or oligomerization initiated by alkoxy CA and Michael addition to another

enoate. If so, the use of preformed borate (**7d**; Fig. 2C) would obviate the need for excess metal alkoxide (*35*) and minimize side product formation. Borate **7d** could be directly converted to **prenyl–Cu** through **iii**, a complex that would be less sterically demanding than **i** or **ii**.

We prepared borate **7d** (Fig. 2C) by treating of **7c** for 15 min at ambient temperature with an equivalent of sodium ethoxide (NaOEt) and 10 mol % ethanol (EtOH; to increase the solubility of the metal alkoxide). The EPCA was also complete after just 15 min at the same temperature, and **8b** was obtained with similar regio- and enantioselectivity as before (97:3 **8b:9**, 90:10 e.r. for **8b**). However, the yield improved to 82% (versus 62% yield with **7c**). By adopting the foregoing conditions (Fig. 2D) and with **imid-3**, we prepared cyclohexyl **10** and lactone **11** in 74% yield (in both cases), 97:3 and >98:2 γ:α ratio, and 94:6 and 94.5:5.5 e.r., respectively. With **imid-5** and **imid-4**, we synthesized cycloheptyl **12** and cyclooctyl **13** in 78 and 72% yield, >98:2 and 93:7 γ:α ratio, and 98.5:1.5 and 95:5 e.r., respectively. Borate **7d** can be used as the initially generated solution or isolated and stored for later use without diminution in efficiency or selectivity.

The high regio- and enantioselectivity (93:7 to >98:2 γ:α and 94:6 to 98.5:1.5 e.r.) for six- to eight-membered ring enones was good news. Ironically, however, the e.r. for cyclopentyl **8b**, one of the more important substrates in regard to enantioselective PPAP synthesis, was the lowest (90:10 e.r.). As the first step toward addressing this issue, we turned to computational studies (Fig. 3). We also wanted to probe the validity of the foregoing hypotheses vis-à-vis the organoboron reagent that can transfer a prenyl group to generate a corresponding copper complex.

Mechanistic studies

The initial question was in regard to relative favorability of the routes that generate **prenyl–Cu** via various prenyl– and reverse-prenyl–B (pin) reagents. The results of computational studies (Fig. 3A) indicated that the copper complex originating from ethoxy-borate of reverse-prenyl–B(pin) **7d** (in blue), namely **iii**, is more stable than that derived from reverse-prenyl–B(pin) **7c** (**ii**, in red) but higher in energy than the complex generated from alkoxy-borate of prenyl–B(pin) **7b** (**iv**, in brown). Nevertheless, and more importantly, the activation barrier for **iii** converting to **prenyl–Cu** is lower compared to the complexes derived from **7c/ii** or **7b/iv** ($\Delta\Delta G^{\ddagger}$ = 19.7 kcal/mol for **7d/iii**, 27.6 kcal/mol for **7b/iv**, and 21.8 kcal/mol for **7c/ii**); this is consistent with the observed efficiency trend.

Coordination of **prenyl–Cu** with a substrate can lead to two diastereomeric complexes. Enone-derived **v** can undergo CA via **ts•v**

to furnish **8b•Na(THF)**, which is converted to **8b** after workup. This implies that the energy gap between **ts•v** (Fig. 3A, in blue) and **ts•v'** (in gray) arises from greater steric pressure in the latter, likely engendered by the propinquity of the ligand's N-aryl ring and the prenyl group's *E*-methyl group. As a counter measure, in **ts•v'**, the ligand's N-aryl ring can rotate, but this would bring its methyl unit closer to the substrate's alkyl ester fragment. These destabilizing interactions are absent in a Cu–allyl complex (Fig. 3B), and **ts•allyl** and **ts•allyl'** are consequently more similar in energy (0.7 kcal/mol difference), resulting in lower e.r.

The perceived influence of steric repulsion between the chiral ligand and the carboxylic ester substituent suggested that bulkier ligand substituents and/or ester moiety could lead to more effective stereocontrol. Increasing the size of the substituents on the NHC's N-aryl group did not improve selectivity (see the supplementary materials for details); we attributed this to more facile rotation around the C–N bond linkage. There was a diminution in e.r. for methyl ester **8c** (72:28 e.r., Fig. 3C; compared to 90:10 e.r. with ethyl ester **8b**). In contrast, and as foreshadowed by the computational work, enantioselectivity improved for *i*-propyl and *t*-butyl esters **8d** and **8e** (Fig. 3C; 96:4 and 95:5 e.r., 79 and 80% yield, respectively).

PPAP precursors in high enantiomeric purity

As noted, EPCA offers efficient access to enantiomerically enriched samples of intermediates formerly converted to racemic PPAPs. An illustrative case relates to synthesis of **16** (Fig. 4A), utilized before in a synthesis of racemic hyperforin (*22*). *t*-Butyl ester **14** (95:5 e.r.) was converted to enolsilane **15**, and the ensuing alkylation afforded multifunctional cyclopentanone **16** (70% yield). Analogously, **15** was transformed to bis-prenyl cyclopentanone **17** (two steps, 95:5 e.r., 75% yield), the racemic variant of which has been transformed to garsubellin A (*21*). The d.r. values for **16** and **17** (75:25 and >98:2, respectively) are the same as those in the prior disclosures. Equally important, applicability of the method extends to enantiomerically enriched β-prenyl cyclohexanones, such as that obtained from EPCA to **18**. To effect ketone transposition, we adopted the recent protocol introduced by Dong and colleagues (*25*). Thus, **19** was first converted to enol triflate **20**; net ketone isomerization and diastereoselective incorporation of the second prenyl unit afforded **21**, a known precursor to nemorosone (*36, 37*), in >98:2 d.r. and 96:4 e.r.

As mentioned earlier, CA with allyl–B(pin) is minimally enantioselective (Fig. 2). However, a cyclic β-allyl ketone can be easily obtained by catalytic ethenolysis (*38*) of an EPCA product, as indicated by the conversion of **8d** to **8f** (Fig.

4B). Therefore, the more effective way of synthesizing an enantiomerically enriched β-allyl ketone turns out to be by ethenolysis of an enantiomerically enriched β-prenyl ketone and not, as we had in the beginning envisioned, by cross-metathesis involving isobutene and a β-allyl ketone. High e.r. values can be obtained much more easily through EPCA.

We have shown previously that a prenyl moiety is optimal for stereoretentive cross-metathesis (*39*), meaning that an EPCA product can be transformed to other alkenes that are stereochemically defined and diversifiable. This advantage is underscored by the conversion of **22** to *E*-enoate **23** (67% yield, >98:2 *E:Z*) and *Z*-alkenyl chloride **24** (82% yield, 94:6 *Z:E*), respectively. Stereoretentive cross-metathesis with purchasable *Z*-2-bromo-2-butene (*40*) followed by modification by cross-coupling afforded β-geranyl cyclopentanone **25** in 48% yield (two steps), >98:2 *E:Z*, and 96:4 e.r. Although the cross-metathesis furnished the alkenyl bromide with 84:16 *E:Z* ratio, the desired and major *E*-alkene underwent cross-coupling faster, allowing us to obtain **25** as a single isomer. Cyclic β-geranyl ketones may be used for enantioselective synthesis of bioactive oblongifolins (*6, 41*) and guttiferones (*42*), NPs that, in addition to prenyl groups, also contain one or more geranyl side chains.

Enantioselective synthesis of nemorosonol

To probe the utility of the approach further, we chose to apply it to enantioselective synthesis of nemorosonol (Fig. 5A), a PPAP recently prepared in the racemic form by Kigoshi and co-workers in 0.7% overall yield (*43*). A more efficient sequence and access to workable amounts of enantiomerically enriched nemorosonol would facilitate biological testing and analog preparation. What is more, it would then be possible to explore the possibility of converting enantiomerically enriched nemorosonol to other PPAPs.

To initiate the total synthesis, we converted 5.0 g of **6d** (prepared on decagram scale in three steps and 84% yield) to 6.3 g of **8d** (89% yield, 96:4 e.r.). At this point, we needed to find a way to form the quaternary carbon stereogenic center adjacent to the carbonyl group in high d.r. (see **27**). Judging from previous work (for instance, see formation of *rac-***2** in Fig. 1C), we surmised that trapping of the enolate generated by EPCA would likely not be diastereoselective. Instead, we opted for decarboxylation catalyzed by an enantiomerically pure Pd complex (*23*). We converted **8d** to allyl ester **26** in two steps: formation of the allyl ester (*44*) and alkylation with iodomethane (84:16 d.r., inconsequential to the total synthesis). To excise the carboxyl group (*45*), we subjected **26** to 2.75 mol % Pd$_2$(dba)$_3$ (dba, dibenzylideneacetone) and 6.0 mol % (*R*)-*i*-Pr-PHOX (PHOX, phosphinooxazoline), both purchasable, in

toluene for 12 hours at ambient temperature. We isolated **27** in 95% yield (3.0 g; 98:2 e.r.) and 94:6 d.r. With (S)-i-Pr-PHOX, presumably the mismatched ligand isomer, the yield and stereoselectivity were markedly lower (60% yield; 36.5:63.5 d.r.).

Next, we prepared **28** (95:5 d.r., 96% yield) by reacting the lithium enolate derived from **27** with prenyl bromide. Oxidation according to the Grubbs modification of the Wacker method (46) afforded **29** with 97:3 aldehyde: ketone selectivity (73% yield). Intramolecular aldol addition of keto-aldehyde **29** delivered the desired bicyclic β-hydroxyketone (91% yield), which by catalytic aerobic oxidation, based on the procedure outlined by Steves and Stahl (47), was converted to diketone **30** (94% yield). Prenylation of **30** proved to be more complicated, owing to instability of the metal enolate (for example, C–C bond rupture accompanied by ketene cyclopentyl enolate formation). In the reported nemorosonol synthesis (43), a related transformation with allyl bromide did not yield any of the desired product. We addressed this problem by simultaneously subjecting **30** to LDA (lithium diisopropylamide) and prenyl bromide. We were thus able to trap the sensitive enolate before it could decompose and secure tri-prenyl diketone **31** in 77% yield (0.62 g from 0.65 g of **30**; >98:2 d.r.).

The final three steps in the published synthesis of racemic nemorosonol (43), comprising sequential Michael and aldol addition, acidic hydrolysis, and multiple cross-metatheses with 2-methylpropene to convert three allyl units to their corresponding prenyl groups, were reported to proceed in 10% yield. The yield for each step was not provided, but, per precedent, it is unlikely that the cross-metathesis was an issue. For a more efficient endgame, we needed to find a more efficient way of converting the β-ethoxy ketone to the corresponding hydroxy derivative. After some experimentation, we found that a Michael addition/aldol cyclization sequence, based on the protocol reported by Kigoshi and co-workers (43), can be used to obtain **32** in 89% yield (0.1 g **31** → 0.13 g **32**). Next, instead of subjecting **32** to HCl at 70°C, we treated it with excess LiCl in dimethylsulfoxide at 120°C, which allowed us to secure (–)-nemorosonol in 82% yield and 98:2 e.r. While similar in length to the reported racemic synthesis (43) [14-step LLS (longest linear sequence) from cyclopentenone **6d**], the present sequence delivers the NP in high enantiomeric purity (98:2 e.r.) and is substantially more efficient (20% compared with 0.7% overall yield).

Nemorosonol to garcibracteatone

Having access to (–)-nemorosonol meant that we could explore the possibility of converting the synthetic NP to garcibracteatone, a compound found to be cytotoxic against KB cell lines (48). This was an idea inspired by the aforementioned proposal by George and co-workers (12) regarding the involvement of radical species **II** and its conversion to nemorosonol by hydrogen atom abstraction, a process that was not observed (Fig. 1B). We thought that it might be possible to transform nemorosonol to garcibracteatone through **II** by means of an aryl radical/aromatization sequence. This indeed turned out to be the case. Subjection of synthetic (–)-nemorosonol to the same conditions as previously reported (12) delivered ent-garcibracteatone in 52% yield, likely through **II** and **III** (Fig. 5B).

Outlook

The catalytic EPCA processes and the attendant strategies described here, together with the programmable approach of Maimone and colleagues (21), make it possible to prepare PPAPs efficiently and in high enantiomeric purity. Organoboron compounds are readily available, easy to use, robust, functional-group tolerant, and can be utilized in sundry Cu-catalyzed regio-, diastereo-, and enantioselective C–C bond–generating reactions (49). The role of the organo-borate compound in modulating solution basicity as well as being a superior reagent for transferring an allylic moiety meant that products could be isolated in higher yields regardless of the scale. Another attractive feature is ease of preparation and reactivity/robustness balance offered by the borate species. We expect that these positive attributes will influence future research involving boron-based reagents.

Considering the breadth of bioactivity exhibited by PPAPs, catalytic EPCA is likely to facilitate future advances in drug development. Particularly noteworthy is that certain PPAPs, such as nemorosonol, can now be prepared in either naturally occurring enantiomeric form so that biological activity of these compounds can be calibrated more easily and thoroughly. The brevity and efficiency of the nemorosonol total synthesis should empower researchers to synthesize and investigate various peripherally as well as skeletally modified analogs, which may in turn be transformed to the remodeled garcibracteatone derivatives. Preliminary studies in regard to anticancer activity of (–)-nemorosonol and several of its precisely altered skeletal analogs and garcibracteatone measured against 10 cancer cell lines are detailed in the supplementary materials.

REFERENCES AND NOTES

1. B. Schetter, R. Mahrwald, Angew. Chem. Int. Ed. **45**, 7506–7525 (2006).
2. T. P. Stockdale, N. Y. S. Lam, M. J. Anketell, I. Paterson, Bull. Chem. Soc. Jpn. **94**, 713–731 (2021).
3. R. Ciochina, R. B. Grossman, Chem. Rev. **106**, 3963–3986 (2006).
4. J. A. Richard, R. H. Pouwer, D. Y.-K. Chen, Angew. Chem. Int. Ed. **51**, 4536–4561 (2012).
5. C. Guttroff et al., Angew. Chem. Int. Ed. **56**, 15852–15856 (2017).
6. X.-W. Yang, R. B. Grossman, G. Xu, Chem. Rev. **118**, 3508–3558 (2018).
7. A. L. Piccinelli et al., Tetrahedron **61**, 8206–8211 (2005).
8. S. Chen et al., Cell Metab. **33**, 565–580.e7 (2021).
9. O. Cuesta-Rubio, L. Monzote, R. Fernández-Acosta, G. L. Pardo-Andreu, L. Rastrelli, Phytochemistry **210**, 113674 (2023).
10. Y. Cui, Y. Zheng, Neuropsychiatr. Dis. Treat. **12**, 1715–1723 (2016).
11. P. Adam, D. Arigoni, A. Bacher, W. Eisenreich, J. Med. Chem. **45**, 4786–4793 (2002).
12. H. P. Pepper, S. J. Tulip, Y. Nakano, J. H. George, J. Org. Chem. **79**, 2564–2573 (2014).
13. A. Oya et al., J. Nat. Prod. **78**, 258–264 (2015).
14. F. Horeischi, C. Guttroff, B. Plietker, Chem. Commun. **51**, 2259–2261 (2015).
15. Y. Shimizu, S.-L. Shi, H. Usuda, M. Kanai, M. Shibasaki, Angew. Chem. Int. Ed. **49**, 1103–1106 (2010).
16. S. Leisering et al., Org. Lett. **24**, 4305–4309 (2022).
17. M. R. Garnsey, D. Lim, J. M. Yost, D. M. Coltart, Org. Lett. **12**, 5234–5237 (2010).
18. D. Jang, M. Choi, J. Chen, C. Lee, Angew. Chem. Int. Ed. **60**, 22735–22739 (2021).
19. B. A. Sparling, D. C. Moebius, M. D. Shair, J. Am. Chem. Soc. **135**, 644–647 (2013).
20. S. Wen, J. H. Boyce, S. K. Kandappa, J. Sivaguru, J. A. Porco Jr., J. Am. Chem. Soc. **141**, 11315–11321 (2019).
21. X. Shen, C. P. Ting, G. Xu, T. J. Maimone, Nat. Commun. **11**, 508 (2020).
22. C. P. Ting, T. J. Maimone, J. Am. Chem. Soc. **137**, 10516–10519 (2015).
23. O. Pàmies et al., Chem. Rev. **121**, 4373–4505 (2021).
24. S. J. Spessard, B. M. Stoltz, Org. Lett. **4**, 1943–1946 (2002).
25. Z. Wu, X. Xu, J. Wang, G. Dong, Science **374**, 734–740 (2021).
26. A. Alexakis, N. Krause, S. Woodward, in Copper-Catalyzed Asymmetric Synthesis, A. Alexakis, N. Krause, S. Woodward, Eds. (Wiley–VCH, 2008), pp. 33–68.
27. M. Shizuka, M. L. Snapper, Angew. Chem. Int. Ed. **47**, 5049–5051 (2008).
28. Y. Kuang et al., Org. Lett. **13**, 3814–3817 (2011).
29. N. Yoshikai, E. Nakamura, Chem. Rev. **112**, 2339–2372 (2012).
30. B. H. Lipshutz, C. Hackmann, J. Org. Chem. **59**, 7437–7444 (1994).
31. A. Yanagisawa, S. Habaue, K. Yasue, H. Yamamoto, J. Am. Chem. Soc. **116**, 6130–6141 (1994).
32. L.-M. Zhao, S.-Q. Zhang, F. Dou, R. Sun, Org. Lett. **15**, 5154–5157 (2013).
33. M. Berger, D. Carboni, P. Melchiorre, Angew. Chem. Int. Ed. **60**, 26373–26377 (2021).
34. E. M. Vieira, M. L. Snapper, A. H. Hoveyda, J. Am. Chem. Soc. **133**, 3332–3335 (2011).
35. L. R. Mills et al., Angew. Chem. Int. Ed. **62**, e202313848 (2023).
36. N. M. Ahmad, V. Rodeschini, N. S. Simpkins, S. E. Ward, A. J. Blake, J. Org. Chem. **72**, 4803–4815 (2007).
37. N. S. Simpkins, J. D. Taylor, M. D. Weller, C. J. Hayes, Synlett **2010**, 639–643 (2010).
38. J. Bidange, C. Fischmeister, C. Bruneau, Chemistry **22**, 12226–12244 (2016).
39. A. H. Hoveyda et al., Acc. Chem. Res. **56**, 2426–2446 (2023).
40. T. Koengeter et al., J. Am. Chem. Soc. **145**, 3774–3785 (2023).
41. W. Hamed et al., J. Nat. Prod. **69**, 774–777 (2006).
42. K. R. Gustafson et al., Tetrahedron **48**, 10093–10102 (1992).
43. K. Mitsugi, T. Takabayashi, T. Ohyoshi, H. Kigoshi, Org. Lett. **24**, 4635–4639 (2022).
44. P. Baumhof, R. Mazitschek, A. Giannis, Angew. Chem. Int. Ed. **40**, 3672–3674 (2001).
45. D. C. Behenna et al., Chemistry **17**, 14199–14223 (2011).
46. Z. K. Wickens, B. Morandi, R. H. Grubbs, Angew. Chem. Int. Ed. **52**, 11257–11260 (2013).
47. J. E. Steves, S. S. Stahl, J. Am. Chem. Soc. **135**, 15742–15745 (2013).
48. O. Thoison et al., Tetrahedron **61**, 8529–8535 (2005).
49. A. H. Hoveyda et al., Angew. Chem. Int. Ed. **59**, 21304–21359 (2020).

EPIDEMIOLOGY

Predicting pathogen mutual invasibility and co-circulation

Sang Woo Park[1,2]*, Sarah Cobey[2], C. Jessica E. Metcalf[1], Jonathan M. Levine[1], Bryan T. Grenfell[1]*

Observations of pathogen community structure provide evidence for both the coexistence and replacement of related strains. Despite many studies of specific host-pathogen systems, a unifying framework for predicting the outcomes of interactions among pathogens has remained elusive. We address this gap by developing a pathogen invasion theory (PIT) based on modern ecological coexistence theory and testing the resulting framework against empirical systems. Across major human pathogens, PIT predicts near-universal mutual susceptibility of one strain to invasion by another strain. However, predicting co-circulation from mutual invasion also depends on the degree to which susceptible abundance is reduced below the invasion threshold by overcompensatory epidemic dynamics, and the time it takes for susceptibles to replenish. The transmission advantage of an invading strain and the strength and duration of immunity are key determinants of susceptible dynamics. PIT unifies existing ideas about pathogen co-circulation, offering a quantitative framework for predicting the emergence of novel pathogen strains.

Understanding the factors that allow the coexistence or exclusion of pathogen variants is a fundamental challenge in the dynamics and control of infectious diseases (*1–4*). Progress requires better understanding of how competition occurs by means of the host immune response and its implications for the invasion of new strains. Despite many studies of specific host-pathogen interactions (*5–10*), no unifying framework exists to explain invasion dynamics and subsequent competitive outcomes across pathogen systems. The continued emergence of novel pathogens and ongoing vaccine development against antigenically variable threats emphasize the importance of this question (*11*).

A major challenge in developing a unifying framework for pathogen community structure is explaining the heterogeneity of observed competitive outcomes across pathogens. Specifically, understanding what mechanisms promote stable coexistence among some path-

ogen variants (Fig. 1, A to C) but not others (Fig. 1, D to E). For example, notwithstanding heterogeneity in the observed dynamics, stable coexistence in large host populations is maintained between respiratory syncytial viruses (RSV) A and B (Fig. 1A), Coxsackievirus A16, Enterovirus A71 (Fig. 1B), and four dengue serotypes (Fig. 1C). On the other hand, the evolutionary dynamics of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants (Fig. 1D) and seasonal influenza strains of the same subtype (Fig. 1E) exemplify strain replacement patterns, in which the emergence of antigenically distinct variants causes extinction of previously dominant variants (*5, 7, 12*).

To explain these heterogeneous outcomes, we present a unifying framework for predicting whether a pathogen variant can invade a host population in the presence of their competitor. We show that predicting co-circulation from invasion depends on changes in the susceptible pool following invasion. Specifically, the invading pathogen will continue to infect susceptible hosts even as a given epidemic declines past its peak, driving the susceptible pool below the invasion threshold. Therefore, the pathogen overcompensates by using up

more susceptible resources than it requires and can prevent itself and its competitor from reinvading until the susceptible pool is replenished. We quantify this overcompensation using two measures: maximum reduction in the susceptible pool, and the time it takes for the susceptible pool to replenish to permit reinvasion. In doing so, we explore how the transmission advantage of the invading variant and underlying immune structures affect this overcompensatory dynamics.

Pathogen invasion theory

Many models have been proposed to explain observed patterns of competition for specific pathogen systems (*1, 11, 12, 13–19*) and to identify the mechanisms that dictate persistence and coexistence following the invasion of new variants (*7, 20*). However, these models rely on detailed assumptions, often specific to the pathogen, making it difficult to draw general conclusions about pathogen competition across different systems. So far, there is no single overarching framework that allows quantitative comparisons of coexistence potential across different pathogen systems.

To unify heterogeneous patterns of pathogen competition, we have developed the pathogen invasion theory (PIT), inspired by Chesson's modern coexistence theory (MCT) from community ecology [see the supplementary materials (SM) section S1] (*21*). Specifically, PIT allows us to simplify any system of two competing pathogen variants i and j into two quantities that determine their invasibility (SM sections S2 and S3). First, the immunological niche difference $1-\rho$ describes the ability of a variant to escape the host's immune response induced by another variant. This is measured as a ratio of equilibrium values (or long-term averages) of geometric mean of intraspecific susceptibility ($\sqrt{S_{i|i}S_{j|j}}$) versus interspecific susceptibility ($\sqrt{S_{j|i}S_{i|j}}$) at the population level:

$$1-\rho = 1 - \sqrt{\frac{S_{i|i}S_{j|j}}{S_{j|i}S_{i|j}}}, \quad (1)$$

[1]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08540, USA. [2]Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637, USA.
*Corresponding author. Email: swp2@princeton.edu (S.W.P.); grenfell@princeton.edu (B.T.G.)
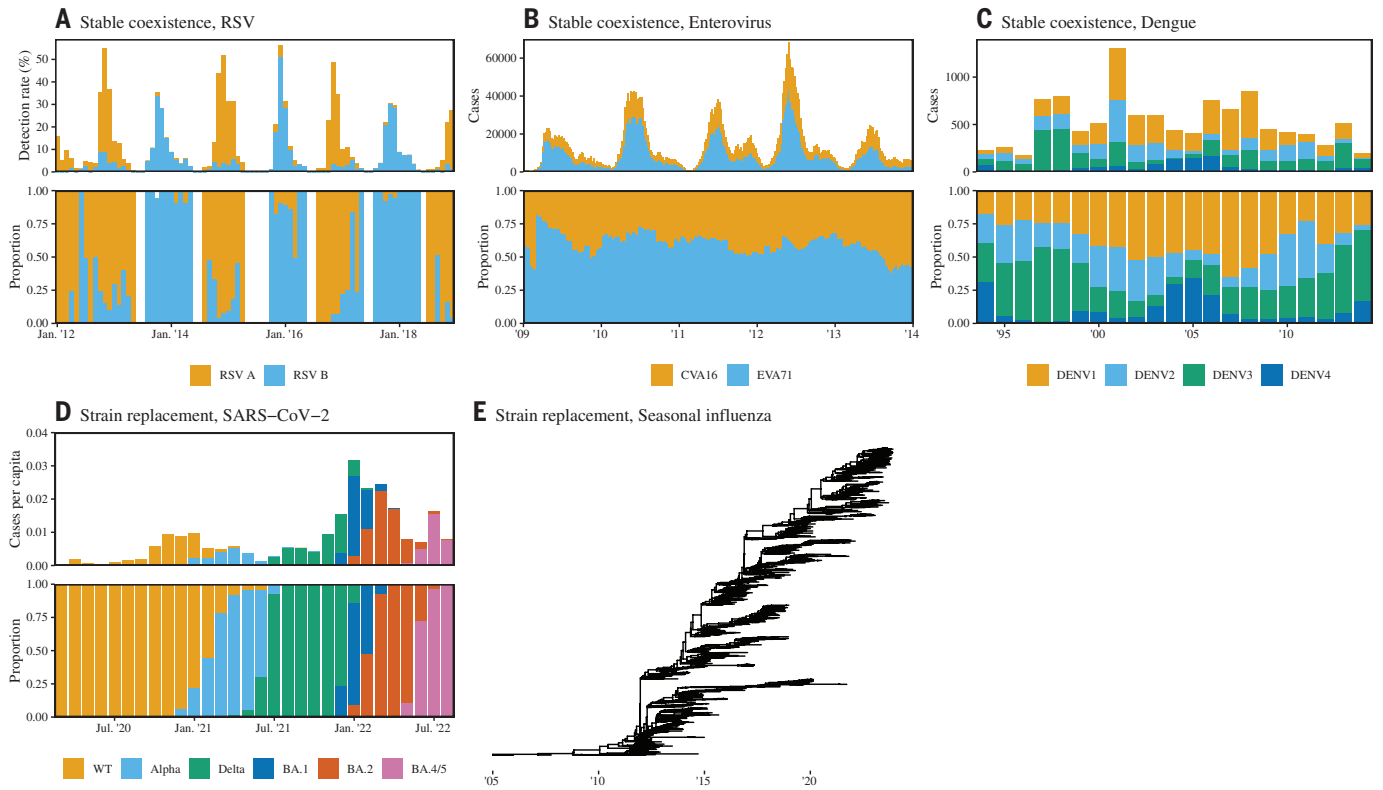
Fig. 1. Observed coexistence patterns of human pathogens. (**A**) Monthly detection rates of RSV A and B (top) and their relative proportions (bottom) from Cheonan, Korea (*41*). (**B**) Weekly cases of hand, foot, and mouth disease caused by CV-A16 and EV-A71 (top) and their relative proportions (bottom) from all 31 provinces of mainland China (*42*). (**C**) Annual cases of four dengue serotypes (top) and their relative proportions (bottom) from Bangkok, Thailand (*43*). (**D**) Monthly per capita cases of SARS-CoV-2 variants (top) and their relative proportions (bottom) from 10 representative countries (*12*). (**E**) Phylogenetic tree of 1578 human influenza A H3N2 subtype genome sampled between 2005 and 2023, publicly available on https://nextstrain.org/. All data are taken from publicly available sources from the above references.

where $S_{j|i}$ represents the effective proportion of individuals who are susceptible to variant j. when variant i is at equilibrium; note that equilibrium conditions arise when the amount of susceptible individuals is balanced with the basic reproduction number: $\mathcal{R}_{0,i}S_{i|i=1}$. Second, the fitness difference, f_i / f_j, measures the ratio of each variant's innate ability to spread ($\mathcal{R}_{0,i}$ and $\mathcal{R}_{0,j}$), discounted by its susceptibility to the host immune response to the other variant:

$$\frac{f_i}{f_j} = \sqrt{\frac{\mathcal{R}_{0,i}S_{i|j}}{\mathcal{R}_{0,j}S_{j|i}}} \qquad (2)$$

Then, similar to MCT (*21*), the immunological niche difference bounds the fitness differences compatible with mutual invasion, the condition under which two competing species (in this case, pathogen variants) can invade one another when the other species is at equilibrium. In short, the greater niche difference $1-\rho$ (likewise, the smaller niche overlap ρ), the more easily competitively imbalanced strains can mutually invade, satisfying the following inequality:

$$\rho < \frac{f_i}{f_j} < \frac{1}{\rho} . \qquad (3)$$

Importantly, all quantities necessary to inform the immunological niche difference and the fitness difference ($\mathcal{R}_{0,i}, S_{i|j}$) can be derived from epidemiological models tailored to individual diseases, allowing for quantitative comparisons across systems.

Mutual invasion is often associated with long-term, stable coexistence in the classical MCT (*21*) and in the broader ecological and evolutionary literature (*22, 23*). However, nonlinear host-pathogen interactions, especially immunological feedbacks, add challenges to translating mutual invasibility to predictions about prolonged pathogen co-circulation, a key requirement for maintaining diversity of pathogen strains (*24*). Here, we begin by comparing invasibility across pathogen systems and then turn to predicting pathogen co-circulation from invasibility.

Comparative pathogen invasion dynamics

We quantify mutual invasibility of various pathogen variant pairs by applying PIT (SM section S4) to a diverse set of empirically motivated models of multistrain pathogen dynamics (*1, 11–19*). The results are plotted in Fig. 2, in which the region of mutual invasibility is

bounded by Eq. 3. The boundaries of mutual invasibility, predicted by PIT, are equivalent to those predicted for ecological communities by MCT (*21*).

For nearly all pathogen variant pairs tested, PIT predicts mutual invasion (Fig. 2). Two main factors contribute to this near universality of mutual invasibility: First, infection with one variant almost always confers weaker heterotypic (cross) immunity against other related pathogens, or their respective variants, than homotypic immunity, creating distinct immunological niches. Second, many closely related pathogens or pathogen variants have similar $\mathcal{R}_0$, which causes limited variation in the resulting fitness difference; when $\mathcal{R}_0$ is similar, even a small advantage in cross immunity ($S_{i|i} < S_{j|i}$, $S_{j|j} < S_{i|j}$) can be sufficient to allow for mutual invasion. For example, these two factors that favor mutual invasion apply to antigenically variable strains of seasonal influenza, for which strain replacement is the norm (*5, 25–27*).

However, not all pathogen variants have similar $\mathcal{R}_0$, in which case larger immunological niche differences are required for mutual invasion. For example, pathogens in the Paramyxovirus
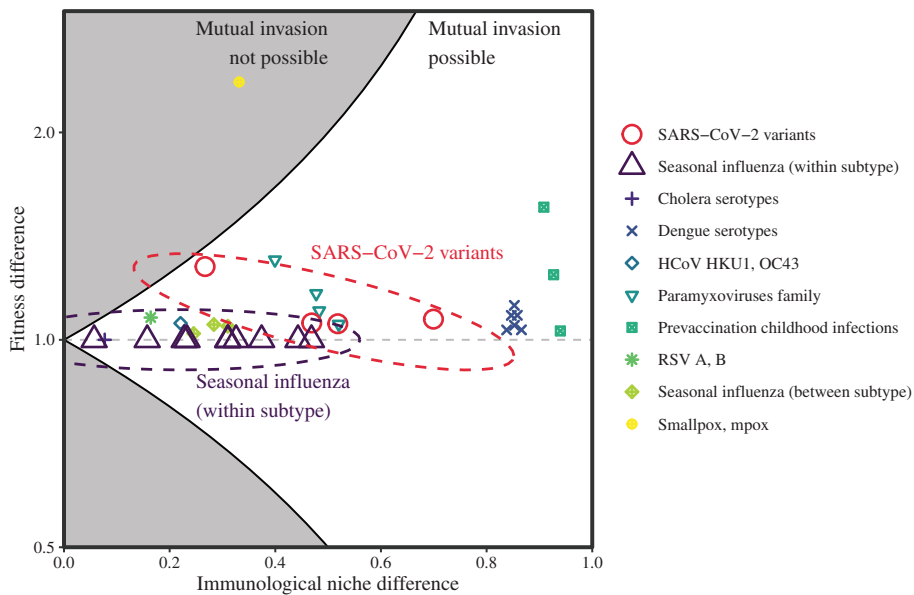
Fig. 2. Quantifying coexistence patterns of different human pathogen variant pairs. Niche and fitness difference estimates across different human pathogen variants in the absence of vaccination. Each point indicates a pairwise comparison of pathogen variants from various systems, indicated by the associated shape. Comparisons for prevaccination childhood infections consider ecological interference between measles versus chickenpox, rubella, and whooping cough (1). Points in the white region allow mutual invasion, meaning that a variant can invade a population when the competing variant is at equilibrium. The yellow point (smallpox and mpox) within the gray region does not allow mutual invasion, meaning that the presence of a more fit variant (smallpox) prevents the invasion of a less fit variant (mpox). Dashed ellipses indicate SARS-CoV-2 variant pairs and within-subtype seasonal influenza strain pairs, which exhibit strain replacement patterns. We plot max $(f_i/f_j, f_j/f_i)$ such that the fitness difference is always greater than 1. The fitness difference is plotted on a log scale for symmetry.

family exhibit 15 to 70% difference in $\mathcal{R}_0$ (10) but still permit mutual invasion. Another key example is provided by SARS-CoV-2 variants, for which we predict mutual invasion despite large differences in $\mathcal{R}_0$ driving strain replacement patterns (28, 29). Accounting for the effects of vaccination (fig. S1), or assuming a greater transmission advantage (fig. S2), prevented the SARS-CoV-2 wild-type variant from invading the SARS-CoV-2 Alpha variant; however, mutual invasibility predictions for all other variants are generally robust across different assumptions about vaccination, increased transmissibility, and cross immunity (SM section S8; figs. S1 to S3).

One exception to mutual invasibility is smallpox in competition with mpox; in this case, PIT predicts that the endemic presence of smallpox would prevent the invasion of mpox. This prediction is consistent with previous hypotheses that mpox invasion was only possible because of smallpox eradication and the eventual waning of immunity against smallpox (2, 18). The apparent contradiction between the predicted mutual invasibility of seasonal influenza strains and SARS-CoV-2 variants (Fig. 2) and the observed patterns of competitive replacement (Fig. 1) raises a critical question of what mechanisms determine the co-circulation

versus replacement of variants that can mutually invade one another.

Mutual invasion, persistence, and co-circulation

Given mutual invasion, the fate of pathogen competition following invasion is determined by the extent of overcompensatory susceptible depletion dynamics, which can prevent both competitors from reinvading the host population (5, 6). When overcompensation is strong, co-circulation requires both resident and invading variants to effectively persist until the susceptible pool is replenished. Although overcompensatory effects can be also found in other ecological systems, including predator-prey systems, most MCT analyses and interpretation of the mutual invasion condition either (1) focus on systems that experience limited overcompensation and do not exhibit large population cycles or (2) allow for long time scales for recovery from vanishingly low densities (21, 30). For overcompensating pathogen systems without persistent reservoirs, such recovery over long time scales is infeasible as a result of demographic stochasticity.

To illustrate how overcompensatory dynamics shape the outcome of pathogen interactions, we begin with a classic RSV model, which pre-

dicts that the invasion of either RSV A or B will cause both competitors to quickly settle into their respective endemic cycles (Fig. 3A). Specifically, these cycles match observed dynamics from outbreaks in Finland (15). This contrasts with the strain replacement pattern observed with SARS-CoV-2 variants and within influenza subtypes (Fig. 1, D and E). What factors allow RSV, but not SARS-CoV-2 or influenza, to overcome these overcompensatory effects?

To synthesize the different competitive outcomes across these three major pathogen systems, we explore how changes in the duration of immunity and transmission advantage of the invading variant affect the competitive outcome (Fig. 3D). In doing so, we summarize changes in the susceptible pool using two key measures: the degree of overcompensation (Fig. 3E) and the return time (Fig. 3F). The degree of overcompensation—defined as the maximum reduction in the effective reproduction number $\mathcal{R}(t)$ [i.e., $1 - \min (\mathcal{R}(t))$]—captures the extent to which the susceptible pool is reduced below an invasion threshold. The return time—defined as the time until $\mathcal{R}(t)$ is increased above 1 for the first time after invasion—measures the amount of time it takes for the susceptible pool to replenish to permit reinvasion. As we illustrate below, the degree of overcompensation is mostly related to the transmission advantage of the invading variant, whereas the return time is affected by both. Using $\mathcal{R}(t)$ allows us to measure changes in the susceptible pool relative to $\mathcal{R}_0$. We specifically focus on measuring the degree of overcompensation and return time using $\mathcal{R}(t)$ of the resident variant to characterize its ability to persist following the invasion of a competing variant. Previous studies have explored how immunological details affect the co-circulation of pathogens (5, 20), but susceptible dynamics often exhibit complex cycles, making them difficult to interpret. Summarizing the susceptible dynamics into the degree of overcompensation and return time both simplifies and reveals how changes in the duration of immunity and transmission advantage of the invading variant translate to changes in susceptible depletion dynamics, which in turn determine the competitive outcome.

We first identify conditions that allow co-circulation of mutually invading pathogens. Short-term immunity and limited transmission advantage (e.g., RSV) cause limited overcompensation (Fig. 3E) and fast return time from the overcompensated state (Fig. 3F), resulting in prolonged co-circulation without local extinction of either variant (Fig. 3D, yellow region). Pneumococcal serotypes, in which the underlying dynamics exhibit limited overcompensation, also experience this regime (8). In the case of limited overcompensation, PIT allows us to quantify how different immune mechanisms contribute to
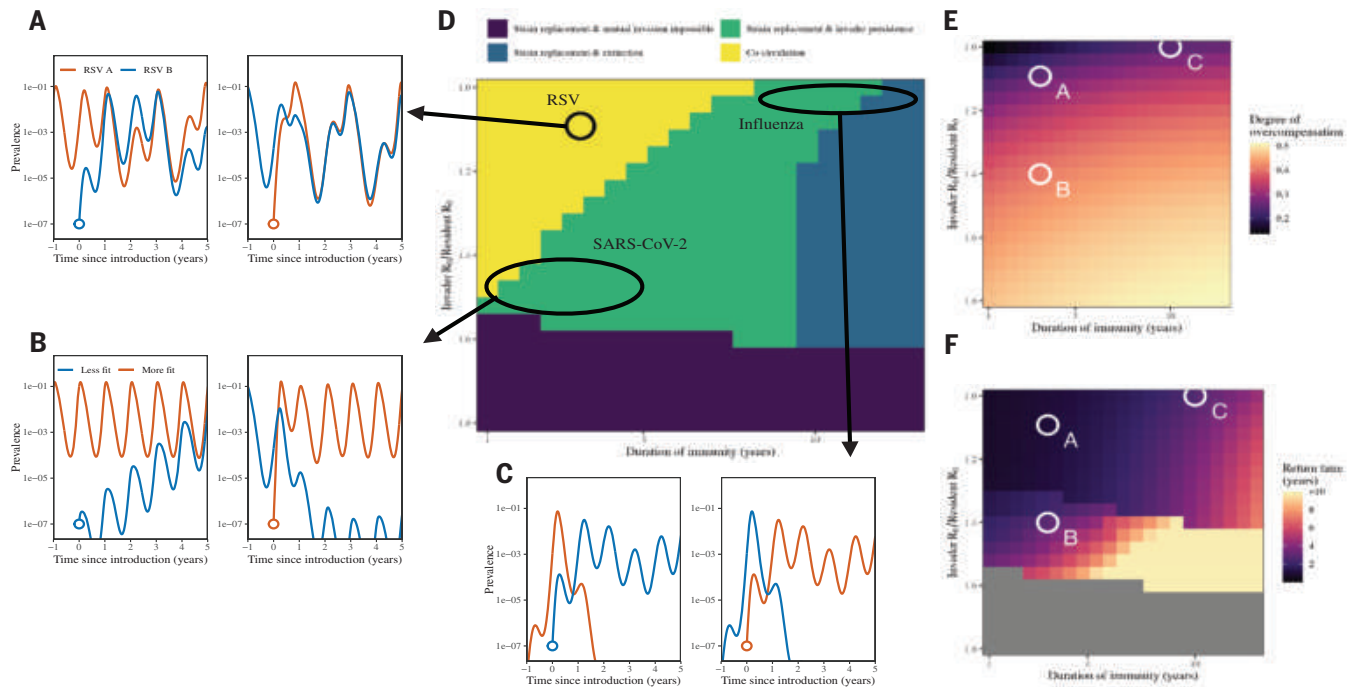
Fig. 3. Effects of epidemiological and immunological factors on overcompensatory dynamics and competitive outcomes. (**A**) to (**C**) Example mutual invasion simulations that capture competitive outcomes of (A) RSV subtypes, (B) SARS-CoV-2 variants, and (C) seasonal influenza strains within each subtype. Simulations assume (A) parameter estimates from (15); (B) 40% higher R0 for the more fit variant (orange); and (C) equal R0 for both variants and 10 years for the average duration of immunity, which is defined as the average time that a recovered individual returns to a fully susceptible state. All other parameters are from (15), including immune escape values which assume stronger homotypic immunity (64% protection) than heterotypic immunity (16% protection). (**D**) Competitive outcomes based on 10 years of simulations following the invasion of a novel variant while the resident is at equilibrium. Mutual invasion is possible in yellow, green, and blue regions. Mutual invasion is not possible in the purple region. 10-year persistence of both resident and invading variants is used as a minimum threshold for long-term co-circulation. Persistence and extinction is distinguished by a prevalence cutoff of 10^{-7}. All other parameters were fixed to previously estimated values. (**E** and **F**) Estimates of the degree of overcompensation and return time for the resident variant, corresponding to simulations in (D). For illustrative purposes, seasonal transmission was not included for simulations in (D) to (F). The standard RSV model by (15) was used for all simulations.

the maintenance of diversity (SM sections S6 and S9; fig. S4). In what follows, we show that the replacement of mutually invading strains can be caused by either severe overcompensation or modest overcompensation coupled with a slow return time.

Under a large transmission advantage, the invasion of a more transmissible variant can result in strong overcompensation (Fig. 3E) and slower return times to a state that permits further invasion (Fig. 3F) even when the duration of immunity of the invading variant is short. This effect can prevent the less transmissible variant from reinvading until the susceptible pool is replenished and causes strain replacement despite mutual invasion (Fig. 3D, green region). This prediction suggests that the increased transmissibility of successive SARS-CoV-2 variants (28, 29, 31) likely allowed for strong overcompensation, thereby preventing past variants from reinvading the population (Fig. 3B). This prediction is further supported by invasion simulations from a realistic SARS-CoV-2 model, which shows that increased transmissibility is required to drive past variants extinct (SM section S10; figs. S5 to S7).

Finally, longer-term immunity can lead to a slow return time (Fig. 3F), which for example explains the replacement among seasonal influenza strains within a subtype (Fig. 3, C and D). In fact, when the duration of immunity is very long, replenishment of the susceptible pool has to rely primarily on the birth of susceptible hosts, which can be very slow. This can prevent both resident and invading strains from reinvading the population after an outbreak, causing the eventual extinction of both strains. We note that whether the invading strains can persist or experience eventual extinction is likely indistinguishable in practice when new strains keep emerging as a result of antigenic drift or shift. These observations are further supported by more detailed influenza simulations, which show that prolonged immunity against individual seasonal influenza strains of the same subtype (32, 33) makes co-circulation rare despite near-universal mutual invasibility (SM sections S7 and S11; figs. S8 to S10).

This conclusion is robust to assumptions about the strength of cross immunity, with weaker cross immunity creating a larger im-

munological niche difference, leading to wider regions of mutual invasibility (fig. S11). Furthermore, faster return time results in wider regions of co-circulation.

Summary

There has been extensive work analyzing the invasion and coexistence of pathogen variants across many individual host-pathogen systems; however, synthesis and comparison across systems have been limited. We present a unifying framework, PIT, that allows for such comparison by quantifying key coexistence metrics—niche and fitness differences—in two-pathogen systems through the analysis of mechanistic models (fig. S13). PIT reveals that mutual invasion of initially rare variants is nearly universal for major pathogen systems, including exemplary systems of strain replacement such as seasonal influenza and SARS-CoV-2. Mutual invasibility implies stable coexistence among pairs of competitors in standard ecological systems (21).

By contrast, mutual invasibility hardly guarantees co-circulation in the pathogen systems we analyzed. Instead, the outcome of pathogen

competition depends on the extent to which susceptible hosts are reduced below the invasion threshold (i.e., the degree of overcompensation) and the amount of time it takes for the susceptible pool to replenish (i.e., the return time) (*6*). Specifically, a greater transmission advantage of an invading variant and more durable immune protection can cause stronger overcompensation and slower return time, which can lead to variant replacement. Comparisons of coexistence dynamics of RSV subtypes, SARS-CoV-2 variants, and seasonal influenza strains demonstrate the importance of understanding the overcompensatory nature of susceptible depletion in predicting outcomes of pathogen competition.

Caveats and future directions

A key limitation of PIT is that it relies on pairwise comparisons. This contrasts with observed competition patterns (Fig. 1), which may depend on multistrain interactions (e.g., dengue serotypes) as well as the evolutionary history of antigenically variable strains (e.g., seasonal influenza and SARS-CoV-2). Therefore, predicting coexistence and replacement for pathogens that exhibit these characteristics would require detailed epidemiological models. In fact, extending pairwise predictions about coexistence to predictions about coexistence of multiple species is a major challenge in community ecology, particularly because of the presence of higher-order interactions (*34*). Predicting the effects of multipathogen interactions and evolutionary history on pathogen invasion and coexistence is an important area for future work (*7*).

Our PIT analysis centers around major, endemic pathogen systems, leaving the applicability of our findings to other pathogen systems as a topic for future work. Nonetheless, we generally expect conclusions about the near-universality of mutual invasibility to be robust given that they hold for exemplary systems of replacement (i.e., seasonal influenza and SARS-CoV-2). While PIT relies on predicting invasion into a resident system at equilibrium (and analogously, during stable epidemic cycles), the dynamics of many pathogens, especially those that exhibit strain replacement patterns, can be far from the equilibrium. Understanding how invasibility changes across time, specifically in relation to the timing of past speciation of major pathogen lineages, will be crucial for predicting the future emergence of new pathogens. Comparing the impact of alternate formulations on niche and fitness differences as well as predictions of the final outcome of the competition will be also an important area for future work.

We focus on transmission advantage and duration of immunity as key variables that determine co-circulation. However, there are many other factors that maintain the diversity of pathogens, including evolutionary constraints (*35*) and other immunological mechanisms, such as strains that transcend immunity (*5*). Although our analysis does not account for these factors, our framework allows us to test whether new pathogens arising under these constraints can mutually invade, providing a tool to directly quantify the effects of various epidemiological and immunological factors on co-circulation.

Our invasion analysis focuses on competition within single, well-mixed host populations and therefore neglects spatial variation. This likely renders our predictions about pathogen coexistence conservative. In reality, heterogeneity in population dynamics across space can play an important role in determining overcompensation and maintaining pathogen persistence (*36*). Seasonality can also affect overcompensation and contribute to pathogen persistence (*37*). Despite these limitations, our work provides a quantitative foundation for understanding pathogen co-circulation.

Conclusion

The recent emergence and reemergence of novel pathogens has caused considerable societal disruption, emphasizing the need for better surveillance and control strategies to prevent future outbreaks. The near universality of pathogen mutual invasibility raises a serious public health concern given the potential recombination among variants and the emergence of new pandemic variants (*38*). Our results further underline the importance of global pathogen and immunological surveillance platforms for clarifying the dynamics of a potentially diverse range of future threats (*39*). A detailed understanding of immuno-epidemiological dynamics will be critical to predicting the persistence and community structure of novel pathogens in the future.

REFERENCES AND NOTES

1. P. Rohani, D. J. Earn, B. Finkenstädt, B. T. Grenfell, *Proc. Biol. Sci.* **265**, 2033–2041 (1998).
2. J. O. Lloyd-Smith, *Philos. Trans. R. Soc. B* **368**, 20120150 (2013).
3. E. W. Seabloom *et al.*, *Ecol. Lett.* **18**, 401–415 (2015).
4. A. J. Sieben, J. R. Mihaljevic, L. G. Shoemaker, *Ecology* **103**, e3819 (2022).
5. N. M. Ferguson, A. P. Galvani, R. M. Bush, *Nature* **422**, 428–433 (2003).
6. O. Restif, B. T. Grenfell, *Proc. Biol. Sci.* **273**, 409–416 (2006).
7. B. Adams, A. Sasaki, *Math. Biosci.* **210**, 680–699 (2007).
8. S. Cobey, M. Lipsitch, *Science* **335**, 1376–1380 (2012).
9. N. G. Reich *et al.*, *J. R. Soc. Interface* **10**, 20130414 (2013).
10. S. Bhattacharyya, P. H. Gesteland, K. Korgenski, O. N. Bjørnstad, F. R. Adler, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13396–13400 (2015).
11. S. M. Kissler, C. Tedijanto, E. Goldstein, Y. H. Grad, M. Lipsitch, *Science* **368**, 860–868 (2020).
12. M. Meijers, D. Ruchnewitz, J. Eberhardt, M. Łuksza, M. Lässig, *Cell* **186**, 5151–5164.e13 (2023).
13. N. Ferguson, R. Anderson, S. Gupta, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 790–794 (1999).
14. D. J. Smith *et al.*, *Science* **305**, 371–376 (2004).
15. L. J. White, M. Waris, P. A. Cane, D. J. Nokes, G. F. Medley, *Epidemiol. Infect.* **133**, 279–289 (2005).
16. K. Koelle, M. Pascual, M. Yunus, *Proc. Biol. Sci.* **273**, 2879–2886 (2006).
17. T. Bedford, A. Rambaut, M. Pascual, *BMC Biol.* **10**, 38 (2012).
18. R. Grant, L. L. Nguyen, R. Breban, *Bull. World Health Organ.* **98**, 638–640 (2020).
19. W. Yang, E. H. Y. Lau, B. J. Cowling, *PLOS Comput. Biol.* **16**, e1007989 (2020).
20. L. Yan, R. A. Neher, B. I. Shraiman, *eLife* **8**, e44205 (2019).
21. P. Chesson, *Annu. Rev. Ecol. Syst.* **31**, 343–366 (2000).
22. M. Turelli, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5085–5089 (1978).
23. S. A. H. Geritz, E. Kisdi, G. Meszena, J. A. J. Metz, *Evol. Ecol.* **12**, 35–57 (1998).
24. R. Poulin, S. Morand, *Q. Rev. Biol.* **75**, 277–293 (2000).
25. J. R. Gog, B. T. Grenfell, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 17209–17214 (2002).
26. K. Koelle, S. Cobey, B. Grenfell, M. Pascual, *Science* **314**, 1898–1903 (2006).
27. T. Bedford *et al.*, *eLife* **3**, e01914 (2014).
28. C. A. B. Pearson *et al.*, Bounding the levels of transmissibility and immune evasion of the Omicron variant in South Africa. medRxiv 2021.12.19.21268038 [Preprint] (2021).
29. R. Earnest *et al.*, *Cell Rep. Med.* **3**, 100583 (2022).
30. S. J. Schreiber, J. M. Levine, O. Godoy, N. J. B. Kraft, S. P. Hart, *Ecology* **104**, e3838 (2023).
31. E. Volz *et al.*, *Nature* **593**, 266–269 (2021).
32. H. C. Kung, K. F. Jen, W. C. Yuan, S. F. Tien, C. M. Chu, *Bull. World Health Organ.* **56**, 913–918 (1978).
33. D. Hannant, J. A. Mumford, D. M. Jessett, *Vet. Rec.* **122**, 125–128 (1988).
34. J. M. Levine, J. Bascompte, P. B. Adler, S. Allesina, *Nature* **546**, 56–64 (2017).
35. A. Kucharski, J. R. Gog, *Proc. R. Soc. B* **279**, 645–652 2012.
36. B. Bolker, B. Grenfell, *Philos. Trans. R. Soc. B.* **348**, 309–320 (1995).
37. V. Andreasen, G. Dwyer, *Am. Nat.* **201**, 639–658 (2023).
38. H. L. Wells *et al.*, *Cell Host Microbe* **31**, 874–889 (2023).
39. M. J. Mina *et al.*, *eLife* **9**, e58989 (2020).
40. S. W. Park, parksw3/PIT: archive, Version v1.0.0, Zenodo (2024); https://doi.org/10.5281/zenodo.12610514.
41. G.-Y. Kim, I. Rheem, Y. H. Joung, J. K. Kim, *Respir. Res.* **21**, 191 (2020).
42. S. Takahashi *et al.*, *PLOS Med.* **13**, e1001958 (2016).
43. L. C. Katznelnick *et al.*, *Science* **358**, 929–932 (2017).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adq0072
Materials and Methods
Supplementary Text
Figs. S1 to S12
References (*44–50*)
MDAR Reproducibility Checklist

MICROSCOPY

Direct optical measurement of intramolecular distances with angstrom precision

Steffen J. Sahl[1]*, Jessica Matthias[2]†‡, Kaushik Inamdar[1,3]‡, Michael Weber[1]‡, Taukeer A. Khan[1], Christian Brüser[1,4], Stefan Jakobs[1,3,4,5], Stefan Becker[6], Christian Griesinger[5,6], Johannes Broichhagen[7]§, Stefan W. Hell[1,2,5]*

Optical investigations of nanometer distances between proteins, their subunits, or other biomolecules have been the exclusive prerogative of Förster resonance energy transfer (FRET) microscopy for decades. In this work, we show that MINFLUX fluorescence nanoscopy measures intramolecular distances down to 1 nanometer—and in planar projections down to 1 angstrom—directly, linearly, and with angstrom precision. Our method was validated by quantifying well-characterized 1- to 10-nanometer distances in polypeptides and proteins. Moreover, we visualized the orientations of immunoglobulin subunits, applied the method in human cells, and revealed specific configurations of a histidine kinase PAS domain dimer. Our results open the door for examining proximities and interactions by direct position measurements at the intramacromolecular scale.

O wing to its minimal invasiveness, fluorescence microscopy has played a central role in the quest for understanding the myriad of functions of biomolecules. However, this endeavor calls for the quantification of sizes, associations, and conformational changes of proteins and other biomolecules, which in turn requires precise measurements of inter- and intramolecular distances, r, equal to 1 to 20 nm. Although measuring such distances is clearly challenged by diffraction, subdiffraction (super)resolution fluorescence microscopy also struggles with this task. Indeed, since the 1960s, macromolecular distances have been inferred indirectly, namely by the phenomenon called Förster resonance energy transfer (FRET) (1). FRET deduces distances r from the probability with which an excited fluorophore transfers its excited state energy to an electronically relaxed fluorophore that features a longer excitation and emission wavelength and is located further away by r. As the only optical method to explore proximities <10 nm in both protein ensembles and on the individual protein level, FRET (2–4) has become one

of the most popular methods in the life sciences. Dubbed a "molecular ruler," FRET has thus led to many insights both in cells and in vitro studies of biomolecules (5–7).

FRET, however, entails experimental provisions stemming from the indirect nature of the measurement (5). Because the energy transfer occurs through an interaction of the two transition dipole moments of the fluorophores, the result varies inversely with the sixth power of the distance, $1/r^6$, rendering FRET a highly nonlinear ruler (6, 7). Moreover, although reproducible results for distance measurements have been demonstrated with a range of carefully established correction factors (8, 9), the FRET process fundamentally depends on the usually uncontrolled relative orientation of the two transition dipoles and on the polarizability of the molecules between them. This altogether makes reliable determinations of r without prior knowledge about the molecular details of the system challenging and occasionally impossible. The situation is exacerbated by the fact that FRET measurements are, in practice, limited to (at most) 2 to 8 nm, because outside this range the dipole-dipole interaction hardly modulates with r. All these factors can lead to considerable uncertainties, making FRET much better at providing "yes" or "no" answers regarding colocalization in cells or large molecular ensembles than at determining exact (intra)molecular distances.

In principle, nanometer-scale proximities between two spectrally shifted fluorophores with different excitation and/or emission wavelengths can be measured (10–12) because the spectral shift allows for the separation of the fluorophores despite diffraction. Provided the spectral shift is sufficient, each fluorophore can be localized by calculating the centroid of its fluorescence diffraction spot rendered on a camera. Ironically, in the sought-after range of $r \lesssim 8$ nm, the spectral shift undesirably elicits

FRET between the fluorophores, thus hampering their independent emission. Independence can be gained by ensuring that the fluorophores emit sequentially, through an on-off switching of their fluorescence, as in the super-resolution (13) method called photoactivated localization microscopy (PALM) or stochastic optical reconstruction microscopy (STORM) (14–16). However, reversible on-off switching does not decouple the fluorophores entirely, and for distances <10 nm, this residual coupling (17) elevates bleaching (18).

Improved emission independence can be gained by binding fluorophores transiently to the sites of interest, as in the method called points accumulation for imaging in nanoscale topography (PAINT) (19), which is highly effective when mediated by DNA hybridization (20, 21). Although subnanometer localization precisions can be obtained (22), DNA-PAINT requires that the sites of interest be decorated with and accessed by labeled single strands of DNA. This labeling procedure is, by far, more restricting than directly attaching a fluorophore at the protein site of interest because not all biomolecules can be labeled by DNA oligonucleotides. Moreover, the oligonucleotides do not pass the living cell membrane, and the fluorophore labels need constant replenishment by fresh labels from the surrounding solution. Finally, the highly charged DNA may affect the molecular conformation (23).

Conceptually even more serious at molecular distances are the fundamental limitations of localizing a fluorophore by establishing the centroid of its fluorescence diffraction spot on a camera. First, centroid calculation critically relies on the detection of many fluorescence photons because the localization precision σ scales with $1/\sqrt{N}$, with N being the total number of detected photons in the camera spot. Thus, the need to wait for many emissions entails frequent excursions of the fluorophore to transient, nonemitting states, sample drift, and bleaching. Second, unless the fluorophore emits isotropically in space, which is virtually impossible given that it is target-bound, the dipole nature of emission leads to systematic deviations of the centroid position from the actual position of the molecule (24–27), which can frequently exceed 10 nm. Because the rotational freedom and the emission (an)isotropy are a priori unknown, these deviations mandate special caution regarding the accuracy of any method based on centroid determination.

In this work, we show that MINFLUX localization (28) determines (intra)molecular distances linearly and with subnanometer precision at room temperature. In MINFLUX localization, the fluorophore position is established by relating, and ideally matching, the known position of a central zero of a doughnut-shaped excitation beam to the unknown position of the fluorophore (Fig. 1A). Thus, MINFLUX implies

a minimization rather than a maximization of the number of required fluorescence photons. Because in MINFLUX localization the localization precision σ scales exponentially with the number of detected photons e^{-N}, the issues of camera-based localization are overcome. Indeed, MINFLUX typically requires about 100 times fewer detected photons to reach the camera localization precision (28, 29). By using photoactivatable (caged) fluorophores, emission orthogonality can also be ensured such that one fluorophore emits while the other one remains fully inert to the excitation. Additionally, the localization of the fluorophores with a circularly polarized doughnut zero renders the localization largely independent of their orientation. The resulting reliable determination of each fluorophore's coordinates has enabled us to measure static molecular distances directly, that is, without mediation by a multifactorial process such as FRET. Quantifying 1- to 20-nm distances has thus

allowed us to visualize, for example, the end-to-end distance of a small 16-kDa protein or provide access to conformational details of a larger 150-kDa protein by direct positional measurements of the reporter fluorophores.

MINFLUX as a linear ruler in the <10-nm (intra)molecular range

To explore the potential of the chosen MINFLUX protocol (Fig. 1A and methods), we initially turned to the polyproline helix secondary structure as a distance reference. The widely assumed rigidity of this polypeptide (30, 31), especially for low to intermediate numbers (<20) of repeats, allows the positioning of fluorophores at defined distances. Indeed, polyprolines enabled Stryer and Haugland to demonstrate (32) the $1/r^6$ FRET dependence in the 1960s, which was put forward by Förster in 1948. Polyprolines also allowed Schuler, Eaton, and co-workers to demonstrate distance-dependent FRET at the single-molecule level (33).

Conjugation of the photoactivatable dye DiMeO-ONB-SiR637 (34) to both ends of polyproline chains P-n, where n is the number of proline residues (P), with N- and C-terminal glycine or lysine residues produced putatively linear unimolecular systems (Fig. 1B). The synthesis was controlled to yield peptides without variation in the proline repeat number, meaning that the interfluorophore distance d was tunable from <3 nm (for P-5) to >10 nm (for P-40), taking as an expectation the highly extended type II (PPII) helix lengths with all-trans peptide bonds (Fig. 1C). The peptides were nonspecifically immobilized on the surface of a poly-L-lysine coated cover glass at high dilution and examined individually (see materials and methods).

To quantify the localization precision for individual fluorophores at the end of the polyprolines, we initially imaged samples of a separated fraction of mono-labeled P-15 polyproline. Fluorophores that exhibited noticeable positional

Fig. 1. Angstrom-precise intramolecular distance measurements based on direct optical measurements of positions in and beyond the FRET range. (**A**) MINFLUX approaches a fluorophore (red star, not to scale) with an illumination intensity minimum (center of doughnut pattern) to achieve its precise localization. (**B**) Tunable distances provided by polyproline, a peptide composed of relatively rigid helices of defined repeat number, terminal glycine and lysine residues, and identical photoactivatable fluorophores conjugated to both ends. The angstrom-range MINFLUX localization precision enables measurements of 1- to 10-nm intramolecular distances. (**C**) Prepared polyproline constructs: P-n, with n = 5, 7, 10, 12, 15, 20, 25, 30, 35, or 40, constituting a distance ladder. Expected end-to-end distances, assuming a type II helix in molecular dynamics simulations, range from ~2.5 to ~11 nm (33). (**D**) Sequential photoactivation used for addressing the fluorophores individually and in molecular proximity. (**E** to **G**) MINFLUX reconstructions of polyproline end-to-end distances, displayed as (E) localization sets (red points) and the localized dye coordinates (black crosses), (F) 2D histogram representation, and (G) inferred dye positions and their respective measurement precision. The ellipses correspond to 3σ contours. The extracted distance d is indicated below each example. (**H**) Distributions of measured distances d from P-5 to P-40. For 20 or more residues, long distances can be discerned in a broad second peak (indicated by arrows), but most end-to-end distances observed on the surface were found to be shorter (first peak). Red curves are Gaussian fits. Scale bars are 200 nm (A), 1 nm (B), 3 nm (C), and 5 nm [(E) to (G)].

drifts owing to insufficient immobilization were excluded from analysis. Photoactivation was ensured by a regularly focused 405-nm beam that was co-aligned with the 640-nm doughnut-shaped excitation beam. After activation of DiMeO-ONB-SiR637 by cleavage of the *ortho*-nitrobenzyl carbamates, the fluorophore was transferred from an inactive off state to an active on state with photostable red emission. Once an active fluorophore was identified, the MINFLUX localization algorithm ensured that the excitation doughnut rapidly zoomed in on the emitter (*35, 36*). The final two MINFLUX iterations (*36*) were repeated until photobleaching, which rendered extended sets of successive fluorophore position estimates (fig. S1). From these raw localization data, the statistical localization precision could be inferred directly. For example, for only 150 photons in the final MINFLUX iteration, we obtained $\sigma_{raw} = 2.2 \pm 0.2$ nm. Aggregation of groups of 5 and 10 consecutive localizations yielded datasets with a smaller spatial spread, namely $\sigma_{5\ combined} = 1.2 \pm 0.2$ nm and $\sigma_{10\ combined} = 0.9 \pm 0.2$ nm, respectively.

The mean position of the fluorophore, that is, its time-averaged center of mass, is subject to an even lower uncertainty owing to the combined information from even more equivalent position determinations obtained for long on times. In practice, the localization precision was estimated to be on the order of the microscope's active-feedback stabilization per measurement time interval (0.1 to 0.3 nm) for the >1000 localizations obtained on average. Under the assumption of stationary fluorophore centers, the mean of all localizations can be assigned with (sub)angstrom precision depending on the numbers of localizations accrued (fig. S1). Note that the stabilized MINFLUX system (*36*), including beamline monitoring, did not require further processing of localization data to correct for drifts.

Moving to polyprolines with two DiMeO-ONB-SiR637 dyes, we observed that, apart from a fraction of already activated fluorophores at the beginning of the measurement, the emitters were indeed activated and localized independently (Fig. 1D). Localization of a fluorophore was terminated by a photobleaching event. Because the activation probability, controlled by the 405-nm laser power, was kept low, the activation of the second fluorophore frequently occurred seconds to minutes later (fig. S2), showing that the activation of the two fluorophores was indeed independent.

Precise and accurate distance measurements in the FRET range

MINFLUX allowed for quantification of the fluorophore-fluorophore distance (Fig. 1, E to G, and fig. S3) from P-40, with examples in the range $d \approx 10$ to 12 nm, over intermediate spacings (e.g., P-30; $d \approx 8$ to 9 nm) down to

spacings deep within the FRET range (e.g., P-15; $d \approx 5$ nm); the shortest tested peptide was P-5 ($d \lesssim 3$ nm). The subnanometer precision of individual position measurements produced data that were easy to evaluate even at small values of d.

The distributions of measured distances (Fig. 1H) were generally broad, and especially so for the longer polyprolines. This broadening, interpreted to result from a considerable flexibility in the peptide chain, had also been observed in the single-molecule FRET data and was described in simulations (*33*). For oligomers with $n \geq 20$, we observed rather broad peaks in the respective ranges that were expected for largely extended helices from ~6 to ~10 nm (highlighted by arrows in Fig. 1H). A large subpopulation of polyprolines, however, exhibited much shorter end-to-end distances contained in the respective first peak of the distribution. We speculate that these may arise in part because of pronounced bending and uncontrolled interactions with the surface, which differed from the solution-phase environment of previous studies.

Up to 20 P, the average measured distances exhibited a close-to-linear scaling with the number of proline residues (Fig. 2A). Linear regression analysis in this range (P-5 to P-20) suggested an increment of $\Delta d \approx 0.21$ nm per proline residue. Although molecular dynam-

ics simulations would also explain moderate reductions of average end-to-end distance (*33*) for oligomers in this size range owing to chain bending, the observed increase of 2.1 Å, on average, per residue is in disagreement with PPII conformation of the helices in their entirety. For such all-trans bonding in the PPII conformation, which we expect to be generally favored under our experimental conditions, the helical pitch of 9.3 Å per turn and 3.0 residues per turn would result in an increase of 3.1 Å per residue. Instead, our data (Fig. 2A), falling short of the expected end-to-end distances of all-trans polyproline helices, may suggest heterogeneous compositions of cis and trans peptide-bond structural segments (*37, 38*), which, in combination, lead to the average linear scaling that was observed.

The constant offset (*y* intercept) of 1.9 ± 0.2 nm may be attributed in part to the fluorophores' displacement by the terminal glycine and lysine residues, which act as linkers, and to the fluorophores' physical size, which contributes additional distance to the "centers" of emission at both ends. In general, further contributions to distance variabilities may stem from motion or nonhorizontal attachments to the substrate and the two-dimensional (2D) projection measurement. Altogether, a horizontal alignment of example data shows that the entire "ladder" of intramolecular distances

Fig. 2. Intramolecular MINFLUX imaging of polyproline end-to-end distances. (**A**) Measured average distances versus number of proline residues P-*n*. For P-5 to P-15, the averages were extracted as the positions of the peaks of the distributions in Fig. 1H. For P-20 and longer, an indication of the average more extended distances is given from the positions of the second, broader peaks. A linear fit over the P-5 to P-20 range is shown (solid line), suggesting an increase of, on average, 2.1 ± 0.2 Å per residue. The range of distances from ~2 to 8 nm has been traditionally addressed by (single-molecule) FRET methods (gray region). The dashed line indicates a continuation of the linear scaling from the fit. R^2, goodness of fit. (**B**) Example data from Fig. 1, E to G, aligned by rotation to show the complete distance ladder (compare with Fig. 1C). The ellipses represent 3σ contours of localization uncertainty. Scale bar is 5 nm.

In panel A: $d = (0.21 * \text{P-}n + 1.88)$ nm, $R^2 = 0.97$. Axis labels: Measured average end-to-end distance d (nm); Number of residues P-*n*. "FRET distance range" marked in gray region.

In panel B: $d =$ 11.3 nm P-40; 9.5 nm P-35; 9.0 nm P-30; 8.3 nm P-25; 6.5 nm P-20; 5.1 nm P-15; 4.5 nm P-12; 4.0 nm P-10; 3.3 nm P-7; 2.8 nm P-5.
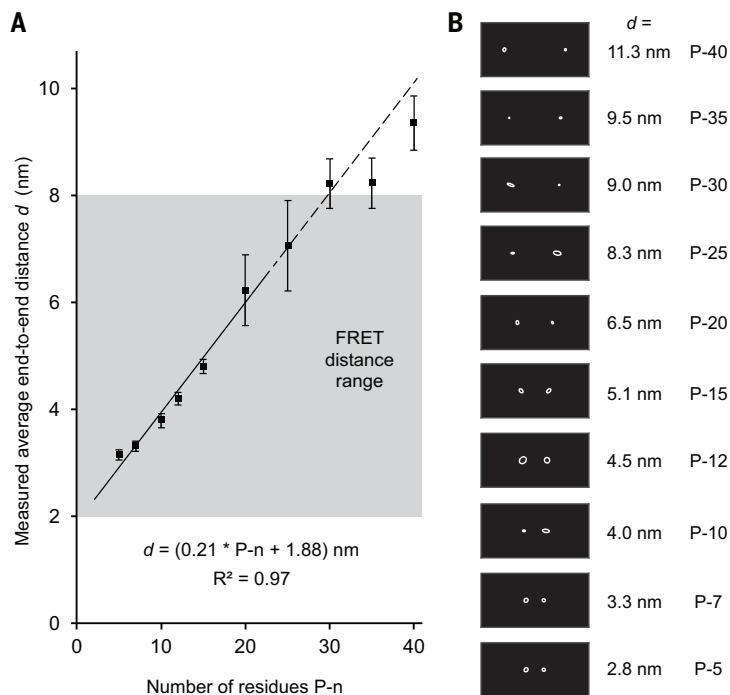
can be resolved and quantified (compare Figs. 1C and 2B) by MINFLUX.

Distance measurements on a small protein

We next turned to examine C- and N-terminally labeled camelid nanobodies for end-to-end distance measurements. The core of this compact 16-kDa protein positions its C and N termini at a ~3.7-nm distance (*39*) (Fig. 3A). The termini were labeled with a less hydrophobic photo-activatable dye (*40*), Abberior CAGE 635, by maleimide-coupling to the terminal cysteines. MINFLUX readily resolved this spacing, measuring (projected) distances predominantly in

the range of 2.5 to 5 nm (Fig. 3B). The distribution of distances (Fig. 3C) peaked at ~3.8 nm.

Interestingly, the distribution was accompanied by a tail of longer spacings, extending out to ~16 nm in rare instances. In particular, distances in the ranges of 7 to 9 nm and 10 to 13 nm (see examples in Fig. 3E) indicated the presence of nanobody dimers and trimers, respectively, whose end-to-end distances are expected to fall in this range. We hypothesized that dimers or trimers are formed by disulfide bridging (Fig. 3, D and E) of the free cysteines during the maleimide labeling protocol. A gradient-gel SDS–polyacrylamide gel electrophoresis (SDS-PAGE) analysis indeed revealed

dimers and trimers both in Coomassie stain and in the fluorescence observed after ultraviolet (UV) photoactivation (Fig. 3F and fig. S4). Owing to its high spatial resolution, MINFLUX had obviously identified the dimers and trimers directly by distance determinations despite their low abundance in the originally measured sample. Measurements of a mixture that was enriched in oligomers by size exclusion chromatography allowed us to assign subpopulations up to the size of tetramers at approximately multiples of the monomer end-to-end distance (Fig. 3H). Whereas the corresponding gel fluorescence (Fig. 3G) suggested less monomer, MINFLUX identified monomers as the most

Fig. 3. Intramacromolecular distance measurements: Sites on proteins and positioning of protein subunits. (**A**) Camelid nanobody (~16 kDa) with N and C termini cysteine-labeled with a photoactivatable dye. The dimensions of the prolate-shaped nanobody along the longer and shorter axes are ~4 nm by ~3 nm, and its N and C termini are 3.7 nm apart based on crystallographic data (*39*). (**B**) Corresponding MINFLUX example data. (**C**) Distribution of distances *d* obtained. The long tail of the distribution suggests the existence of nanobody dimers and trimers. The red curve is a Gaussian fit to estimate the peak position. (**D**) Distances attributed to dimers and trimers and (**E**) measurements corresponding to dimers (left) and trimers (right). (**F**) Coomassie stain and fluorescence of a gradient gel of the solution as imaged, with dimer and trimer bands highlighted by the red rectangles. (**G**) Gel fluorescence of a nanobody sample with a relative enrichment of dimers, trimers, and tetramers. (**H**) Distribution of measured distances for oligomer-enriched samples. A fit to a Gaussian model with four components (monomers, dimers, trimers, and tetramers) is shown (red curve). (**I**) IgG molecule, with the kappa lights chains decorated by N- and C-terminal dye-labeled nanobodies, thus providing up to four positional fluorophore marks (indicated by stars) on the antibody molecule. (**J**) Examples of different orientations of the two IgG arms. Scale bars are 1 nm [(A), (D), and (I)] and 5 nm [(B), (E), and (J)].
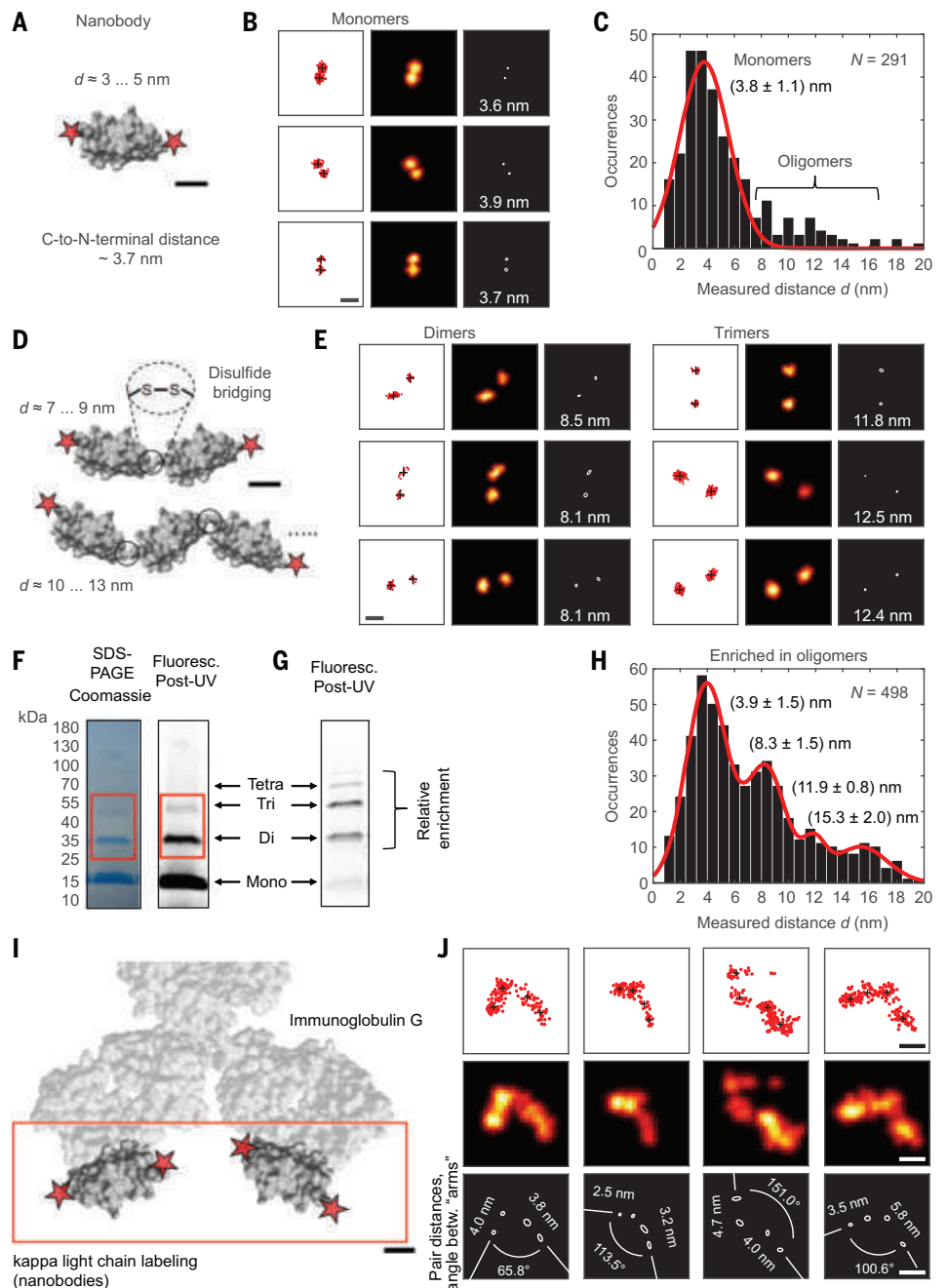
abundant species, possibly as a result of different kinetics of surface adsorption among the different nanobody species. Note that the spacing of dimers, trimers, and tetramers (fig. S5) falls outside the FRET range, highlighting the ability of MINFLUX to measure distances >10 nm as well.

Subunit structural arrangements of a large protein

Next, we extended the demonstration of our method to more reporter fluorophore sites. To this end, we labeled immunoglobulin G (IgG) with C- and N-terminally dye-labeled (Abberior CAGE 635) nanobodies that have affinity for the IgG's kappa light chain. Up to four fluorophores were localized on the surface of the IgG. The fluorophores formed two pairs aligned with either of the two flexible "arms" of the IgG

(Fig. 3I). Owing to incomplete labeling and/or sampling (fig. S6A), not all fluorophore positions could always be extracted. Prolonged observation of individual dyes also indicated incomplete immobilization, as suggested in examples of relative displacements of one arm during acquisition (fig. S6B). However, those antibodies in which all four dyes were registered clearly displayed the orientation of the two IgG arms, with the pair distances of each assigned arm falling in the range of the (2D-projected) nanobody distances (Fig. 3J).

Measurements in cells

Next, we stably expressed a translational fusion of HaloTag to the N terminus of Lamin A/C in human osteosarcoma (U-2 OS) cells. Nuclear lamins are fibrous proteins that form

intermediate filaments, providing structural stability to the nucleus and influencing gene expression (41). The ~3.5-nm-diameter filaments formed by lamins are substantially thinner than microtubular (24 nm), vimentin (10 nm), or actin (8 nm) filaments (42). Present models of the lamin structure (42, 43) suggest an assembly based on antiparallel pairings of dimers due to coiled-coil domain interactions (43). This arrangement explains the uniformly thick filament, along which globular Ig-like domains are positioned pairwise at the C termini of the individual lamins, on average every ~20 nm (Fig. 4A). The pairs of N termini in this arrangement are spaced at two unequal distances along the filament that add up to ~40 nm.

In our imaging experiments at the basal nuclear membrane, a dye-conjugated ligand bound

Fig. 4. Quantification of labeled protein distances in cells. MINFLUX imaging was used to examine the spatial distribution of HaloTag-Lamin A/C in the basal nuclear membrane of U-2 OS cells, where lamins form a filamentous meshwork. (**A**) Model of lamin filament molecular architecture (42, 43). An arrangement based on antiparallel pairings of dimers (43) is indicated. The N-terminally fused and fluorescently labeled (Abberior CAGE 635) pairs of HaloTags are positioned unevenly along the direction of the filament with a full period of ~40 nm, on either side of the C-terminal Ig-like domains that are known from electron microscopy to be spaced ~20 nm apart. (**B**) Examples of image regions. Arrows indicate sub-10-nm distances. Scale bars are 25 nm. (**C**) Examples of <10-nm distance measurements. Scale bars are 5 nm. (**D**) Histogram of individually quantified distances <20 nm.
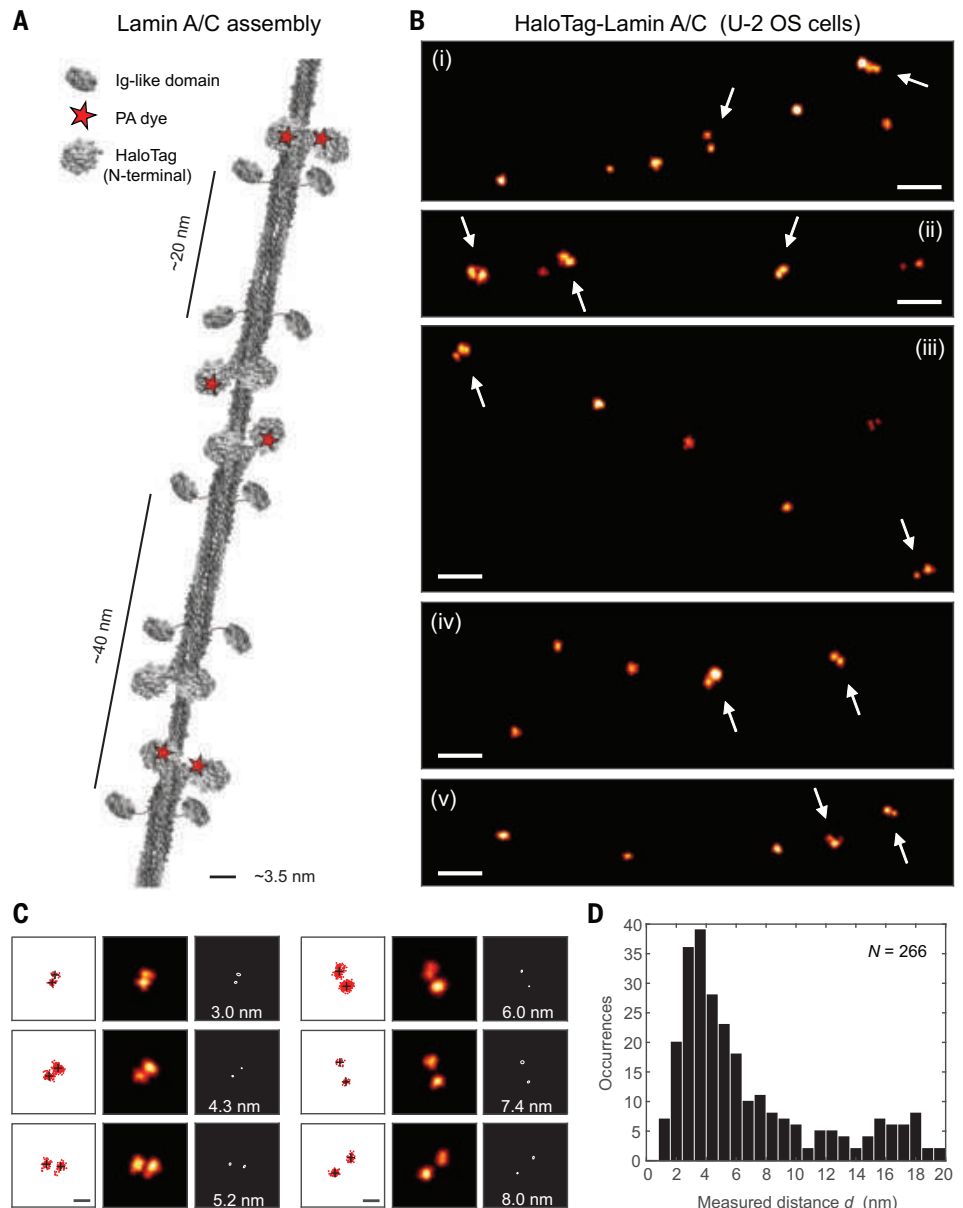
Fig. 5. Conformational measurements at and below FRET-range distances, and angstrom distance measurements: Protein domain dimer in parallel and antiparallel configuration.

(**A**) Crystallography data of the PASc domain of the bacterial citrate sensor histidine kinase (CitA) reveal that this dimer of protein domains exists in both antiparallel and parallel configurations. The separation of the C termini, labeled with two photoactivatable dyes (Abberior CAGE 635), differs between ~1 nm for the parallel dimeric arrangement and ~4 nm for the antiparallel dimer. (**B**) Examples of ~4- and ~1-nm distances. (**C**) Projected instances lead to distances <1 nm. The ellipses in the enlarged views on the right represent the 1σ (white), 2σ (yellow), and 3σ (red) contours of the position measurement uncertainty. (**D**) Distribution of all (2D-projected) C-terminal–to–C-terminal distances measured. Simulated projected distance distributions to capture the effects of the putatively random orientation of the dimers on the glass substrate can be found in fig. S10. Modeling with the assumption of isotropic orientations (red curve, see materials and methods) indicates a mixture of about half parallel, half antiparallel dimers.

(**E**) Measured 3D distances clearly identify both populations. The red curve is a two-component Gaussian model fit. (**F** and **G**) Examples of orientations and distances in 3D. The data are shown in projection to a vertically (z) oriented plane through the two measured 3D coordinates. Distance measurements (1σ, 2σ, and 3σ contours indicated) and the respective angle with the horizontal are shown. The corresponding 2D-projected distance is also indicated. Scale bars are 1 nm (A), 5 nm and 2 nm (enlarged view) (B), 5 nm and 5 Å (enlarged view) (C), 5 nm and 2 nm (enlarged view) (F), and 5 nm and 5 Å (enlarged view) (G).
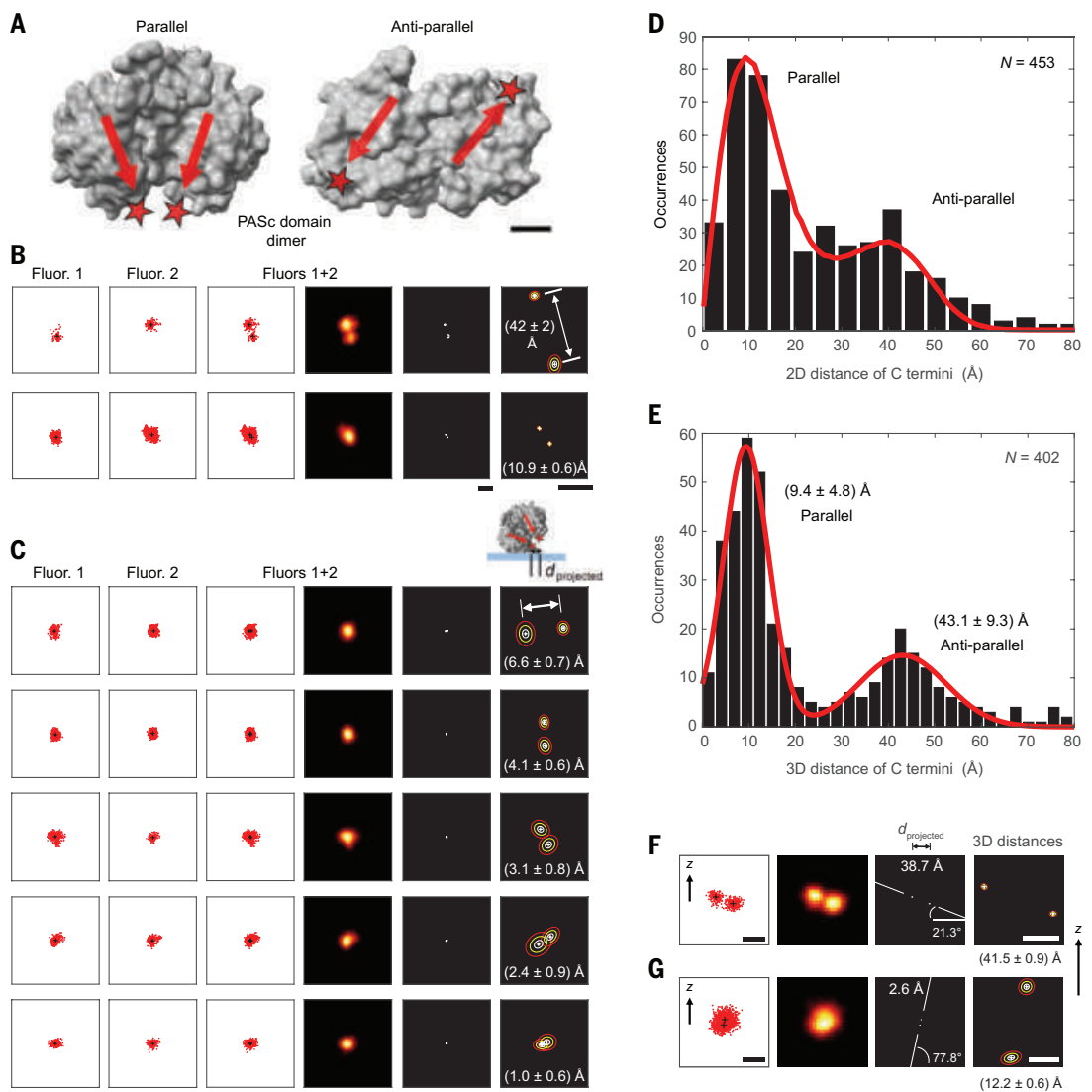
the HaloTag. Examples of 2D MINFLUX image regions (Fig. 4B) contained geometric arrangements of the labeled sites appearing to lie along a line, with pairs at small separation also seen. Incidentally, the shown regions merely span one diffraction zone (~300 nm across). Histograms of first nearest neighbor distances for the identified tag positions contain evidence for periodicities that can be accounted for by combinations of individual ~15-, ~25-, and ~40-nm separations (fig. S7). Numerous pairs of sites appear to correspond to the pairs of N termini (HaloTags) that are positioned directly adjacent. Because many of the first nearest neighbor distances also fell below 10 nm and could be reliably measured with high precision (Fig. 4, C and D; N = 266), these experiments demonstrate that distances <10 nm can be directly quantified in cells.

Dimeric protein domains in parallel and antiparallel configuration: Down to angstrom distances

One of the exciting prospects of the newly gained resolution capability in the intramolecular range is the direct imaging of distance distributions from an ensemble of macromolecules. Biological structures do not adopt one rigid and nonvarying arrangement. Molecules are flexible, and this flexibility should become visible in the many snapshots acquired. The resulting dataset is expected to contain the mean as the dominant structure, which is also observed by the averaging pursued in crystallography or cryo–electron microscopy. However, because the imaging of individual macromolecules maps out their conformational space, structural subpopulations become apparent by MINFLUX.

To demonstrate this capability, we investigated a protein domain dimer, the cytosolic PAS domain (PASc) of the bacterial citrate sensor histidine kinase (CitA) (*44*). Such a homodimeric system would be especially difficult to study by FRET because the approach typically relies on two different dyes. Crystallography data of PASc reveal that this dimer can assume both an antiparallel and parallel arrangement (*45*) (Fig. 5A). A pronounced quaternary structure rearrangement in the form of an antiparallel-to-parallel transition upon citrate binding is now thought to have a central role in transmitting and amplifying an

initially small structural change as part of the transmembrane signaling process (45).

The distance between the C termini, labeled with photoactivatable dyes (40), is expected to differ between ~1 nm (~10 Å) for the parallel dimeric arrangement and ~4 nm for the antiparallel dimer. In the imaging experiments, after sequential photoactivations (figs. S8 and S9), distance data were observed that ranged from the shortest distances of <1 nm, representing the parallel dimer at its full distance (or partially 2D-projected and thus reduced in length), to distances in the range of 4 nm, corresponding to the antiparallel arrangement (Fig. 5B). To appreciate the effects of the putatively random orientation of the dimers on the glass substrate, we simulated distributions of 2D-projected distances for species mixtures of various proportions (fig. S10). The data clearly contain a population of tilted dimers, for which the axis between the two fluorophores has a perpendicular component to the plane and for which distances between 0 and 1 nm were extracted with subangstrom measurement precision (Fig. 5C). The complete distribution (Fig. 5D; N = 453) indicates that about equal amounts of measured dimers were in the parallel and antiparallel states.

Equally angstrom-precise 3D position measurements (see materials and methods and fig. S11) provided additional information, allowing us to quantify the distribution of C-terminal separations as Euclidian 3D distances across the ensemble of sufficiently immobile dimers (Fig. 5E; N = 402). Uncompromised by planar projection, two distinct populations at 9.4 ± 4.8 and 43.1 ± 9.3 Å were identified, representing the parallel and antiparallel configurations in a nearly 2:1 mixture. MINFLUX thus quantified the relative abundance of the subpopulations with average C-terminal distances of ~1 and ~4 nm.

Individual examples of dimers measured to reside at different 3D orientations provide clear illustrations of the aforementioned 2D projection effects (Fig. 5, F and G): Species with mutual dye-dye separations of ~4 and ~1 nm are shortened in their x-y projected views to a reduced apparent distance, depending on orientation (fig. S12). Although the horizontally positioned molecules virtually agree in two and three dimensions, in the most extreme case, the projection shortens what is a ~12-Å species to ~2 Å in projection (Fig. 5G). The situation is readily interpreted as a parallel dimer standing almost orthonormal with the surface.

Conclusions

MINFLUX fluorophore localization down to ~0.1-nm precision enables accurate distance measurements down to the ~1-nm physical extent of the fluorophores (Fig. 5). For fluorophores positioned on macromolecules that lie at an angle and hence do not obstruct each other along the distance to be measured, we quantified distances well below 1 nm. Resolving the end-to-end distance of small proteins and oligomers, these measurements provide optical access to the intra-macromolecular scale. As illustrated by distinguishing distances of 1 to 4 nm in parallel and antiparallel dimer configurations, MINFLUX also directly resolves subunits of proteins, as well as their relative orientation and conformation.

Our study further shows that intramolecular resolution can be directly obtained by observing just the two fluorophores at the end of the spatial interval in question. Unlike in camera-based localization, hundreds of copies of fresh fluorophores binding sequentially to the two sites over a long time are not needed.

Ensuring independent dye emissions (figs. S13 and S14), MINFLUX enables linear distance measurements not only over the FRET range (2 to 8 nm) but also below 2 to 3 nm, where FRET usually falters. For distances >8 nm, MINFLUX measurements are fully viable (Fig. 3, D and E). Indeed, because MINFLUX treats all distances equally, any distance from the diffraction scale (>200 nm) down to smallest values (<1 nm) is directly measurable with angstrom precision.

The present experiments certainly average over any rotational details in the combined localization information from the multiple individual position estimates, which are each obtained in ~1 ms (with the probing beam dwelling in one of the positions for only tens of microseconds at any given time in the MINFLUX implementation used). We anticipate that further developments of stable, bright dyes with near-unity photoactivation yields will further improve the applicability of our approach for studies of distance distributions within protein and other biomolecular systems, including in cells. Improved immobilization strategies to minimize the surface influence on proteins and the demonstrated 3D capability will be especially valuable in this regard. We also anticipate extensions to larger numbers of molecular sites.

With photoactivation as the on-off mechanism, we observe that no "10-nm resolution barrier" (18) was encountered, as for the STORM-type thiol-based blinking (46), nor did the approach require cryogenic temperatures (47, 48). Unlike sample expansion (49), which involves harsh chemical treatment, MINFLUX does not risk sample alterations at the sub-10-nm scale. The power of MINFLUX stems from the fact that it (i) can use established fluorescence labeling procedures and (ii) uses the fluorescence photon budget about 100 times more effectively than camera-based localization while also essentially avoiding (iii) molecular-orientation effects and (iv) FRET dipolar coupling.

Because it offers a high degree of spatial parallelization, a remaining advantage of camera-based FRET imaging is the large field of view, which allows for the examination of many molecules simultaneously. However, ongoing efforts for parallelization of MINFLUX should translate the virtues of MINFLUX to larger fields of view and living cells. Finally, our work shows that fluorescence microscopy is undergoing a seminal transition from a method that merely maps out biomolecular spatial distributions in cells to one that directly reveals biomolecular function through precise determination of positions and distances with minimal invasiveness.

REFERENCES AND NOTES

1. T. Förster, *Ann. Phys.* **437**, 55–75 (1948).
2. T. Ha *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6264–6268 (1996).
3. G. J. Schütz, W. Trabesinger, T. Schmidt, *Biophys. J.* **74**, 2223–2226 (1998).
4. X. Zhuang *et al.*, *Science* **288**, 2048–2051 (2000).
5. R. Roy, S. Hohng, T. Ha, *Nat. Methods* **5**, 507–516 (2008).
6. L. Stryer, *Annu. Rev. Biochem.* **47**, 819–846 (1978).
7. W. R. Algar, N. Hildebrandt, S. S. Vogel, I. L. Medintz, *Nat. Methods* **16**, 815–829 (2019).
8. B. Hellenkamp *et al.*, *Nat. Methods* **15**, 669–676 (2018).
9. G. Agam *et al.*, *Nat. Methods* **20**, 523–535 (2023).
10. L. S. Churchman, Z. Ökten, R. S. Rock, J. F. Dawson, J. A. Spudich, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 1419–1423 (2005).
11. A. Pertsinidis, Y. Zhang, S. Chu, *Nature* **466**, 647–651 (2010).
12. S. Niekamp *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 4275–4284 (2019).
13. S. W. Hell, *Science* **316**, 1153–1158 (2007).
14. E. Betzig *et al.*, *Science* **313**, 1642–1645 (2006).
15. M. J. Rust, M. Bates, X. Zhuang, *Nat. Methods* **3**, 793–795 (2006).
16. S. T. Hess, T. P. K. Girirajan, M. D. Mason, *Biophys. J.* **91**, 4258–4272 (2006).
17. T. Ha, P. Tinnefeld, *Annu. Rev. Phys. Chem.* **63**, 595–617 (2012).
18. D. A. Helmerich *et al.*, *Nat. Methods* **19**, 986–994 (2022).
19. A. Sharonov, R. M. Hochstrasser, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18911–18916 (2006).
20. R. Jungmann *et al.*, *Nano Lett.* **10**, 4756–4761 (2010).
21. M. Dai, R. Jungmann, P. Yin, *Nat. Nanotechnol.* **11**, 798–807 (2016).
22. S. C. M. Reinhardt *et al.*, *Nature* **617**, 711–716 (2023).
23. S. M. Früh *et al.*, *ACS Nano* **15**, 12161–12170 (2021).
24. J. Enderlein, E. Toprak, P. R. Selvin, *Opt. Express* **14**, 8111–8120 (2006).
25. S. Stallinga, B. Rieger, *Opt. Express* **18**, 24461–24476 (2010).
26. J. Engelhardt *et al.*, *Nano Lett.* **11**, 209–213 (2011).
27. M. D. Lew, M. P. Backlund, W. E. Moerner, *Nano Lett.* **13**, 3967–3972 (2013).
28. F. Balzarotti *et al.*, *Science* **355**, 606–612 (2017).
29. Y. Eilers, H. Ta, K. C. Gwosch, F. Balzarotti, S. W. Hell, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 6117–6122 (2018).
30. P. M. Cowan, S. McGavin, *Nature* **176**, 501–503 (1955).
31. A. A. Adzhubei, M. J. E. Sternberg, A. A. Makarov, *J. Mol. Biol.* **425**, 2100–2132 (2013).
32. L. Stryer, R. P. Haugland, *Proc. Natl. Acad. Sci. U.S.A.* **58**, 719–726 (1967).
33. B. Schuler, E. A. Lipman, P. J. Steinbach, M. Kumke, W. A. Eaton, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 2754–2759 (2005).
34. M. Weber *et al.*, *Chem. Eur. J.* **27**, 451–458 (2021).
35. K. C. Gwosch *et al.*, *Nat. Methods* **17**, 217–224 (2020).
36. R. Schmidt *et al.*, *Nat. Commun.* **12**, 1478 (2021).
37. S. Doose, H. Neuweiler, H. Barsch, M. Sauer, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 17400–17405 (2007).
38. R. B. Best *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 18964–18969 (2007).
39. H. Götzke *et al.*, *Nat. Commun.* **10**, 4403 (2019).
40. K. Kolmakov *et al.*, *Photochem. Photobiol. Sci.* **11**, 522–532 (2012).
41. T. Dechat *et al.*, *Genes Dev.* **22**, 832–853 (2008).
42. Y. Turgay *et al.*, *Nature* **543**, 261–264 (2017).

43. J. Ahn *et al.*, *Nat. Commun.* **10**, 3757 (2019).
44. M. Salvi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 3115–3120 (2017).
45. X. C. Zhang *et al.*, Mechanism of sensor kinase CitA transmembrane signaling. bioRxiv 2023.2002.2006.527302 [Preprint] (2023); https://doi.org/10.1101/2023.02.06.527302.
46. G. T. Dempsey *et al.*, *J. Am. Chem. Soc.* **131**, 18192–18193 (2009).
47. S. Weisenburger *et al.*, *Nat. Methods* **14**, 141–144 (2017).
48. P. D. Dahlberg *et al.*, *J. Am. Chem. Soc.* **140**, 12310–12313 (2018).
49. A. H. Shaib *et al.*, Expansion microscopy at one nanometer resolution. bioRxiv 2022.2008.2003.502284 [Preprint] (2022); https://doi.org/10.1101/2022.08.03.502284.
50. S. J. Sahl, Matlab scripts for: Direct optical measurement of intra-molecular distances with Ångström precision. Zenodo (2024); https://doi.org/10.5281/zenodo.12744422.

SOLAR CELLS

Long-term stability in perovskite solar cells through atomic layer deposition of tin oxide

Danpeng Gao[1]†, Bo Li[1]†, Qi Liu[2]†, Chunlei Zhang[1], Zexin Yu[1], Shuai Li[1], Jianqiu Gong[1], Liangchen Qian[1], Francesco Vanin[1,3], Kelly Schutt[4], Melissa A. Davis[4], Axel F. Palmstrom[4], Steven P. Harvey[4], Nicholas J. Long[3], Joseph M. Luther[4], Xiao Cheng Zeng[2], Zonglong Zhu[1,5]*

Robust contact schemes that boost stability and simplify the production process are needed for perovskite solar cells (PSCs). We codeposited perovskite and hole-selective contact while protecting the perovskite to enable deposition of SnO_x/Ag without the use of a fullerene. The SnO_x, prepared through atomic layer deposition, serves as a durable inorganic electron transport layer. Tailoring the oxygen vacancy defects in the SnO_x layer led to power conversion efficiencies (PCEs) of >25%. Our devices exhibit superior stability over conventional p-i-n PSCs, successfully meeting several benchmark stability tests. They retained >95% PCE after 2000 hours of continuous operation at their maximum power point under simulated AM1.5 illumination at 65°C. Additionally, they boast a certified T_{97} lifetime exceeding 1000 hours.

D espite surpassing the power conversion efficiency (PCE) of many conventional thin-film solar technologies (*1–4*), perovskite solar cells (PSCs) struggle to achieve long-term stability because of fragile interfaces (*5–8*). Some contacts degrade under the combination of various environmental stressors, such as humidity, oxygen, temperature changes, and light exposure, which diminishes cell performance and life span (*9–13*). In inverted p-i-n PSCs, the perovskite–electron transport layer (ETL) interface notably contributes to efficiency losses, forming deep trap states (*14, 15*). The high cost and poor mechanical properties associated with fullerene ETLs drive the search for alternatives.

However, replacing fullerenes with alternatives, such as metal oxides, has been challenging (*16, 17*). The thermal sensitivity of hybrid perovskites is problematic, with high-temperature annealing often used for optimal metal oxide processing. Although sputtered oxides damage the perovskite surface, atomic layer deposition (ALD) for SnO_x directly on perovskite creates chemical reactions and interfacial barriers that lead to devices with PCEs below 1% (*18–20*).

In this work, we redesigned an inverted PSC device architecture with a codeposited perovskite and hole-selective contact deposited as active layers directly on ITO/glass (where ITO is indium tin oxide). The SnO_x prepared through ALD served as a durable inorganic ETL without use of a fullerene. To eliminate detrimental chemical reactions at the perovskite/SnO_x interface (*21*), we introduced aliphatic amine-functionalized perylene-diimide (PDINN) to 3-fluoro-phenethylammonium iodide (m-F-PEAI) surface passivation to provide a one-step passivating and buffering interlayer compatible with ALD deposition.

To enhance device performance, we manipulated the interlayer between the active layers and SnO_x by introducing trace amounts of oxygen vacancies to activate interfacial carriers and facilitate electron extraction, which resulted in a PCE up to 25.1%. The devices also demonstrate excellent stability compared with conventional p-i-n PSCs. Even without encapsulation, the devices perform exceptionally well in various International Summit on Organic and Hybrid Photovoltaics Stability (ISOS) tests, with aging attenuation amplitude less than one-tenth of typical p-i-n devices. Under a prolonged maximum power point tracking (MPPT) stability test at 65°C, the device exhibits a T_{95} aging lifetime exceeding 2000 hours (with third-party certification of a T_{97} stability lifetime surpassing 1000 hours).

[1]Department of Chemistry, City University of Hong Kong, Kowloon 999077, Hong Kong. [2]Department of Materials Science and Engineering, City University of Hong Kong, Kowloon 999077, Hong Kong. [3]Department of Chemistry, Imperial College London, London W12 0BZ, UK. [4]National Renewable Energy Laboratory, Golden, CO 80401, USA. [5]Hong Kong Institute for Clean Energy, City University of Hong Kong, Kowloon 999077, Hong Kong.
*Corresponding author. Email: zonglzhu@cityu.edu.hk
†These authors contributed equally to this work.
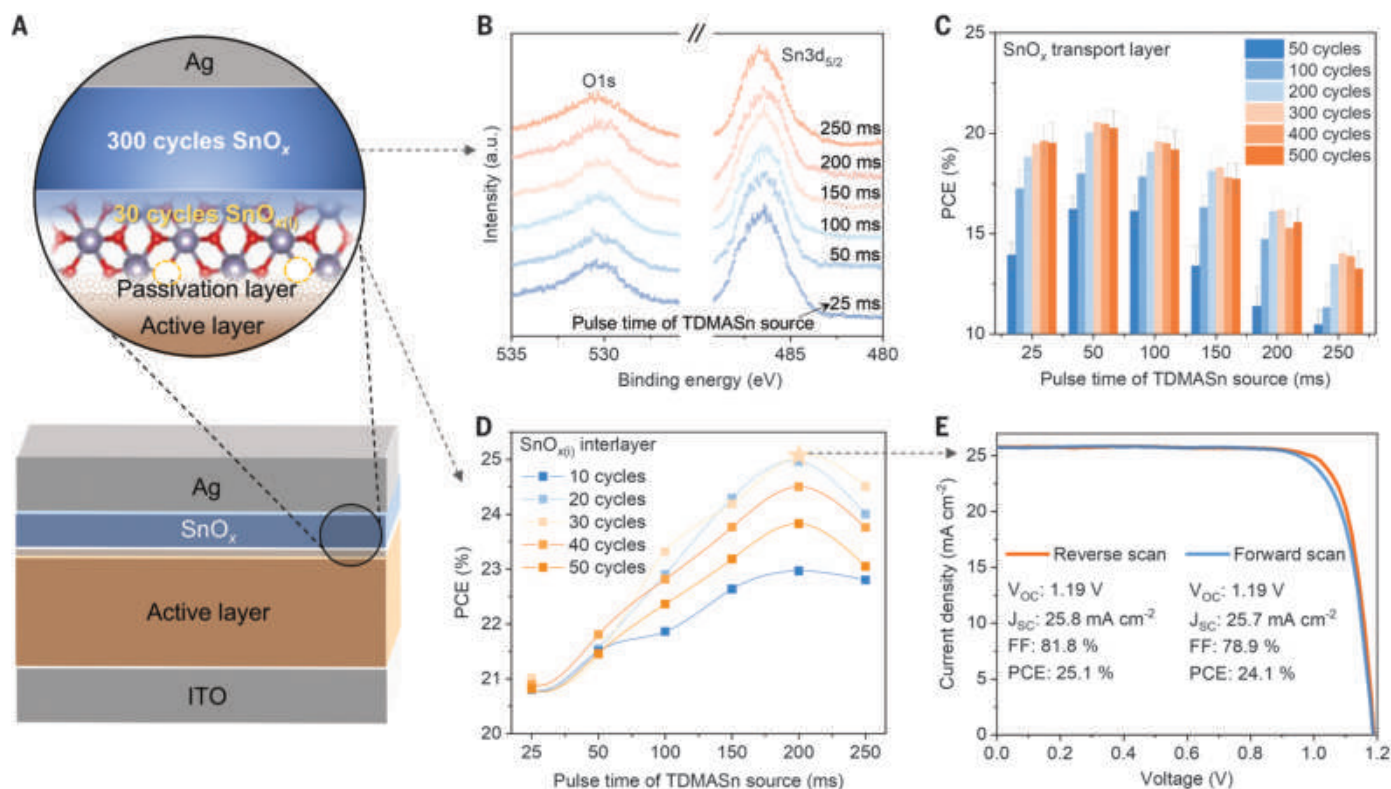
Fig. 1. Device fabrication and performance. (**A**) Schematic illustration of the device structure. (**B**) XPS spectra of SnO_x films obtained using different ALD dosing ratios of TDMASn:H_2O, with H_2O fixed at 20 ms. a.u., arbitrary units. (**C**) The impact of SnO_x transport layer on device performance under different ALD dosing ratios of TDMASn:H_2O (H_2O fixed at 20 ms) and different ALD cycles. (**D**) The effect of $SnO_{x(i)}$ interlayer on device performance under different ALD dosing ratios of TDMASn:H_2O (H_2O fixed at 20 ms) and varied ALD cycles. (**E**) J-V curves of the best-performing SnO_x-based device with the structure of ITO/active layer/ passivation layer/30 cycles $SnO_{x(i)}$ (ALD dosing ratio of TDMASn:H_2O is 200:20) interlayer/300 cycles SnO_x (ALD dosing ratio of TDMASn:H_2O is 50:20)/Ag.

Deposition of SnO_x

The device structure of the PSCs consisted of a simple stack of glass/ITO/active layer/passivation layer/SnO_x/Ag (Fig. 1A). In this structure, the active layers are composed of $Cs_{0.05}FA_{0.95}$-$PbI_{2.94}Br_{0.06}$ (where FA is formamidinium) doped with [4-(3,6-dimethyl-9H-carbazol-9-yl)butyl]phosphonic acid (Me-4PACz), which was prepared by a one-step chemical deposition that has been shown to be effective for constructing inverted PSCs (22). The ETL SnO_x was deposited on the active layer with ALD.

To achieve efficient electron transport, we modulated the tetrakisdimethylamino tin(IV) (TDMASn) to water dosing ratios [TDMASn dose (ms):H_2O dose (ms); 25:20, 50:20, 100:20, 150:20, 200:20, and 250:20] and deposition cycles (from 50 to 500 cycles). As shown by x-ray photoelectron spectroscopy (XPS) in Fig. 1B, figs. S1 to S3, and table S1, SnO_x at the 25:20 ratio was closest to the stoichiometric ratio of SnO_2. As the TDMASn:H_2O ratio increased, the x value of SnO_x gradually decreased.

We further studied the device performance at different TDMASn:H_2O ratios with varying deposition cycles (Fig. 1C and figs. S4 to S12), and the passivation layer was also optimized (figs. S13 and S14). The investigation revealed champion performance with an ALD TDMASn:

H_2O dose ratio of 50:20 and a deposition of 300 cycles. Specifically, glass/ITO/active layers/300 cycles SnO_x (Sn:O=50:20)/Ag, achieved a PCE of 21.8% [as shown in fig. S15; open-circuit voltage (V_{OC}) of 1.16 V, short-circuit current density (J_{SC}) of 25.5 mA/cm^2, and fill factor (FF) of 73.7%].

We noticed a FF deficit for the device using SnO_x as the ETL compared with the C_{60}-based device (fig. S16). A potential cause of the FF loss is the transport resistance, although we found no notable difference between them by comparing the conductivity of C_{60} and SnO_x in fig. S17. Moreover, ultraviolet-visible (UV-vis) spectroscopy and ultraviolet photoelectron spectroscopy (UPS) in figs. S18 and S19 showed that the conduction band minimum (CBM) of SnO_x matched that of the active layers better compared with commercial C_{60}, which suggests that the FF loss was independent of the energy misalignment between the ETL and the perovskite active layer.

In addition to energy level alignment and transport capabilities, the interface between the perovskite and ETL can also determine device performance by affecting carrier extraction efficiency and interfacial defect chemistry. In this case, we further engineered the interface of SnO_x in contact with the perovskite. Before preparing the SnO_x ETL, we deposited

an interlayer, named $SnO_{x(i)}$, with different TDMASn:H_2O ratios and cycle numbers for 10 to 50 cycles. When the TDMASn:H_2O ratio of the $SnO_{x(i)}$ interlayer is 200:20 and the deposition 30 cycles (Fig. 1D and figs. S20 and S21; specific process curve shown in fig. S22), the device achieved the highest PCE of 25.1%.

The current density–voltage (J-V) characteristics of the champion device (Fig. 1E) with reverse and forward scans yielded PCEs of 25.1% (V_{OC} of 1.19 V, J_{SC} of 25.8 mA/cm^2, and FF of 81.8%) and 24.1% (V_{OC} of 1.19 V, J_{SC} of 25.7 mA/cm^2, and FF of 78.9%), respectively. The corresponding stable power output reached up to 24.95% (as shown in fig. S23). Corresponding external quantum efficiency (EQE) spectra in fig. S24 yielded integrated J_{SC} with negligible variation from the values obtained from J-V measurements. Moreover, the process was reproducible, as can be seen from the statistical distribution of all photovoltaic parameters for 20 devices in fig. S25. The performance comparison of different structural devices involved in this study is shown in fig. S26.

Interlayer characterization

To study the specific effects of the interlayer, we performed electroluminescence quantum efficiency (EQE-EL) and defect density measurements on the devices obtained by depositing the
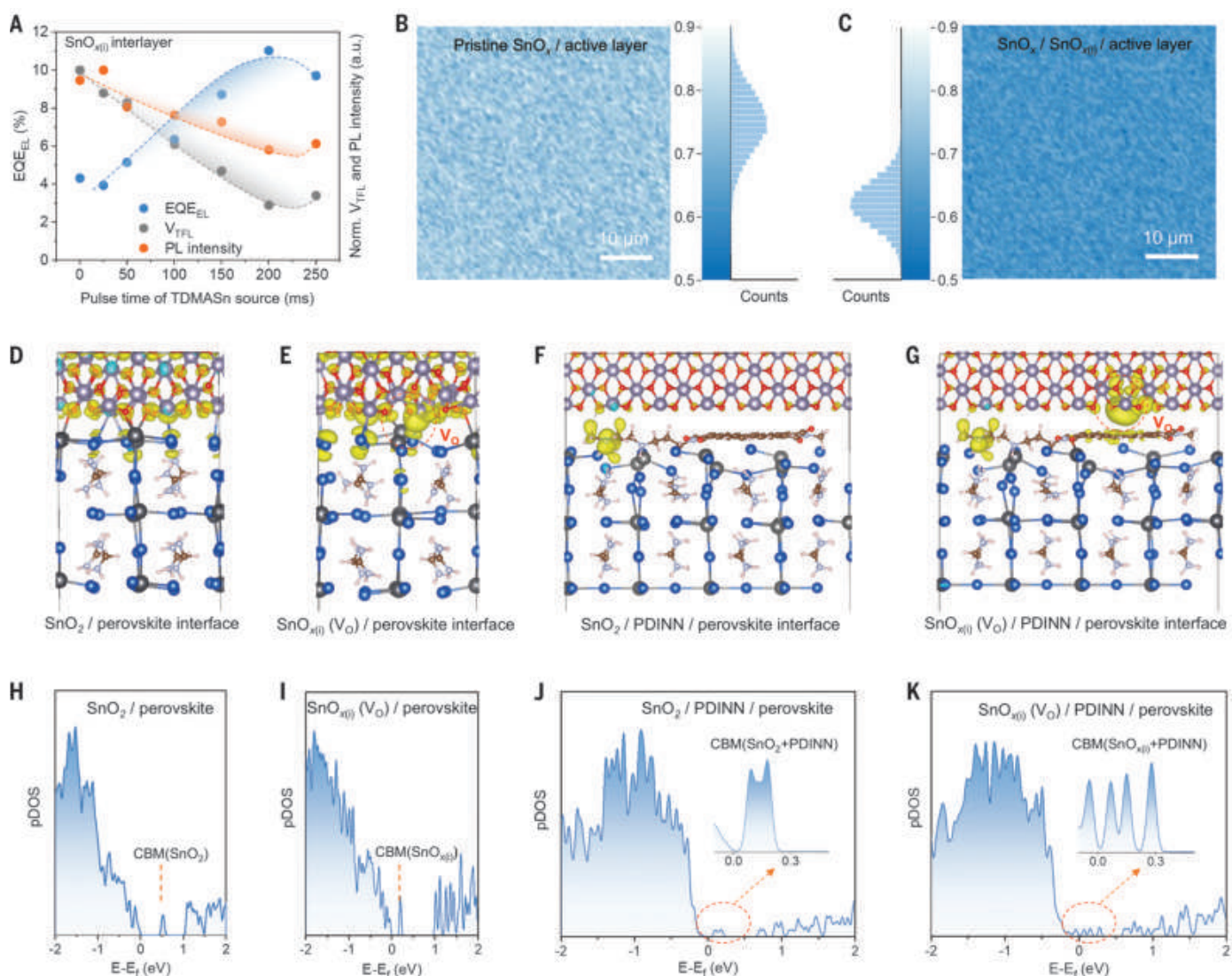
Fig. 2. Interlayer modulation. (**A**) EQE-EL, V_{TFL}, and PL intensity of the $SnO_{x(i)}$ interlayer obtained with different pulse time of TDMASn source (the pulse time of H_2O sources is fixed at 20 ms). (**B** and **C**) PL mapping (incident from the SnO_x side) of the films, including the pristine SnO_x/active layer (B) and the $SnO_x/SnO_{x(i)}$/active layer (C). (**D** to **G**) Illustration of the simulated charge distribution (corresponding to CBM) at different interfaces, including the pristine SnO_2/perovskite interface (D), the $SnO_{x(i)}$ (V_O)/perovskite interface (E), the SnO_2/PDINN/perovskite interface (F), and the $SnO_{x(i)}$ (V_O)/PDINN/ perovskite interface (G). (**H** to **K**) Computed pDOS plots of the pristine SnO_2/perovskite interface (H), the $SnO_{x(i)}$ (V_O)/perovskite interface (I), the SnO_2/ PDINN/perovskite interface (J), and the $SnO_{x(i)}$ (V_O)/PDINN/perovskite interface (K).

$SnO_{x(i)}$ interlayer with different $TDMASn:H_2O$ ratios, as well as photoluminescence (PL) characterization on the active layer/ETL films (Fig. 2A). Compared with the device without the interlayer, the EQE-EL value increased with the increase of the $TDMASn:H_2O$ ratio of the $SnO_{x(i)}$ interlayer, from 3.9% at the $TDMASn:H_2O$ ratio of 25:20 to 11.0% at the $TDMASn:H_2O$ ratio of 200:20 (fig. S27), which we attributed to reduced nonradiative recombination and energy loss (23–25).

Moreover, the trap-filled limit voltage (V_{TFL}) and PL intensity showed opposite trends. Both decreased gradually with the increase of the $TDMASn:H_2O$ ratio of the $SnO_{x(i)}$ interlayer (specific results are shown in figs. S28 and

S29), which suggests that the oxygen-deficient condition in the $SnO_{x(i)}$ interlayer suppressed defects and promoted carrier extraction (26–28). Additionally, PL mapping on the SnO_x/active layers film and the $SnO_x/SnO_{x(i)}$/active layers film, respectively (Fig. 2, B and C), revealed a wide PL intensity distribution for the SnO_x/active layers indicative of uneven charge extraction. By contrast, the introduction of the $SnO_{x(i)}$ interlayer led to a decrease in PL intensity, and the film showed more uniform PL emission, which further verifies that the carrier extraction in $SnO_x/SnO_{x(i)}$/active layers was improved and homogenized (29–32). The surface morphology of the films treated with SnO_x and $SnO_{x(i)}$, the cross section image of the device, and the thick-

ness of the SnO_x and $SnO_{x(i)}$ films are shown in figs. S30 to S33.

Theoretical studies

We performed density functional theory (DFT) calculations on various interfaces involving SnO_x and perovskite. The calculations were based on the uncharged heterostructures at the level of Perdew-Burke-Ernzerhof (PBE) + spin-orbit coupling (SOC) as implemented in Vienna Ab initio Simulation Package (VASP) 6.4 (33–35). The results are presented in Fig. 2, D to K, and the detailed data are shown in figs. S34 to S37, where we compared the charge distribution and partial density of states (pDOS) at the interfaces, both with and without the
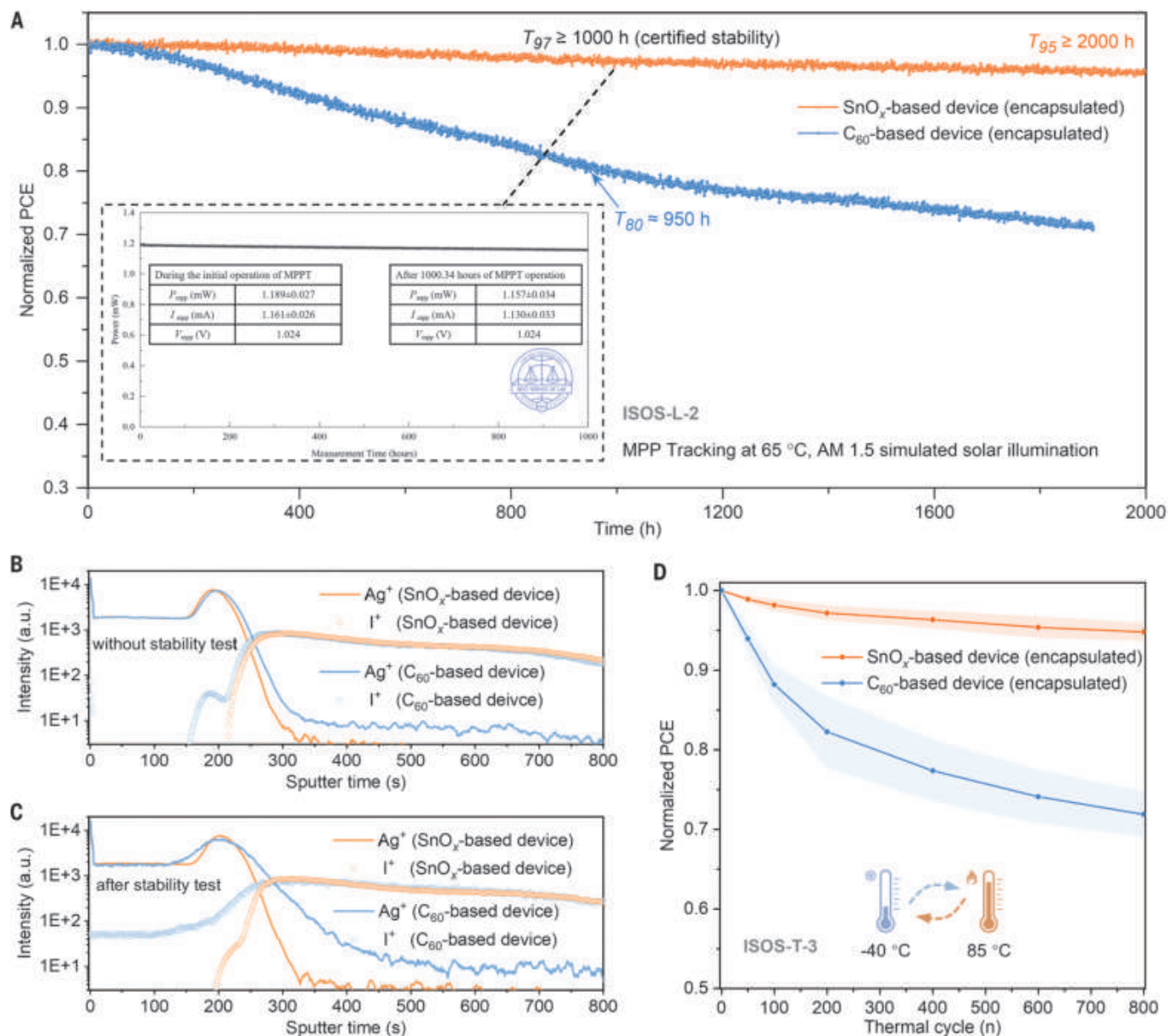
Fig. 3. Operational stability. (**A**) Operational stability of devices at temperature of 65°C. (**B**) TOF-SIMS characterization of original device without stability test. (**C**) TOF-SIMS characterization of device after the operational stability test in (A). (**D**) Devices' efficiency evolution under repeated thermal cycling (−40°C to 85°C) in the dark in air (ISOS-T-3). We conducted thermal cycling tests using eight individual devices and obtained the average performance change along with the SD as the error measure.

$SnO_{x(i)}$ (V_O) slab and the PDINN buffering layer (where V_O refers to the oxygen vacancy). The selection of both pristine SnO_2 and $SnO_{x(i)}$ (V_O) with oxygen vacancies is to clarify the effect of the formation of oxygen vacancies (i.e., $TDMASn:H_2O$ ratios from 25:20 to 200:20) on the charge carriers. For the PDINN molecule, which may give very large supercell and strong lattice mismatch, we keep only one amino side chain instead of the two side chains in the original designation—on the other side, we adopt a methyl group instead.

The external-adjacent Sn^{4+} and I^- usually have a high degree of chemical reactivity, especially when they are directly combined with

hetero components at interface. The formation of V_O structure (Fig. 2E) can be considered a result of partial reduction of Sn^{4+} cations (*36*), which exhibits higher spatial charge density distribution at the conducting band (CB) edge compared with the pristine SnO_2 (Fig. 2D). With the strong interaction between the perovskite and SnO_2 layers, the fiercely distorted Pb-I frameworks tend to directly bind with the SnO_2 layer (as shown in fig. S34). As such, the interfacial perovskite structures bonded with SnO_2 have the same CB edge level as SnO_2 (Fig. 2, D and H, and figs. S35A and S36A), removing the band offset between SnO_2 and perovskite layers while weakening

the carrier transfer trend between the interface. Additionally, with the V_O defects, the strong surface interaction further induces the bandgap downshift between the CB edge of $SnO_{x(i)}$ and the valence band (VB) edge of the perovskite (Fig. 2, E and I, and figs. S35B and S37A).

With the introduction of a PDINN molecular buffer layer (as shown in Fig. 2F), it is remarkable that there is no effective spatial charge distribution at the interface within the planar polycyclic aromatic backbone of PDINN in the vacancy-free structure while only partial effective spatial charge distribution arises at the amino side chains. Similarly, the pDOS in Fig. 2J

Fig. 4. Environmental stability. (**A**) Devices' efficiency evolution under storage in dark under indoor air conditions (ISOS-D-1). (**B**) Devices' efficiency evolution under damp heat testing at 85°C and 85% RH in the dark in air (ISOS-D-3). (**C**) Devices' efficiency evolution under light on-off cycle test (12 hours–12 hours) using light-emitting diode (LED) lamps simulated 1-sun illumination (ISOS-LC-1). (**D**) Devices' efficiency evolution under storage in sunlight under outdoor condition (ISOS-O-1). (The devices are in an open-circuit state. Environmental data are quoted from the Hong Kong Observatory, https://www.hko.gov.hk/tc/wxinfo/pastwx/mws2023/mws202309.htm.) We conducted stability tests for each condition [(A) to (D)] using 10 to 14 individual devices and obtained the average performance change along with the SD as the error measure.
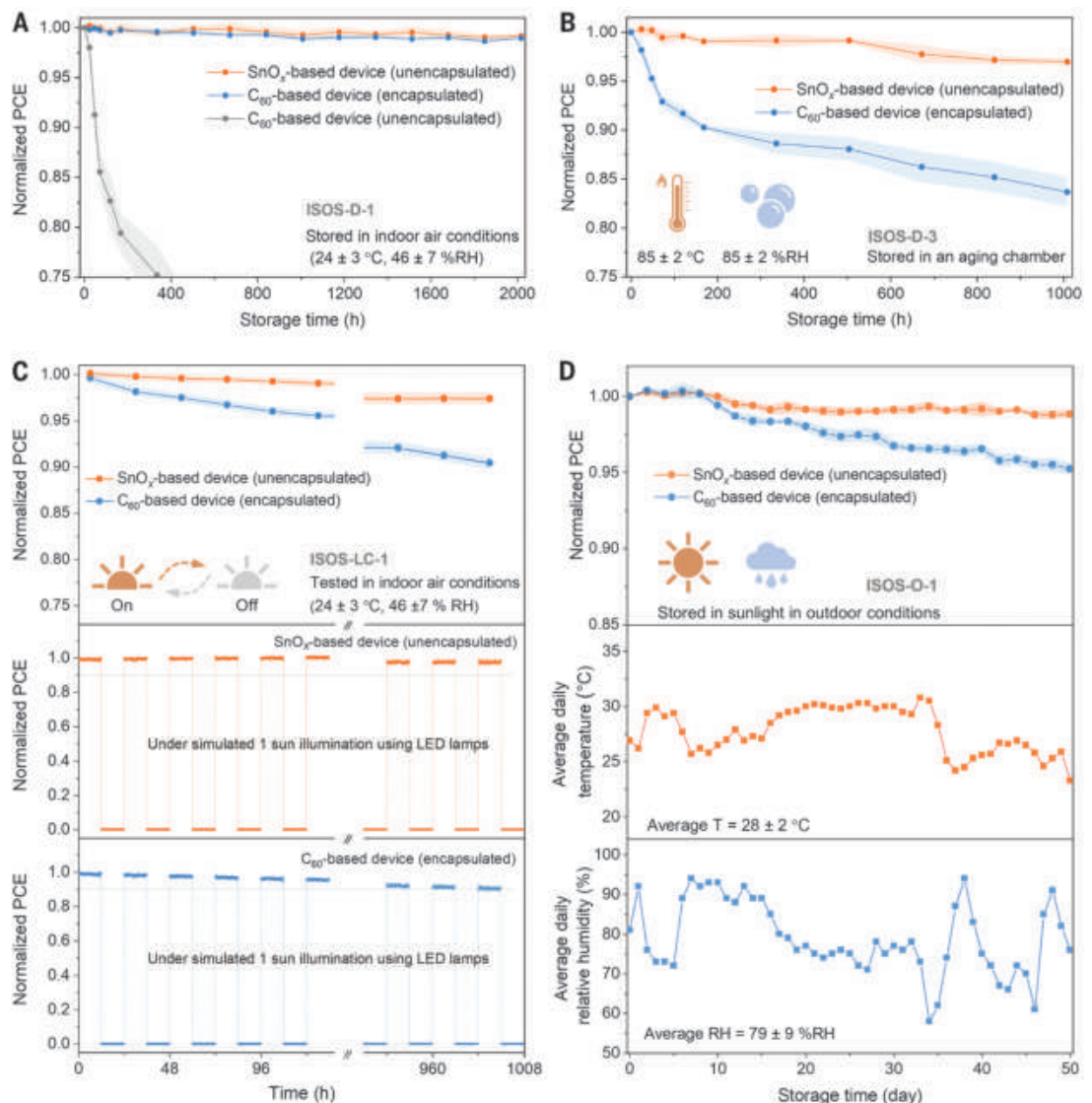
and figs. S35C and S36 shows a substantial decrease in the DOS peak at the edge of the CBM, indicating a reduced possibility of carrier distribution in this region. This result suggests that polycyclic aromatic backbones of PDINN do not promote charge transport but rather act as a buffering agent for the ALD process owing to its hydrophobic properties, which can prevent H_2O from directly contacting and reacting with the halide perovskite as demonstrated by XPS analyses in fig. S38.

For the V_O-existence $SnO_{x(i)}$/PDINN/perovskite interface (Fig. 2G), the presence of oxygen vacancies leads to an appearance of spatial charge density contribution of $SnO_{x(i)}$ adjacent with the polycyclic aromatic part of PDINN. Similarly, the CBM peaks in pDOS (Fig. 2K and figs. S35D and S37) exhibit stronger dispersion than those in the vacancy-free benchmark. Without the direct contact between the $SnO_{x(i)}$ and perovskite layers due to the PDINN layer, the strongly dispersed CBM peak ranges

result in the reduction of the negative "cliff-like" band offset compared with the benchmark (Fig. 2, F and J), which further reduces the energy loss during carrier extraction and transfer (37–39). These first-principles computational results collectively suggest that the appearance of the V_O defects increased the carrier transfer, thereby activating the $SnO_{x(i)}$/PDINN/perovskite interface by inducing carrier distribution into the polycyclic aromatic backbone parts of the PDINN layer. This conclusion is also applicable to the mF-PEAI and PDINN dual-additive interface because PDINN dominates the charge density contribution owing to its stronger conjugation effect compared with the isolated single aromatic ring of mF-PEAI, as proven in fig. S39.

Operational stability

The encapsulated devices were placed in air under continuous 1-sun illumination (following the ISOS-L-2 protocol), and the performance of

the devices was monitored at 65°C under MPPT. The SnO_x-based devices show excellent stability, maintaining >95% of the initial PCE after 2000 hours of continuous operation ($T_{95} \geq$ 2000 hours; Fig. 3A). Moreover, to obtain certified stability results, the SnO_x-based devices were sent to a third-party institution for stability testing and demonstrated an exceptional stability of >97% of the initial PCE after 1000 hours of continuous operation (shown in the dashed box in Fig. 3A and fig. S40). Table S2 details a favorable comparison with other recently reported stability results. By contrast, the control device maintains a relatively stable performance in the first 100 hours, but the PCE dropped to 80% of the initial value after 950 hours of continuous operation. Similar trends were observed even during accelerated testing at 85°C and under 1-sun illumination (fig. S41).

To gain insight into the difference in device operational stability, we performed time-of-flight secondary ion mass spectrometry (TOF-SIMS)

tests on the devices before and after the stability tests. In Fig. 3, B and C, for the untested control device, a small amount of I^- ions penetrated into the C_{60}/BCP layer and contacted the Ag electrode. After the stability test, the diffusion of I^- ions intensified, manifested as a large number of I^- ions penetrating into the C_{60}/BCP layer and entering the Ag electrode (figs. S42 and S43). For the SnO_x-based device, under the protection of SnO_x, I^- ions did not show any obvious diffusion in the device before and after aging (figs. S44 and S45). This result indicates that the photothermal stability of the SnO_x-based device comes from the protection of the SnO_x layer to the perovskite layer, which makes it difficult to undergo ion diffusion and potential phase separation under the dual influence of high temperature and light.

The thermal cycling test, which is a common test used by the International Electrotechnical Commission (IEC) for outdoor conditions and commercialization, was also performed. For the test (−40°C to 85°C, dark; ISOS-T-3), the SnO_x-based devices showed an average degradation of ~4.7% after 800 thermal cycles (as shown in Fig. 3D), which exceeded the 200 thermal cycles usually performed by this test protocol. The control devices showed an average efficiency reduction of 28.1% after experiencing the same conditions.

Environmental stabilities were studied on the SnO_x-based PSCs and conventional inverted PSCs (referred to as control devices) according to multiple sets of standard environmental stability tests, including ISOS-D-1, ISOS-D-3, ISOS-LC-1, and ISOS-O-1. For the ISOS-D-1 stability test shown in Fig. 4A and fig. S46A, after storing for 2000 hours at 23° ± 4°C and 46 ± 7% relative humidity (RH), both encapsulated and unencapsulated SnO_x-based devices showed almost no performance degradation and maintained >99.1% of their initial PCE for 2000 hours. The unencapsulated control device exhibited obvious performance degradation (the PCE dropped to 76.4% of the initial value after 300 hours), whereas the encapsulated control device maintained 98.9% of the initial PCE after storing for 2000 hours. The result indicates that the SnO_x layer can effectively diminish water and oxygen damage around the perovskite, effectively achieving self-encapsulation of the devices.

We further performed high-temperature and high-humidity storage (ISOS-D-3) at 85°C and 85% RH. As shown in Fig. 4B, after storing for 1000 hours, the unencapsulated SnO_x-based device showed a decay of <3.0%. Moreover, encapsulating it did not appreciably improve its stability (the attenuation was reduced from 3.0% to 2.1% in fig. S46B). In comparison, the encapsulated control device showed a decay of >16.4%, indicating that external encapsulation alone was insufficient to resist the invasion of extreme environments. In addition, we performed light on-off stability experiments (ISOS-LC-1) in indoor environments. As shown in Fig. 4C, after 42 cycles of 12 hours–12 hours light on-off test at 23° ± 4°C and 46 ± 7% RH, the unencapsulated SnO_x-based device showed a PCE decay of 2.5%, whereas the encapsulated control device attenuates by 9.7%. We further studied the stability of the devices under outdoor aging conditions (ISOS-O-1).

Outdoor aging tests

Because outdoor temperature, relative humidity, and solar radiation intensity all affect the stability of the devices, we tracked the changes of these parameters during the outdoor aging test (Fig. 4D and fig. S47). After placing in the outdoor environment for 50 days, the unencapsulated SnO_x-based device shows a decay of only 1.1%, indicating that the self-encapsulated SnO_x-based device is almost unaffected by the outdoor environment. In comparison, the encapsulated control device shows a relatively large attenuation of 4.8%. The stability results validate the self-encapsulation mechanism of SnO_x devices. It inhibits external moisture and oxygen penetration, extending device lifetime. It also suppresses ion diffusion from the perovskite layer to the ETL or electrode, maintaining device performance (fig. S48A). By contrast, control devices with an additional encapsulation layer experience adhesive aging, potentially leading to encapsulation failure (fig. S48B). The photos of the original device and the encapsulated device are shown in fig. S49. The encapsulation layer above the electrode fails to prevent ion diffusion that may occur during long-term operation.

Discussion

We demonstrated the effectiveness of a simplified fullerene-free PSC device structure featuring SnO_x by modulating oxygen vacancy defect concentration in the interlayer. We reveal that the fine-tuning of oxygen defects reduces the negative cliff-like band offset between perovskites and ETLs, thereby decreasing energy losses during carrier extraction and transfer. The modulated device exhibits excellent efficiency up to 25.1% and superior stability. The attenuation amplitude of unencapsulated devices in various standard stability tests, such as ISOS-D-1, ISOS-D-3, ISOS-LC-1, and ISOS-O-1, is less than one-tenth of that of traditional p-i-n devices. Under long-term MPPT stability tests at 65°C, the T_{95} aging life of the device exceeds 2000 hours (third-party certified T_{97} stable life exceeds 1000 hours). This simplified, efficient, and stable device architecture provides insight into the development of low-cost and reproducible PSCs.

REFERENCES AND NOTES

1. H. Chen et al., Science 384, 189–193 (2024).
2. Z. Li et al., Science 376, 416–420 (2022).
3. H. Gao et al., Science 383, 855–859 (2024).
4. National Renewable Energy Laboratory, "Best Research-Cell Efficiency Chart"; https://www.nrel.gov/pv/cell-efficiency.html.
5. H. Zhu et al., Nat. Rev. Mater. 8, 569–586 (2023).
6. L. Duan et al., Nat. Rev. Mater. 8, 261–281 (2023).
7. B. Li et al., Chem 10, 35–47 (2024).
8. S. Zhang et al., Adv. Energy Mater. 10, 2001610 (2020).
9. P. Chen et al., Nature 625, 516–522 (2024).
10. S. M. Park et al., Science 381, 209–215 (2023).
11. Z. Li et al., Science 382, 284–289 (2023).
12. Q. Jiang et al., Nature 623, 313–318 (2023).
13. R. Azmi et al., Science 376, 73–77 (2022).
14. J. Warby et al., Adv. Energy Mater. 12, 2103567 (2022).
15. A. A. Said et al., Nat. Commun. 15, 708 (2024).
16. J. P. Bastos et al., ACS Appl. Mater. Interfaces 8, 9798–9805 (2016).
17. Z. Li et al., Joule 3, 1559–1572 (2018).
18. A. E. A. Bracesco et al., ACS Appl. Mater. Interfaces 15, 38018–38028 (2023).
19. N. Mallik et al., Nano Energy 126, 109582 (2024).
20. A. Hultqvist et al., ACS Appl. Energy Mater. 4, 510–522 (2021).
21. F. A. Palmstrom et al., Adv. Energy Mater. 8, 1800591 (2018).
22. X. Zheng et al., Nat. Energy 8, 462–472 (2023).
23. G. Yang et al., Nat. Photonics 15, 681–689 (2021).
24. J. Zhu et al., Nat. Energy 8, 714–724 (2023).
25. W. Peng et al., Science 379, 683–690 (2023).
26. J. Wang et al., Nat. Commun. 11, 177 (2020).
27. Q. Jiang et al., Nat. Energy 2, 16177 (2016).
28. D. Yang et al., Nat. Commun. 9, 3239 (2018).
29. Y. Bai et al., Science 378, 747–754 (2022).
30. C. Liu et al., Nat. Commun. 12, 6394 (2021).
31. A. Krishna et al., Energy Environ. Sci. 14, 5552–5562 (2021).
32. T. Zhang et al., Joule 2, 2706–2721 (2018).
33. G. Kresse, J. Non-Cryst. Solids 192–193, 222–229 (1995).
34. P. E. Blöchl, Phys. Rev. B 50, 17953–17979 (1994).
35. J. P. Perdew, K. Burke, M. Ernzerhof, Phys. Rev. Lett. 77, 3865–3868 (1996).
36. Q. Liu, M.-G. Ju, W. Liang, Phys. Chem. Chem. Phys. 22, 20553–20561 (2020).
37. H. Chen et al., Nature 613, 676–681 (2023).
38. Z. Yu et al., Nat. Energy 5, 657–665 (2020).
39. M. Stolterfoht et al., Energy Environ. Sci. 12, 2778–2788 (2019).

SUPPLEMENTARY MATERIALS

RANGE SHIFTS

Unexpected westward range shifts in European forest plants link to nitrogen deposition

Pieter Sanczuk[1]*, Kris Verheyen [1], Jonathan Lenoir[2], Florian Zellweger[3], Jonas J. Lembrechts[4,5], Francisco Rodríguez-Sánchez[6], Lander Baeten[1], Markus Bernhardt-Römermann[7,8], Karen De Pauw[1], Pieter Vangansbeke[1], Michael P. Perring[9,10], Imre Berki[11], Anne D. Bjorkman[12,13], Jörg Brunet[14], Markéta Chudomelová[15], Emiel De Lombaerde[1]†, Guillaume Decocq[2], Thomas Dirnböck[16], Tomasz Durak[17], Caroline Greiser[18,19], Radim Hédl[15,20], Thilo Heinken[21], Ute Jandt[8,22], Bogdan Jaroszewicz[23], Martin Kopecký[24,25], Dries Landuyt[1], Martin Macek[24], František Máliš[26,27], Tobias Naaf[28], Thomas A. Nagel[29], Petr Petřík[15,30], Kamila Reczyńska[31], Wolfgang Schmidt[32], Tibor Standovár[33], Ingmar R. Staude[8,34], Krzysztof Świerkosz[35], Balázs Teleki[36], Thomas Vanneste[1], Ondrej Vild[15], Donald Waller[37], Pieter De Frenne[1]

Climate change is commonly assumed to induce species' range shifts toward the poles. Yet, other environmental changes may affect the geographical distribution of species in unexpected ways. Here, we quantify multidecadal shifts in the distribution of European forest plants and link these shifts to key drivers of forest biodiversity change: climate change, atmospheric deposition (nitrogen and sulfur), and forest canopy dynamics. Surprisingly, westward distribution shifts were 2.6 times more likely than northward ones. Not climate change, but nitrogen-mediated colonization events, possibly facilitated by the recovery from past acidifying deposition, best explain westward movements. Biodiversity redistribution patterns appear complex and are more likely driven by the interplay among several environmental changes than due to the exclusive effects of climate change alone.

One of the most prominent biogeographical changes of the 21st century is the large-scale redistribution of plants and animals in response to changes in the climate system (*1*). Warming temperatures are causing many terrestrial species to move toward higher latitudes and elevations, which results in a reordering of species' distributions (*1–3*) and the emergence of novel communities (*4*). Empirical evidence has been reported for a wide range of ecosystems and taxa—from poleward and upslope range shifts in temperate regions (*5, 6*) and high-latitude boreal biomes (*7*) to upslope shifts in mountain vegetation (*8*)—which suggests an emerging link with anthropogenic climate warming (*2, 9*).

According to the most recent global synthesis (*1*), terrestrial species are shifting toward higher latitudes at an average rate of 1.11 km year⁻¹. This trend, however, lacks statistical significance, possibly because estimates are often blurred by variation in methodological attributes (*1, 10*). Alternatively, species' redistributions in geographical directions that are orthogonal (i.e., west–east oriented) or even inverse (e.g., equatorward) to the moving isotherms are less likely to be detected from commonly studied range boundary shifts along thermal transects of latitude and elevation alone (*11, 12*). Indeed, other prominent environmental changes such as atmospheric (nitrogen and sulfur) deposition and forest disturbances show spatial patterns that are weakly correlated to the geographic direction of climate change (*13–15*) and can also influence demographic processes of colonization and local extinction (*1, 12, 16*). To what extent these other environmental changes are contributing to species range shifts remains largely unquantified (*17–19*).

Here, we quantify the rate and geographic direction of range shifts in 266 European forest understory plant species using multidecadal community data collected in mature forest stands across 2954 resurveyed semi-permanent vegetation plots (*20*) (Fig. 1A). Plant community data were derived from baseline surveys recorded between 1933 and 1994 and paired resurveys carried out after the baseline surveys between 1987 and 2017 [median (minimum to maximum) intersurvey interval: 39 (13 to 67) years]. We quantified the shifts of species' distributions within the spatial extent of the study area on the basis of range centroids, i.e., the abundance-weighted geometric center of a species' distribution (fig. S1). In contrast to the more frequent quantification of range boundary shifts at the trailing or leading edges, analyzing centroid shifts allows us to obtain more robust estimates of the magnitude and geographic direction of complex distribution shifts (*6, 11, 21*). This is important because range shift estimates from leading and trailing edges alone are more prone to bias from stochastic processes or low sample sizes that may blur overall biogeographical trends (*11*).

The rate and geographic direction of centroid shifts

To calculate the centroid shift of each species, we first located the position of the abundance-weighted centroid at the time of the baseline survey and the resurvey and assessed the magnitude (i.e., the distance) and geographical direction (i.e., the bearing) of the centroid shift over time. Centroid shifts were expressed as the absolute shift rate (km year⁻¹) as well as the projected south–north (km north year⁻¹) and west–east (km east year⁻¹) rate (schematically explained in fig. S1). Centroid shifts were calculated for the 266 species that were

Fig. 1. Spatial and environmental gradients of the study. (**A**) Map of modeled nitrogen (N) deposition rate (sum of oxidized and reduced wet and dry deposition expressed in kg N ha^{-1} year^{-1}; dry deposition accounted for deciduous forest surface) at 0.1° resolution for the reference year 2000 and distribution of the 2954 resurveyed vegetation plots (gray crosses, spatially jittered for clarity) across the European temperate forest biome (shaded green background) (20). (**B**) Observed environmental variation across the 2954 vegetation plots (gray dots) of three key drivers of forest biodiversity over the course of the study period investigated here: nitrogen deposition rate (total of oxidized and reduced wet and dry N), climate-change velocity (realized changes of both temperature and precipitation, expressed in km north year^{-1}), and the rate of canopy cover change (average annual rate; expressed in % canopy cover increase year^{-1}). In all boxplots, we present the median (horizontal line), first and third quantile (lower and upper hinges), and 1.5 times the interquartile range (whiskers). Half violin plots represent the density distributions of the environmental change values. The gray dashed lines represent no changes (not shown for nitrogen deposition rates). Negative values in the case of climate and canopy cover change indicate southward velocities and canopy opening, respectively.
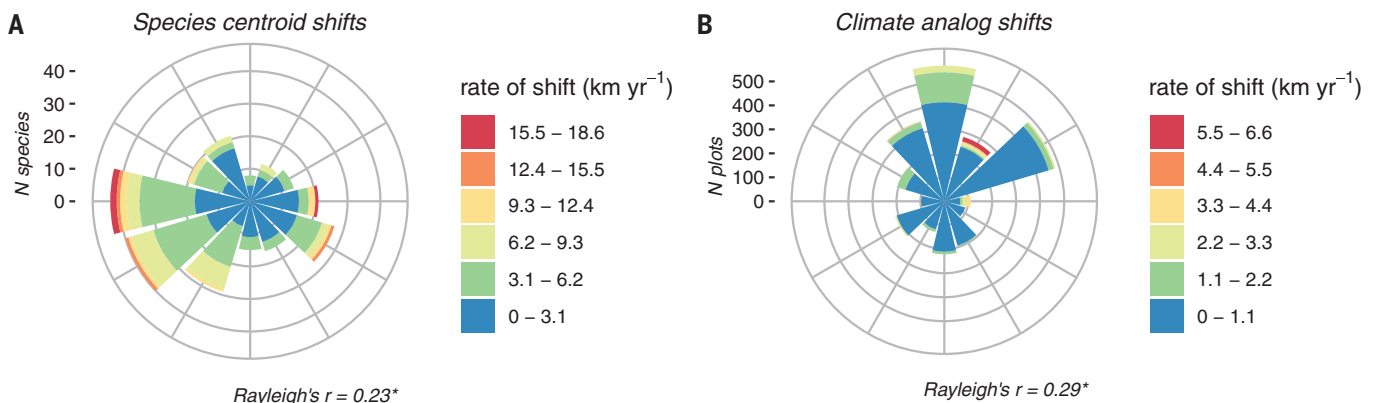
Fig. 2. Rate and geographic direction of species centroid shifts and climate analog shifts. (**A**) Rate and geographic direction of species range centroid shifts (*n* species = 266). (**B**) Velocity and geographic direction of climate analog shifts (*n* plots = 2,954). In all graphs, the Rayleigh's *r* statistic represents a test of uniformity that compares the bearings of shifts to a uniform circular distribution (null hypothesis). Larger values indicate more directional shifts. Asterisk (*) indicates significant deviations from the null hypothesis (*P* < 0.05). See figs. S3 and S4 for results of the analyses including rare species and per biogeographic region.

recorded in ≥1% of the plots to increase robustness of the estimates. The directionality (i.e., angular dispersion of the directions of centroids shifts) across all species was tested by using the Rayleigh's *r* coefficient, a circular regression coefficient that quantifies how uniform and isotropic the directions of shifts are (Rayleigh's *r* = 1 if all species are moving in

the same direction, whereas Rayleigh's *r* = 0 with random directional movements, i.e., anisotropic, meaning that directions of shifts can be drawn from a uniform circular distribution.

Species' centroid shifts were first compared to the velocity and direction of climate change realized over the course of the study period. Spatially explicit climate-change velocities were

calculated by climate analog mapping (22, 23), an approach that is theoretically equivalent to the mapping of species centroid shifts (fig. S1). In contrast to the frequent calculation of climate-change velocities on the basis of gradients of isotherms alone (3, 24), climate analog mapping allows us to consider consolidated changes of multiple bioclimatic variables at
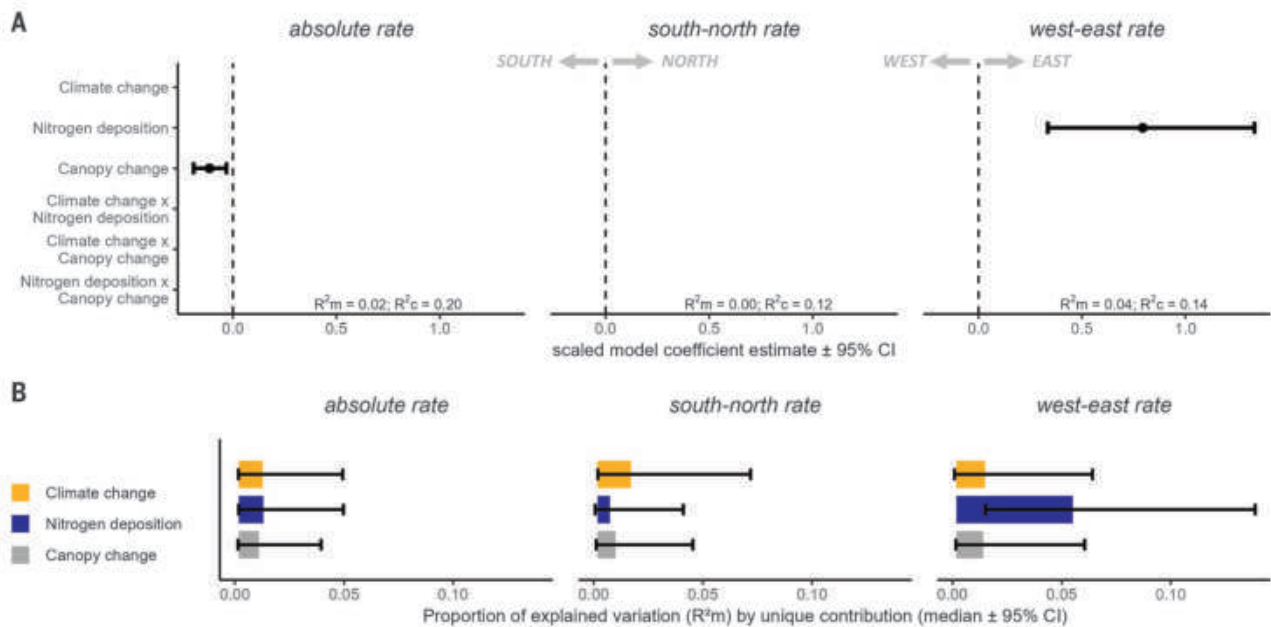
Fig. 3. Effects of environmental changes on centroid shifts based on the most parsimonious model. (**A**) Results of the mixed-effects models (*n* species = 266) indicating coefficient estimates and 95% confidence intervals (CIs) of the effects of the velocity of climate change (km year^{-1}, km north year^{-1}, km east year^{-1}), average nitrogen deposition rate between the baseline survey and resurvey (kg N ha^{-1} year^{-1}), and rate of canopy change (% cover increase year^{-1}), as well as all pairwise interactions (indicated with "×") on the absolute rate of centroid shifts (km year^{-1}) and the south–north rate (km north year^{-1}; negative values indicate southward shifts) and west–east rate (km east year^{-1}; negative values indicate westward shifts) extracted from the most parsimonious model structure (empty rows were not included in the final selected model). All the predictor variables were *z*-transformed to increase comparability. Rates of canopy opening (negative values of canopy change) are associated with greater absolute rates of centroid shifts. The west–east rate was exclusively linked to nitrogen deposition, with faster westward shifts in species with lower rates of nitrogen deposition across their distributions. Model fit is presented as the proportion of variation explained by the fixed effect (marginal R^2, R^2_m) and the proportion of variation explained by the fixed and random effects (conditional R^2, R^2_c). Models accounted for plant growth form as random intercept (five levels: forbs, graminoids, pteridophytes, shrubs, and trees). (**B**) Results of the variation partitioning analyses representing the individual contribution of each environmental predictor. Bar plots are proportional to the variation explained by the contribution of each fixed effect (expressed as R^2_m). In all graphs, estimates and error bars represent the median value and 2.5 to 97.5 percentiles across 1000 bootstrap samples. See fig. S7 for results on the analyses that also included rare species and fig. S8 for a direct comparison with the estimated effects of acidifying deposition.

the same time. For example, we simultaneously considered changes in maximum growing-season temperatures, minimum winter temperatures, and growing-season precipitation as one measure of the climate-change velocity between the baseline survey and resurvey periods (fig. S2). This is highly relevant because plants respond not only to warming temperatures, but also to alterations in precipitation regimes. In this method, for all resurveyed vegetation plots, a grid search (at 4 km by 4 km resolution) was performed to map all raster cells within the study area in which the climatic conditions in the resurvey period are similar (i.e., show no statistical difference) to a given plot's climate during the baseline period (i.e., "analog climate conditions"). For each plot, we then located the position of the nearest raster cell with analog climatic conditions to calculate the velocity and geographic direction of climate change over time. Identical to the centroid shifts, the climate-change velocity for each plot was expressed as the absolute shift velocity (km year^{-1}), as well as the projected south–north (km north year^{-1}) and west–east (km east year^{-1}) velocity. The directionality of climate analog shifts was tested by using Rayleigh's *r* coefficient as described above.

Centroid shifts across the 266 understory plant species varied from 0.006 km year^{-1} (*Symphytum cordatum*) to 18.27 km year^{-1} (*Abies alba* seedlings) and occurred at a mean rate of 3.56 (5% to 95% quantile: 0.39 to 9.80) km year^{-1} (Fig. 2A and data S1). Unexpectedly, two-thirds of the studied plant species showed directional shifts along the west–east axis (Rayleigh's *r* = 0.23; df = 265; *P* < 0.05). Most of these shifts were westward (39% of species), but we also noted many eastward shifts (23%). Southward shifts (23%) were more frequent than northward shifts (15%). Westward range centroid shifts were thus 2.6 times more likely than the northward range shifts expected in response to climate change. The average south–north rate of centroid shifts was slow but significantly equatorward [−0.63 (−4.30 to 2.89) km north year^{-1}; one-sample *t* test: *t* = −4.36, df = 265, *P* < 0.001], whereas the rate of west–east shifts was 1.8 times faster and significantly westward (−1.17 [−6.95 to 4.17] km east year^{-1}; one-sample *t* test: *t* = −4.90, df = 265, *P* < 0.001). The observed rates of centroid shifts toward each cardinal direction ranged from 62% (southward) to 70% (eastward) faster than expected by chance alone, as confirmed by a null-model approach (fig. S5).

The climate significantly changed over the course of the study period in 2949 of the 2954 resurveyed vegetation plots (99.8%) based on climate analog mapping. Maximum growing-season temperatures increased by an average of 1.59 (1.15 to 2.21) °C. Climate change took place at an average absolute velocity of 0.66 (0.07 to 1.67) km year^{-1} (Fig. 2B). Unsurprisingly, the dominant geographic direction of climate change was north (40% of the plots; Rayleigh's *r* = 0.29; df = 2,948; *P* < 0.05). The south–north velocity of climate change was also significantly northward [0.24 (−0.72 to 1.94) km north year^{-1}; one-sample *t* test: *t* = 18.15, df = 2,948, *P* < 0.001]. The west–east climate-change velocity was marginal but significantly eastward [0.06 (−0.92 to 0.72) km east year^{-1}; one-sample *t* test: *t* = 5.55, df = 2,948, *P* < 0.001] and thus

Fig. 4. Distance and geographic direction of colonization and extinction centroids.
(**A**) Distance and geographic direction of colonization centroids (*n* species = 202).
(**B**) Distance and geographic direction of extinction centroids (*n* = 246). Longer distances reflect that colonization or local extinction are happening in one preferred direction relative to the baseline centroid. Rayleigh's *r* statistic represents a test of uniformity that compares the bearings of shifts to a uniform

circular distribution (null hypothesis). Larger values indicate more directed shifts. Asterisk (*) indicates significant deviations from the null hypothesis (*P* < 0.05). Results of the analyses that also included rare species are presented in fig. S9. Results for each biogeographic region are presented in fig. S10.
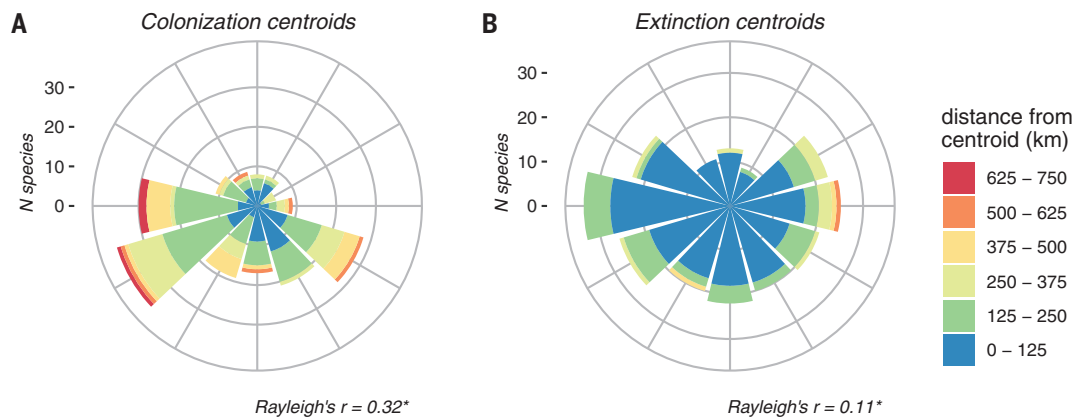
opposite to the most common cardinal direction of centroid shifts of European forest plants.

Neither the geographic direction nor the velocity of climate change was reflected in the species' centroid shifts, which hints toward the importance of other environmental changes. Two prominent alternative drivers of forest plant community changes are elevated atmospheric inputs of nitrogen and forest canopy cover changes (*16*, *25–28*), with eutrophying effects of nitrogen complicated by recovery dynamics from past acidification caused by the combined deposition of nitrogen and sulfur compounds (*29*, *30*). These drivers show spatial patterns not confounded with the velocity of climate change (across all plots, pairwise Spearman correlations between the south–north velocity of climate change versus the rates of nitrogen deposition and forest canopy cover changes were only 0.04 and 0.01, respectively).

To better understand the potential drivers associated with the reported changes in the geographical distribution of species, we related the rates of centroid shifts to the average rate of nitrogen deposition between the baseline survey and resurvey (kg N ha^{-1} year^{-1}) extracted from atmospheric deposition maps at 0.1° resolution (~8 km by 8 km within the study area) (Fig. 1A) and to the observed rate of forest canopy cover change at each site (percent canopy cover increase year^{-1}) while also accounting for the velocity of climate change (the absolute, south–north, and west–east velocity) in a linear mixed-effects modeling framework. We furthermore tested for the potential confounding effect of past acidifying deposition, considering the known adverse effects on European forest plant communities (*31*). The acidifying deposition rate, however, was highly correlated to the rate of nitrogen deposition over the course of the study because of partially shared emission sources (Spearman correlation: 0.87; *n* = 2954 plots), and their individual effects are therefore difficult to tease apart in an observational study. We calculated species-specific

experienced rates of atmospheric (nitrogen and acidifying) deposition and forest canopy cover changes as the average rate across all plots where the species was observed, weighted by the species' original abundance in the baseline time period survey (fig. S6 for a data flow chart).

Model outputs show that the absolute rate of centroid shifts was weakly but exclusively linked to the rate of forest canopy change, with greater opening of the canopy enhancing centroid shifts (Fig. 3). The velocity of climate change was not associated with the rates of centroid shifts. By contrast, the rate of nitrogen deposition was significantly linked to the west–east rate of centroid shifts, with species that initially experienced a lower nitrogen deposition rate across their distributions showing faster westward shifts (Fig. 3A). Variation partitioning revealed that the nitrogen deposition rate, rather than the climate-change velocity, explained most of the variation in the species' centroid shifts, albeit the proportion of variation explained was small (Fig. 3B and fig. S7). The estimated effects of past acidifying deposition on species centroid shifts were nearly identical to the effects of the spatially correlated nitrogen deposition rate (fig. S8). We are therefore unable to distinguish with certainty whether centroid shifts were brought about by eutrophying rather than changes in acidifying deposition or a combination of both. In either case, however, atmospheric deposition rates—and not the climate-change velocities—were the superior predictors of westward species movements.

Colonization and extinction centroids

To shed light on the mechanisms that drive centroid shifts, we decomposed centroid shifts into shifts attributed to the individual contribution of colonization and local extinction. For this analysis, we introduce the concept of colonization centroids (the centroid of plots newly colonized by a species, abundance weighted by the percentage cover in the resurvey) and extinction centroids (the centroid of plots in

which a species became extinct, abundance weighted by the percentage cover in the baseline survey). Colonization and extinction centroids were expressed as the projected distance from the species' baseline centroid position in each geographic direction (kilometers north and kilometers east). From a biogeographical point of view, longer distances reflect that colonization or local extinction events took place farther away from the baseline and that these processes occurred mostly in one preferred direction (schematically explained in fig. S1C).

The average absolute distance of colonization centroids [202.20 (28.30 to 478.46) km; *n* species = 202] was larger than the average absolute distance of extinction centroids [82.22 (4.22 to 249.48) km; *n* species = 246; Fig. 4 and figs. S9 and S10]. Colonization centroids were also more isotropic across species (Rayleigh's *r* = 0.32; df = 201; *P* < 0.05) than extinction centroids (Rayleigh's *r* = 0.11; df = 245; *P* < 0.05). This suggests that colonization events were happening more in one preferred direction and occurred further away from baseline centroids (i.e., closer to range boundaries) compared with local extinctions.

Colonization along the west–east axis was most closely related to the nitrogen deposition rate (figs. S11 and S12). Because the observed dominant direction of species' colonization was westward (39% of the species colonized west, 21% east, 28% south, and 12% north), colonization occurred more frequently for species that initially experienced a lower rate of nitrogen depositions across their distributions. This pattern is possibly associated with the westward colonization of nitrogen generalist species that can take advantage of eutrophic conditions such as observed in large parts of Western Europe (Fig. 1A and fig. S13). Indeed, linking the colonization centroids to each species' ecological indicator value for nitrogen niche width (with larger values indicating generalist species with a broader niche) (*32*) revealed that, for nitrogen generalists in

Fig. 5. Effects of nitrogen deposition rates and species nitrogen niche width on colonization centroids. (**A** and **B**) Results of the mixed-effects model testing for the interaction effect between the average nitrogen deposition rate between the baseline survey and resurvey across each species' distribution (kg N ha^{-1} year^{-1}) and species' ecological indicator values for nitrogen niche width (an index integrating the intra- and inter-regional variability in the nitrogen niche, with higher values indicating more generalist species and smaller values specialist species)

on the south–north and west–east colonization centroids. Negative distances indicate southward (A) or westward colonization (B). Interaction effects between species nitrogen niche width and the average nitrogen deposition rate between the baseline survey and resurvey across each species' distribution are plotted along the color gradient. Model predictions were plotted to a maximum of 35 kg N ha^{-1} year^{-1} to avoid extrapolation uncertainty for deposition values where observations were scarce. Modeling results without outlier data
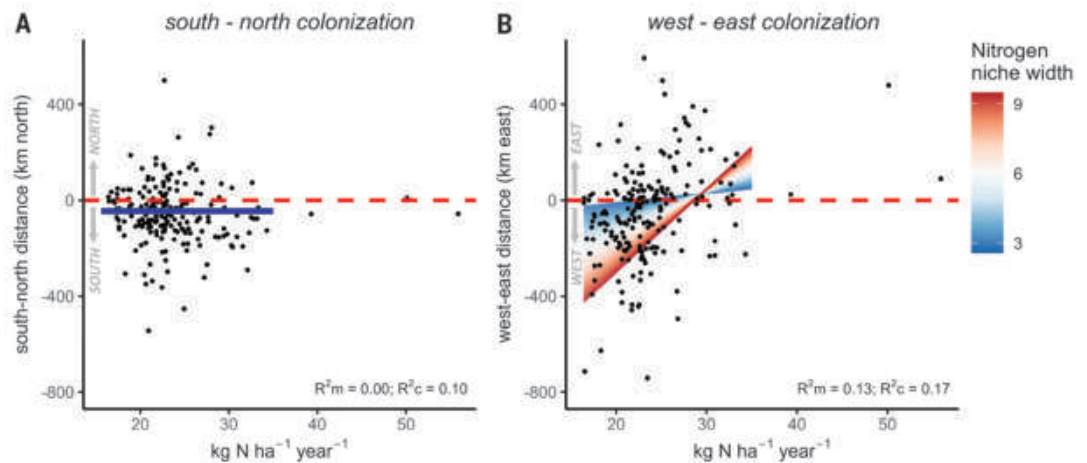
(n = 3 data points) are provided in fig. S13. The most parsimonious model of south–north colonization was an intercept-only model (blue solid line). Westward colonization depended on species nitrogen niche width. Nitrogen generalists that initially occurred in areas with lower rates of nitrogen deposition moved more westward. Colonization in the more specialist species was generally suppressed, irrespective of the nitrogen deposition rate. In all plots, the red dashed line represents the zero line.

particular, those that initially occurred in areas with a lower rate of nitrogen deposition have taken advantage to move more westward (Fig. 5 and fig. S14). Nitrogen generalists that already occurred in areas with higher nitrogen deposition (Western Europe; Fig. 1A and fig. S13) tended to remain in place without necessarily moving westward. More specialist species— i.e., those with narrow niche widths for nitrogen and that often also have smaller range sizes (*33*)—have shown lower colonization rates across temperate Europe, allowing generalist species to replace specialists (*14*). Also, the decreasing levels of acidifying deposition [since the peak in the 1980s (*34*)] may have facilitated the recovery of species' ranges in formerly polluted regions (*30*). Using our observational data, we cannot fully disentangle these recovery effects after past acidification caused by both nitrogen and sulfur pollution from dynamics of eutrophication chiefly involving nitrogen deposition. However, eutrophication may be the most likely driver because (i) we show that west–east colonization distances were statistically better linked to nitrogen deposition and the species' nitrogen niche width than to acidifying deposition and the species acidity niche width (fig. S15) and (ii) most of the vegetation plots appear to be relatively well-buffered against soil acidification (fig. S16). Regardless of whether the driver of westward colonization chiefly involves eutrophying deposition or facilitated by the recovery from past acidification, forest plant species native to regions with lower atmospheric deposition rates are more vulnerable to unanticipated range shifts in response to

atmospheric pollution—a key finding for forest biodiversity conservation policy.

Local extinction events along the south–north axis were preferentially located southward relative to the species' baseline centroid position and thus closer to species' warm range limits (25% south versus 13% north) (Fig. 4). This trend was significantly associated with climate change and a higher rate of nitrogen deposition (figs. S11 and S12). Local extinction events along the west–east axis, however, occurred more often (28% east and 31% west). Eastward local extinctions occurred more commonly in species that experienced a higher rate of nitrogen deposition across their distribution. Such nitrogen-mediated local extinctions were amplified when forest canopies became more open. The velocities of climate change also interacted with the rate of forest canopy cover change in that local extinctions due to climate change occurred more often in forests where the canopy cover decreased. This confirms the importance of tree canopies in buffering the impacts of environmental changes (*35*).

Our findings suggest that atmospheric deposition and forest canopy cover dynamics interact to determine how forest plant species are shifting their ranges and that these environmental changes induce shifts that can be independent from isotherm shifts. This contradicts the idea that species have shifted ranges mainly in response to warming air temperatures. Rather, other environmental changes, especially rates of atmospheric deposition and forest canopy cover dynamics, have likely induced unexpected westward range

shifts in European forest plants. Although it remains unclear whether the effect of atmospheric deposition chiefly involves eutrophication or a recovery effect from past acidification due to both nitrogen and sulfur pollution, our findings point at nitrogen deposition as the most likely driver explaining the westward range shifts in European forest plants. Since the continued success of the United Nations air convention and the European Union Emissions Ceiling Directive in reducing nitrogen and sulfur emission levels, prospective trends in climate change and atmospheric deposition are unlikely parallel, with climate change outpacing the effects of atmospheric deposition on future species' range shifts. Accurate and recent range shift data will be key to adequately anticipating the respective impacts of climate change and atmospheric deposition on biodiversity and ecosystem functioning. It is already clear, however, that biodiversity redistribution patterns appear complex and are more likely driven by the interplay among several environmental changes than due to the exclusive effect of climate change alone.

REFERENCES AND NOTES

1. J. Lenoir *et al.*, *Nat. Ecol. Evol.* **4**, 1044–1059 (2020).
2. I.-C. Chen, J. K. Hill, R. Ohlemüller, D. B. Roy, C. D. Thomas, *Science* **333**, 1024–1026 (2011).
3. M. T. Burrows *et al.*, *Science* **334**, 652–655 (2011).
4. J. W. Williams, S. T. Jackson, *Front. Ecol. Environ.* **5**, 475–482 (2007).
5. C. Parmesan *et al.*, *Nature* **399**, 579–583 (1999).
6. J. Lenoir, J. C. Gégout, P. A. Marquet, P. de Ruffray, H. Brisse, *Science* **320**, 1768–1771 (2008).
7. R. Virkkala, A. Lehikoinen, *Glob. Change Biol.* **20**, 2995–3003 (2014).
8. S. Dullinger *et al.*, *Nat. Clim. Chang.* **2**, 619–622 (2012).
9. C. Parmesan, G. Yohe, *Nature* **421**, 37–42 (2003).

10. M. A. Rubenstein et al., Environ. Evid. **12**, 7 (2023).
11. L. P. Shoo, S. E. Williams, J. Hero, Austral Ecol. **31**, 22–29 (2006).
12. J. D. Ash, T. J. Givnish, D. M. Waller, Glob. Change Biol. **23**, 1305–1315 (2017).
13. T. L. Greaver et al., Nat. Clim. Chang. **6**, 836–843 (2016).
14. I. R. Staude et al., Nat. Ecol. Evol. **4**, 802–808 (2020).
15. C. Senf, R. Seidl, Nat. Sustain. **4**, 63–70 (2021).
16. P. Sanczuk et al., Nat. Clim. Chang. **13**, 840–847 (2023).
17. S. M. Crimmins, S. Z. Dobrowski, J. A. Greenberg, J. T. Abatzoglou, A. R. Mynsberge, Science **331**, 324–327 (2011).
18. D. Jacobsen, Front. Ecol. Environ. **18**, 211–218 (2020).
19. J. Lenoir et al., Ecography **33**, 295–303 (2010).
20. D. M. Olson et al., Bioscience **51**, 933–938 (2001).
21. J. Vanderwal et al., Nat. Clim. Chang. **2**, 1–5 (2012).
22. M. C. Fitzpatrick, R. R. Dunn, Nat. Commun. **10**, 614 (2019).
23. J. W. Williams, S. T. Jackson, J. E. Kutzbach, Proc. Natl. Acad. Sci. U.S.A. **104**, 5738–5742 (2007).
24. S. R. Loarie et al., Nature **462**, 1052–1055 (2009).
25. P. De Frenne et al., J. Ecol. **101**, 784–795 (2013).
26. J. Segar et al., F. Nat. Commun. **13**, 7837 (2022).
27. K. Verheyen et al., J. Ecol. **100**, 352–365 (2012).
28. P.-O. Hedwall et al., Glob. Ecol. Biogeogr. **30**, 1765–1780 (2021).
29. N. Duarte, L. H. Pardo, M. J. Robin-Abbott, Water Air Soil Pollut. **224**, 1355 (2013).
30. E. Tipping, J. A. C. Davies, P. A. Henrys, S. G. Jarvis, S. M. Smart, Environ. Pollut. **281**, 117017 (2021).
31. G. Riofrío-Dillon, R. Bertrand, J.-C. Gégout, Glob. Change Biol. **18**, 3383–3394 (2012).
32. J. Dengler et al., Veg. Classif. Surv. **4**, 7–29 (2023).
33. S. Kambach et al., Ecography **42**, 467–477 (2019).
34. P. Grennfelt et al., Ambio **49**, 849–864 (2020).
35. P. De Frenne, Nat. Ecol. Evol. **8**, 196–202 (2024).
36. J. T. Abatzoglou, S. Z. Dobrowski, S. A. Parks, K. C. Hegewisch, Sci. Data **5**, 170191 (2018).
37. EMEP, Transboundary particulate matter, photo-oxidants, acidifying and eutrophying components (EMEP Status Report 1/2023, Norwegian Meteorological Institute, 2023); https://emep.int/publ/reports/2023/EMEP_Status_Report_1_2023.pdf.
38. European Environment Agency, https://www.eea.europa.eu/en/datahub/datahubitem-view/.
39. P. Sanczuk et al., Data and Code associated with original research article: Unexpected westward range shifts in European forest plants link to nitrogen deposition [Dataset], Dryad; https://doi.org/10.5061/dryad.4b8gthtmt (2024).

FERROELECTRICS

Polar and quasicrystal vortex observed in twisted-bilayer molybdenum disulfide

Chi Shing Tsang[1,2]†, Xiaodong Zheng[1,3]†, Tong Yang[1]†, Zhangyuan Yan[1,3], Wei Han[1,3], Lok Wing Wong[1,3], Haijun Liu[1,3], Shan Gao[2,4,5], Ka Ho Leung[2,5], Chun-Sing Lee[2], Shu Ping Lau[1], Ming Yang[1,6,7]*, Jiong Zhao[1,3,8]*, Thuc Hue Ly[2,4,5]*

We report the observation of an electric field in twisted-bilayer molybdenum disulfide (MoS$_2$) and elucidate its correlation with local polar domains using four-dimensional scanning transmission electron microscopy (4D-STEM) and first-principles calculations. We reveal the emergence of in-plane topological vortices within the periodic moiré patterns for both commensurate structures at small twist angles and the incommensurate quasicrystal structure that occurs at a 30° twist. The large-angle twist leads to mosaic chiral vortex patterns with tunable characteristics. A twisted quasicrystal bilayer, characterized by its 12-fold rotational symmetry, hosts complex vortex patterns and can be manipulated by picometer-scale interlayer displacement. Our findings highlight that twisting 2D bilayers is a versatile strategy for tailoring local electric polar vortices.

Bilayer two-dimensional (2D) materials such as MoS$_2$ are dominated by a few energy-favorable states. These 2D materials naturally adopt distinct stacking orders: the anti-aligned 2H (P63/mmc) state or the aligned 3R (R3m) state, which differ by a relative interlayer rotation of 60° (1). Recently, twisting of van der Waals (vdW) bilayers and related multilayered homo- or heterostructures has emerged as a transformative approach to access and tailor electronic properties through enhanced quantum coupling effects induced by long-wavelength moiré patterns. For example, moiré superlattices can host charge density waves (2), exhibit unconventional superconductivity (3), and give rise to Mott insulator states in magic-angle twisted-bilayer and trilayer graphene (4). Moreover, unexpected atomic reconstructions (5–8) and moiré excitons (9) have been observed in twisted-bilayer transition metal dichalcogenides (TMDCs).

In bilayer 2D materials, interlayer lattice structures undergo notable transformations as the twist angle is varied. At small twist angles (θ_t < 3°), the two layers maintain commensurability (10–15). However, at larger angles (θ_t > 8°), the lattice structures enter an incommensurate regime in which the superposition of electronic lattices leads to an approximate moiré supercell, albeit with a discernible loss of the long-range order translational symmetry. Previously, the incommensurate bilayer structures were thought to have a minimal impact on the interlayer electronic interaction primarily because of the lack of electronic phase coherence. However, recent studies have demonstrated nontrivial interlayer coupling in the twisted-bilayer quasicrystal, as evidenced by the discovery of Dirac electrons in dodecagonal graphene quasicrystals (16, 17) and the identification of van Hove singularities associated with the rich K valley structures in dodecagonal tungsten diselenide quasicrystals (18). These findings suggest that

[1]Department of Applied Physics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, P. R. China. [2]Department of Chemistry and Center of Super-Diamond & Advanced Films (COSDAF), City University of Hong Kong, Kowloon, Hong Kong, P. R. China. [3]The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen 518057, P. R. China. [4]Department of Chemistry and State Key Laboratory of Marine Pollution, City University of Hong Kong, Kowloon, Hong Kong, P. R. China. [5]City University of Hong Kong Shenzhen Research Institute, Shenzhen 518057, P. R. China. [6]Research Centre on Data Sciences & Artificial Intelligence, The Hong Kong Polytechnic University, Kowloon, Hong Kong, P. R. China. [7]Research Centre for Nanoscience and Nanotechnology, The Hong Kong Polytechnic University, Kowloon, Hong Kong, P. R. China. [8]The Research Institute for Advanced Manufacturing, The Hong Kong Polytechnic University, Kowloon, Hong Kong, P. R. China.
*Corresponding author. Email: kevin.m.yang@polyu.edu.hk (M.Y.); jiong.zhao@polyu.edu.hk (J.Z.); thuchly@cityu.edu.hk (T.H.L.)
†These authors contributed equally to this work.
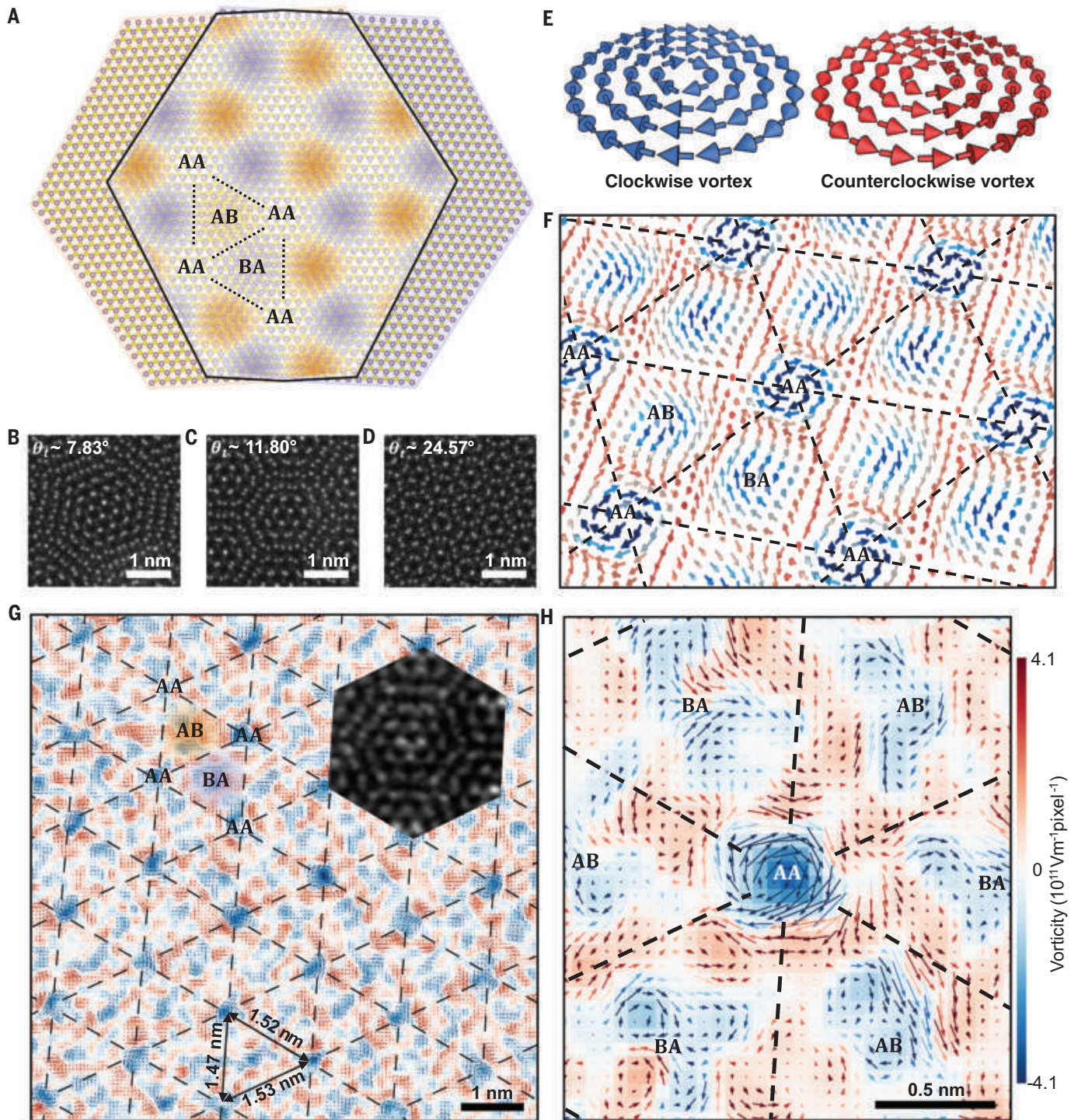
Fig. 1. 4D-STEM results of the polar field in twisted-bilayer MoS_2.
(**A**) Schematic of moiré structure emergence from the twist stacking of two individual layers of MoS_2. The AB and BA stacking area are highlighted by orange and purple, respectively. (**B** to **D**) STEM-ADF images of synthesized moiré structures of bilayer MoS_2 with twist angles (θ_t) of ~7.83° (B), 11.80° (C), and 24.57° (D). (**E**) Scheme of the polar field distribution with clockwise and counterclockwise vortex, respectively. Arrows represent the in-plane component of the polarization. (**F**) 3D map of the polar field distribution of twisted-bilayer MoS_2 at the AA, AB, and BA areas. The arrow denotes the direction of polarity.

(**G**) 4D-STEM curl (vorticity) mapping of the polar field ($-1 \times \nabla \times E_{xy}$)$_{[001]}$ with polar vectors for bilayer MoS_2 with a twist angle of 11.6°. Blue indicates clockwise rotation and red indicates counterclockwise rotation. The inset is a STEM-ADF image corresponding to atomic structure at the hexagonal lattice. The dashed line illustrates the boundaries between the AA, AB, and BA stacking areas in the twisted bilayer. (**H**) Magnified 4D-STEM polar mapping with polar vectors showing noticeable in-plane vortex and antivortex structures at the AA, AB, and BA stacking area. (G) and (H) share the same intensity bar.

Fig. 2. Twist-angle dependence of polar field distributions. (**A** to **D**) Experimental STEM-ADF images of bilayer MoS_2 with different twist angles of 8.3°, 11.6°, −18.4°, and 30°, respectively. (**E** to **H**) Corresponding curl (vorticity) mapping of polar field based on the 4D-STEM experimental results. (**I** to **L**) Multislice simulations of STEM-ADF results with the same twist angles shown in (A) to (D). (**M** to **P**) Corresponding curl (vorticity) mapping of the polar field based on the multislice simulation method. (**Q**) Twist angle–dependent polar field varies across all twist angles in a specified 0.4 × 0.4 nm region. This angular-dependent polar plot presents the variation in vortex tendency (vorticity). A total of 120 models of MoS_2, each featuring a different twist angle ranging from 0° to 119° with an incremental step of 1°, were created. The chosen 0.4 × 0.4 nm region, positioned within the AA stacking area and labeled as the grid mask of MoS_2, serves as the central point for rotation transformation in each twisted sample. The orange line on the plot corresponds to the average vorticity as a function of the twist angle, illustrating how the vorticity evolves with the twist angle. The additional background color is used to indicate vorticity directions. Blue indicates clockwise rotation and red indicates counterclockwise rotation for better recognition of peaks and trends across a wide twist angle range. This comparison allows for an evaluation of how the polar field varies with the twist angle in the specifically selected AA stacking region. The four diagrams at the bottom represent the unique vortex features at twist angles of ~8.3°, ~17.9°, ~30°, and ~50.6°, which show the vorticities of the rotation center changing from clockwise to counterclockwise. Scale bars, 1 nm.
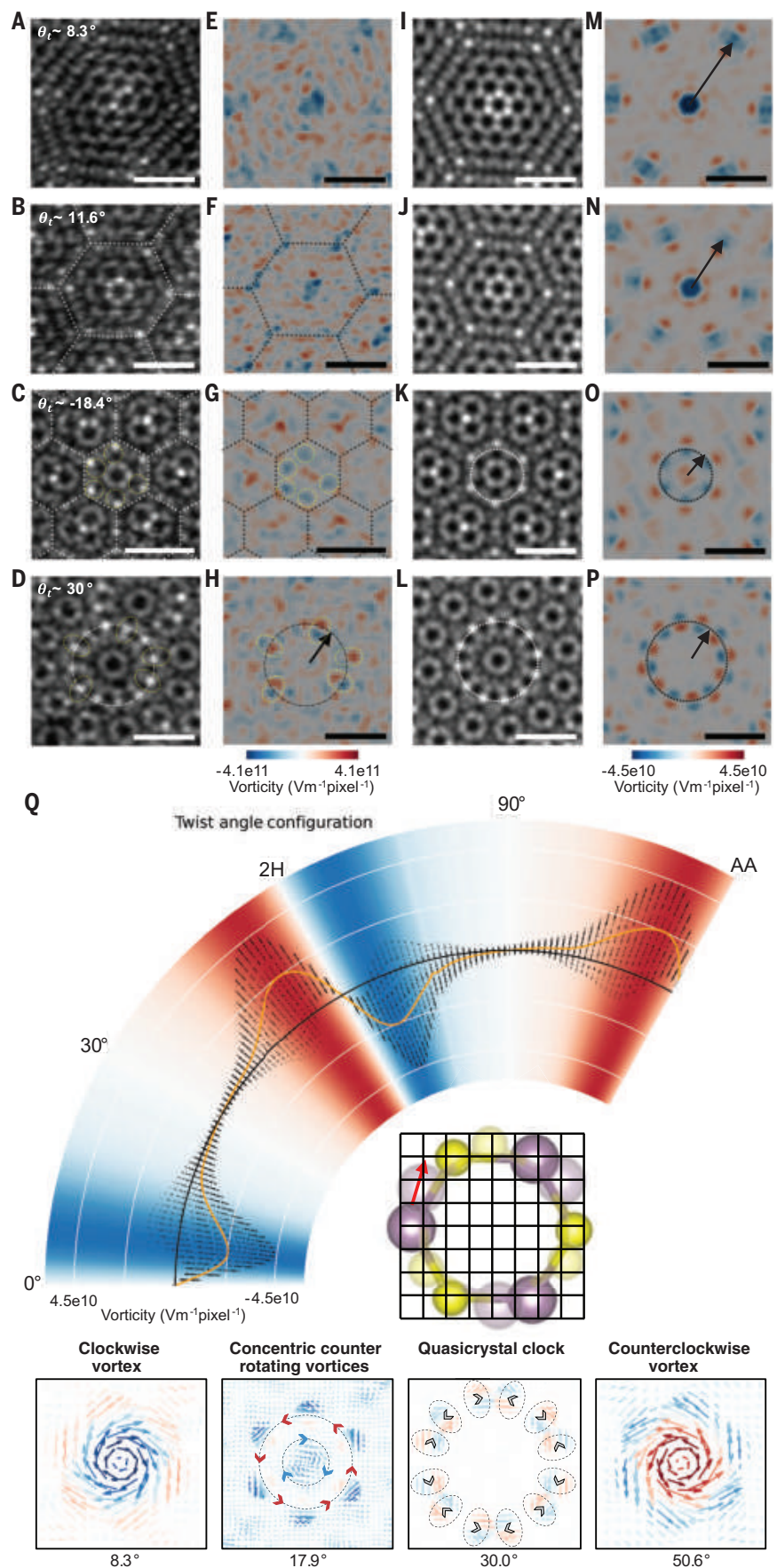
overlapping moiré potential wells may show emergent properties.

Unlike conventional ferroelectric materials, which host distinct multiple polar states at their free energy minima (19), low-dimensional 2D systems may encompass collective local electric dipole moments. These moments give rise to spontaneous polarization domains that emerge from a balance among dimensional constraints, misfit elastic energies, and interfacial charge effects (20, 21). As a result, this balance fosters the evolution of nontrivial topological configurations with continuous polar structures such as skyrmions (22, 23), merons (24), and vortices (21, 25). These local electric fields could be detected and manipulated and may have device applications. However, the observation of topological polar textures within twisted 2D vdW materials has remained elusive.

In this study, we used 4D scanning transmission electron microscopy (4D-STEM) and first-principles calculations to characterize local topological polar electric fields and establish their relation with the twist angle in twisted-bilayer MoS$_2$. In contrast to the "sliding ferroelectricity" observed in the normal direction of bilayer 2D materials such as hexagonal boron nitride (h-BN) (26, 27), TMDCs (28), and other vdW systems (22, 23), the topological vortex fields in twisted-bilayer 2D MoS$_2$ were oriented in-plane and even exhibited tunability of polar patterns over a wide range of twist angles. These intricate results were directly associated with the twist-induced moiré superstructures.

These experimental findings are further supported by our first-principles calculations in collaboration with recent insights into topological interlayer-driven ferroelectricity within moiré domains at minimal twist angles (29) and twisted BaTiO$_3$ freestanding layers (30). Moreover, the vdW interaction in atomically smooth 2D materials provides a considerable degree of turnability in terms of interlayer sliding and twisting. Beyond the complex topological mosaic vortex configurations discovered at lower twist angles, we have detected even more elaborate topological polar patterns at larger angles, including the polar quasicrystal structure with a motif of clock-like patterns and 12-fold rotational symmetry, which recurs throughout the material matrix. These polar patterns within twisted 2D bilayers could be manipulated through picometer-scale interlayer displacements, which could allow additional control over emergent properties.

Polar vortex field in twisted moiré structures

The as-synthesized twisted-bilayer MoS$_2$ samples (see the materials and methods) were investigated with a 4D-STEM experiment, in which a focused electron beam was systematically rastered across the samples under focused-probe differential phase contrast imaging conditions (31–35). These STEM and 4D-STEM techniques were used to map the in-plane atomic polarization fields in the twisted-bilayer MoS$_2$. The distribution of twist angles was analyzed through a series of selected-area electron diffraction (SAED) patterns, which can be found in fig. S1. The atomic structure model of the twisted-bilayer MoS$_2$ is shown in Fig. 1A, and the STEM–annular dark field (STEM-ADF) images are shown in Fig. 1, B to D. The observed topological polar structures, such as the vortices illustrated in Fig. 1, E and F, correspond to the 4D-STEM results shown in Fig. 1, G and H. The 4D-STEM analysis is based on the shift in the center of mass of the electron beam, which correlates to the deflected trajectory of the incident electrons after interaction with the electrostatic field in ultrathin samples. The details of the 4D-STEM method, including the computation of the in-plane vorticity of the measured polarization fields ($-1 \times \nabla \times E_{xy}$), can be found in the materials and methods. The intensity of vorticity is proportional to the curl of the polarization field, which signifies the swirling nature of the local field. In this context, the blue and red hues in Fig. 1, G and H, denote the clockwise and counterclockwise swirling directions, respectively.

The STEM-ADF (fig. S2) image revealed hexagonal arrangements with the moiré periodicity (λ_m ~1.5 nm; fig. S3). In addition, the additional spots that appeared in the fast Fourier transform analysis due to the coherence effect of the STEM-ADF images from two samples (θ_t ~8.3° and θ_t ~11.6° in fig. S4) also confirmed the existence of the moiré pattern and its periodicity. Although slight variations in the equilateral periodicities observed were caused by strain influences or experimental drift, they did not affect the measured features. The characteristic periodic vortex patterns aligned well with the moiré lattice, as shown by our 4D-STEM observations (Fig. 1, G and H).

The most pronounced vortex feature, associated with local AA stacking regions that exhibit mirror symmetry with respect to the middle plane of bilayer, manifested a lateral periodicity of ~1.5 nm in the 11.6° twisted-bilayer sample (Fig. 1G). This vortex feature was also found at the AB/BA stacking domains, showing a hexagonal sublattice pattern with six minor vortices (Fig. 1H). These features were in concert with those of the switchable out-of-plane moiré polarization domains: sliding ferroelectricity in a small twisted 2D structure. The vortex polarization fields were also observed in bilayer MoS$_2$ samples at twist angles of 1.8°, 2.4°, –4.1°, and 8.3° (figs. S5 to S7), where the positive or negative twist angle denotes a clockwise or counterclockwise rotation, respectively, of the top layer with respect to the bottom layer. This further demonstrates the presence of the in-plane vortex polarization patterns in twisted-bilayer MoS$_2$ across a wide twist angle range.

The out-of-plane polarization within the moiré supercells, which are a consequence of the interlayer twisting, could be quantitatively assessed through density functional theory (DFT) calculations. These calculations differentiated potential variations in the out-of-plane direction, which indicates the emergence of downward (–P) and upward (+P) polarization states within the AB and BA stacking domains, respectively. By contrast, AA stacking domains did not show out-of-plane polarization (figs. S8 and S9). By combining these findings, the overall scheme of the polar field distribution can be illustrated (Fig. 1F), where the AB and BA sites exhibited similar in-plane vortex field and opposite out-of-plane polarity, resulting in different chirality.

To elucidate the nature of such nontrivial topological polar textures, it was essential to evaluate their topological charges (36, 37), which can be computed by integrating the angular variation of the normalized local field over a 2D cross-section. The winding number, an integral topological invariant, is defined as:

$$N_{sk} = \frac{1}{4\pi} \int n \cdot \frac{\partial n}{\partial x} \times \frac{\partial n}{\partial y} \, dx \, dy \ (20, 38),$$

where N_{sk} refers to the number of times that the plane surface can wrap around a unit sphere and n is the normalized local dipole moment. The experimental results of the in-plane polar field in twisted-bilayer MoS$_2$ (Fig. 1, G and H) show 2π modulation of the in-plane component. However, the absence of a z-component measurement in the experiment hinders a direct analysis for these meron-like topologies that are commonly seen in ferroelectric materials and are characterized by their discrete swirling core and spherical domain boundaries (24).

The localized polar vortex field that we observed could be ascribed to the moiré periodicity that directly resulted from the twisting of the MoS$_2$ bilayers. The formation of in-plane vortex polar domains in these twisted bilayers was likely to stem from the twist stacking–induced charge redistribution in combination with a weak contribution from small in-plane ionic displacements. This is in stark contrast to the ferroelectricity observed in traditional perovskite oxides, which arises primarily from ionic displacements (23, 24), or in out-of-plane ferroelectricity in twisted-bilayer h-BN and TMDs (5, 26–28, 39), which is induced by interfacial charge transfer only. Specifically, the distinctive in-plane polarization in highly twisted bilayers originates from the local approximant stacking domains. DFT calculations imply small in-plane atomic displacements of within 0.055 Å at the twist angle of 6.0090° upon stacking and tend to further decrease and even diminish with the increasing twist angle (fig. S10). This is indicative of a weak contribution to the in-plane polarization from the in-plane ionic displacement, especially in the large twist angle regime. Conversely, there are noticeable electronic charge redistributions induced by the twist stacking, which are mainly localized around the interfacial S and Mo atoms,
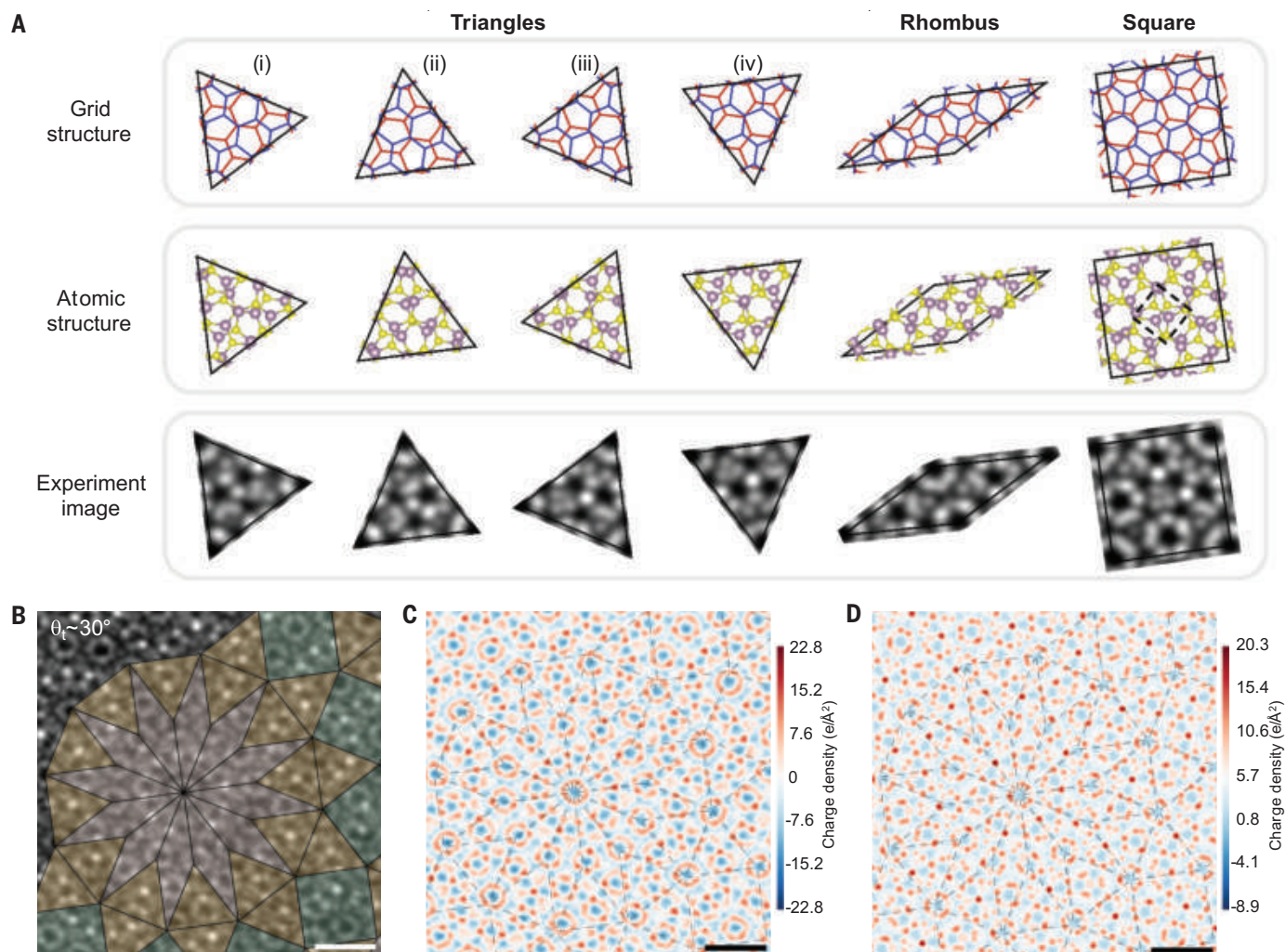
Fig. 3. Basic tiles and charge density distribution of 30° twisted-bilayer quasicrystal MoS₂. (**A**) Grid structure, atomic structure, and STEM-ADF image of Stampfli tiles in 30° twisted-bilayer quasicrystal MoS₂. (**B**) Experimental STEM-ADF image of 30° twisted-bilayer quasicrystal MoS₂ mapped with false-colored 12-fold Stampfli tiles. (**C** and **D**) Experimental (C) and DFT-simulated (D) charge density contour plot in 30° twisted-bilayer quasicrystal MoS₂. The superimposed gridlines refer to the perfect Stampfli tiles. Scale bars, 1 nm.

as displayed in fig. S11. Further details on the in-plane polarization estimation are provided in the supplementary text and in table S1.

Twist angle–dependent polar structures

To explore the impact of twist angles on the topological polarization structures, we undertook a comprehensive spatial evolution analysis. We scrutinized the polar domains in MoS₂ bilayer moiré patterns across a wide range of twist angles using both 4D-STEM measurement and corroborative image simulations, as illustrated in Fig. 2, A to P. We observed that MoS₂ samples subjected to diverse twist angles, namely 8.3°, 11.6°, –18.4°, and 30°, demonstrated different polar field distributions that ranged from intricate vortex formations to 12-fold quasicrystal polar domains resembling clock-like structures.

At lower twist angles ($\theta_t < 3°$), the interlayer twist modulated the electronic properties

through induced moiré superlattice reconstruction (8, 40). At higher twist angles ($\theta_t > 8°$), the bilayers showed complex atomic configurations (Fig. 2C). These structures gave rise to diverse polar textures (Fig. 2G), which were different from those observed at lower twist angles (Fig. 2, E and F). We further discovered both counter-rotational (blue) and anti–counter-rotational (red) vortices around the AA sites (rotation center) (Fig. 2G, yellow circles). The combination of these rotational vortices led to a pattern of concentric counter-rotating vortices (Fig. 2G), which was consistent with the results obtained from multislice image simulations (Fig. 2O).

In the 30° twisted-bilayer MoS₂ quasicrystal, we observed a "clock" motif manifested in both the STEM-ADF and the SAED patterns (Fig. 2D and fig. S12, A and B). Each clock motif could be identified by a dotted circle ring that enclosed an arrangement of 12 lobes or atomic clusters

centering around AA stacking sites exhibiting a 30/90° twist. These clocklike motifs distributed in an aperiodic manner, filling the STEM-ADF image of the 30° twisted-bilayer MoS₂ and leaving space for motifs of the 30° twisted AA stacking "hole."

The polar mapping of the 30° twisted-bilayer MoS₂ (Fig. 2P) unveiled a distinct circular ring formation in the shape of a clock, each featuring 12 lobes that resembled the patterns observed in the STEM-ADF images. Upon closer examination of the atomic structure, it was evident that the atomic columns were closely stacked, which resulted in pairs of blue-red and red-blue lobes relative to the center. These atomic columns exhibited a deviation of ~15° from the zigzag and armchair orientations of MoS₂ (fig. S13). Furthermore, the atoms situated at the top and bottom of the nearly aligned atomic column were laterally separated by ~25 pm. When the clock motif was oriented at
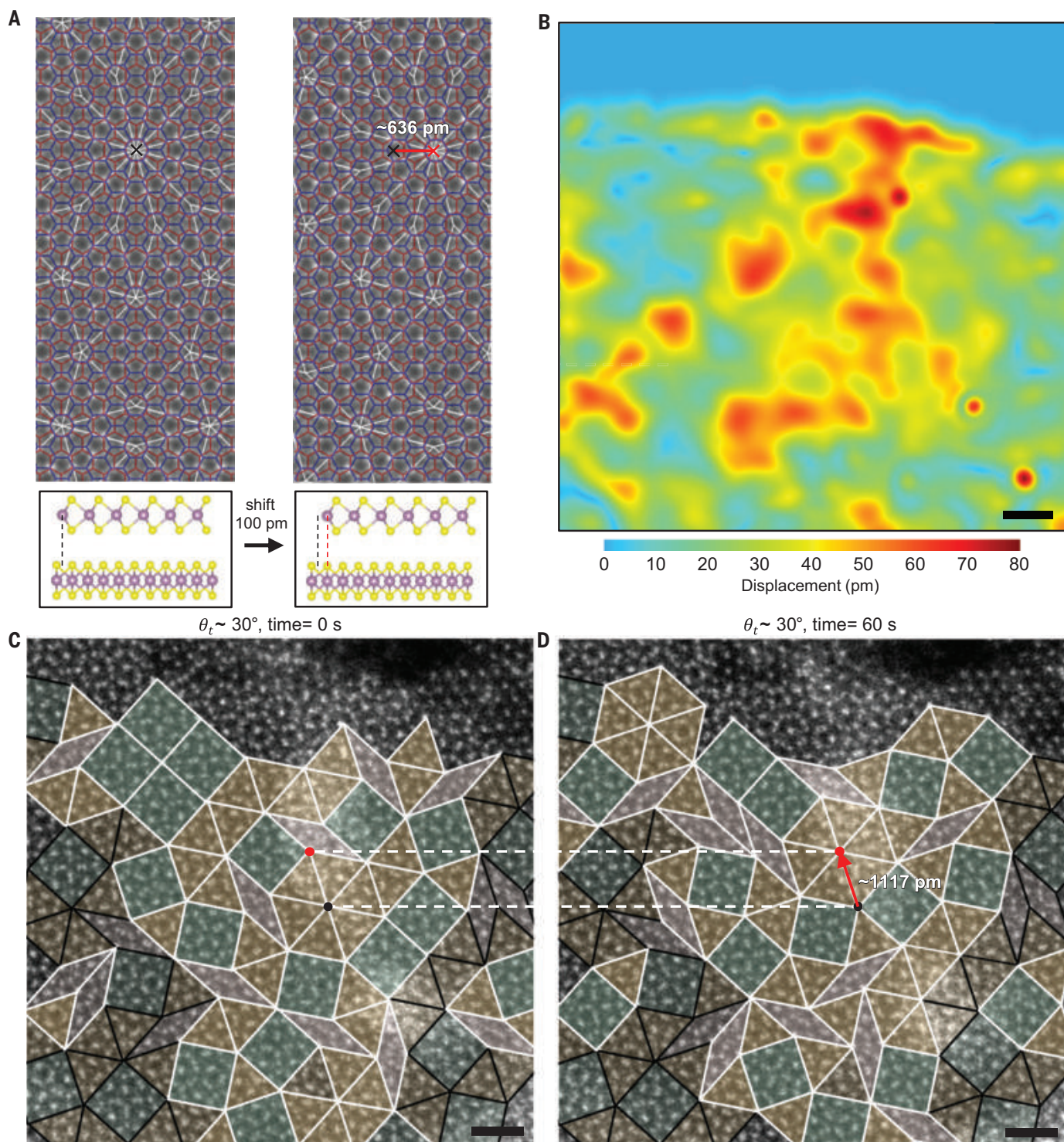
Fig. 4. In situ TEM manipulation of Stampfli tiles in 30° twisted-bilayer quasicrystal MoS$_2$. (**A**) Schematic for the alteration of Stampfli tiles before and after a 100-pm interlayer shift in 30° twisted-bilayer quasicrystal MoS$_2$. A lateral displacement of 100 pm in the top layer relative to the bottom layer results in a notable alteration of the Stampfli tiles as the pattern shifts toward the right by 636 pm. (**B**) Experimental interlayer displacement mapping for bilayer MoS$_2$ upon in situ STEM manipulation, corresponding to (C) and (D). The color bar represents the magnitude of local relative displacements, as determined by the differences between Mo atomic positions before and after manipulation. (**C**) STEM-ADF snapshot of Stampfli tiles in the original 30° twisted-bilayer quasicrystal MoS$_2$. (**D**) STEM-ADF snapshot of Stampfli tiles at the same area as shown in (C) after in situ STEM manipulation. The Stampfli tiles are typed as rhombus (light pink), square (green), and triangle (yellow). The gridlines illustrate patterns with (white) and without (black) severe changes after manipulation. The red arrow in (D) highlights the upward shift of a hexagon by 1117 pm after manipulation. Scale bars in (B) to (D), 1 nm.

~30°, a slight distortion in the atomic structure influenced the polar map, leading to a weakened representation of the symmetric blue-red and red-blue lobes. The curl map for the quasicrystal displayed weak swirling features, as shown in fig. S14. As a result, only a few of these lobes were clearly distinguishable in the experiments (Fig. 2H, yellow dotted lines).

We present the polar field distribution across a wide range of twist angles, ranging from 0° to 120° (Fig. 2Q). During the twisting process, a selected area of the AA core region was used as the rotation center. The intensities of the polar fields at these AA cores were notably prominent at twist angles in the vicinity of ~8° and 52°, as well as 68° and 112°, which corresponded to vortex and antivortex configurations, respectively. In the case of the commensurate structures observed at 0° and 60° twist angles, a marked mirror symmetry was noticeable, manifesting in the vorticity inversion. At certain twist angles, such as 30° and 90°, the emergence of a horizontal plateau was observed, indicating quasicrystal structure formation. These plateaus occurred at high twist angles, underscoring a transition to a more complex structural paradigm. At higher twist angles ($\theta_t > 8°$), a mosaic periodicity emerges, which is often classified as an incommensurate system. This mosaic periodicity involved a periodic variation in polar order spanning 0° to 60°, and it became increasingly pronounced within the intermediate twist angles (15° to 22°), as shown in fig. S15.

The presence of this mosaic periodicity in the high–twist angle regime raised fundamental questions regarding the stacking configuration and moiré wavelength. A closer examination of the STEM-ADF image (fig. S16) revealed saddle domain walls in which AB/BA domains became obscure and gave way to a dominance of the AA stacking configuration. The difference between AA and AB/BA lies in the interlayer slip or shift along the $(11\bar{2}0)$ crystal plane with a magnitude of $\frac{\sqrt{3}}{3}a$, where a is the lattice constant. In the intermediate–twist angle regime, the overall structure could be described as an approximated stacking configuration in which AA regions were separated by saddle domain walls rather than the distinct AA, AB, and BA configurations observed in the low–twist angle region. As the twist angle deviated from 30°, we observed the emergence of a distinct circular pattern, with the shortest periodicity occurring between 23° and 37°. This pattern exhibited characteristics similar to the clock motif observed in the 30° quasicrystal bilayer. The overall periodicity in the twisting tendency transited from atomic and mosaic structures to a more aperiodic or microscale arrangement, rather than strictly following the rules of the moiré equation or wavelength.

To complement our experimental findings, multislice image simulations were used to model the scattering dynamics of incident electron probes interacting with the twisted-bilayer samples (41). The simulation results aligned very well with the experimental observations (Fig. 2, A to P) and also suggested that twist polarization phenomena could occur in other materials. Vortices are observed in various types of rigidly twisted stacking 2D materials, including MoS_2, indium selenide (InSe), and bismuth selenide (Bi_2Se_3), although they are nearly absent in graphene (fig. S17).

Further explorations into the effects of lateral sizes on MoS_2 revealed intriguing insights. The polarization signals between a single hexagonal MoS_2 unit cell and an extended 5-nm unit cell showed negligible variance (fig. S18). This observation indicates that the detected polarization signals are caused by the twisted lattice and electronic structures rather than being effects of moiré pseudoparticles. To verify our results and to mitigate any potential artifacts introduced by the 4D-STEM setup, we conducted a series of controlled experiments with varied collection angles and beam convergence angles to assess their influence on the observed polar field results (figs. S19 and S20) and ferroelectric structures (fig. S21). These comprehensive investigations ensured the robustness and accuracy of our findings.

Quasicrystal tiling and 12-fold symmetry

The 30° rotation between the two MoS_2 layers led to the formation of a dodecagonal quasicrystal moiré superlattice, an incommensurate structure that was characterized by its 12-fold rotational symmetry but also by its lack of translational symmetry (16). Our observations revealed a distinct signal displaying 12-fold symmetry with two side lobes, similar to the optical phase singularities observed in momentum space (42). Using STEM-ADF images obtained with a virtual detector constructed through 4D-STEM, we demonstrated the quasiperiodic ordering of the atomic structure (fig. S12A). The STEM-ADF image overlapped the dodecagonal framework (Fig. 3, A and B), in which the vertices consisted of two-hole-aligned columns. This STEM result (Fig. 3B and fig. S12A) is in contrast to the phase-contrast–based high-resolution transmission electron microscopy images of quasicrystal bilayer graphene (16). The dodecagonal arrays in this quasicrystal bilayer MoS_2 are composed of Stampfli tiles such as squares, rhombuses, and triangles, which fill the space without any voids (Fig. 3, A and B) (43). Furthermore, the triangular tiling pattern can be classified as four distinct sublattices based on the anisotropic atomic configurations of Mo and S_2.

Although both MoS_2 and graphene have hexagonal lattices, the lattice sites occupied by Mo and S in MoS_2 are not equivalent to those of the carbon atoms in graphene. MoS_2 belongs to the space group P63/mmc, whereas graphene belongs to the space group P6/mmm. In twisted-bilayer quasicrystal MoS_2, four distinct types of triangular tiles emerged, each resulting from different stacking combinations of sublattice atoms within the triangular domains. These combinations can be characterized by the overlapping of atom columns at the center, as depicted in Fig. 3A. We identified these combinations as Mo/S_2, S_2/S_2, S_2/Mo, and Mo/Mo stackings, which are denoted as (i), (ii), (iii), and (iv), respectively (Fig. 3A). Such varied stackings are absent in graphene quasicrystals because the two sublattices are symmetry equivalent.

We further analyzed the divergence in the 4D-STEM electric field measurements in bright-field mode (detailed in the materials and methods), which allowed us to map the charge density distribution in the bilayer quasicrystal MoS_2. The mapping results are in good agreement with the DFT simulations (Fig. 3, C and D). It should be noted that the electronic coupling effect was too weak to result in substantial amplitude fluctuations of electron wave functions detectable using the STEM technique. However, it could be indirectly inferred through techniques such as angle-resolved photoemission spectroscopy and scanning tunneling spectroscopy, which can capture photon-electron interactions (16) and weak tunneling current (17), respectively, and might offer the necessary sensitivity to observe this effect.

Within the triangular subtiles of the quasicrystal structure, the polar fields not only exhibit a noticeable degree of similarity, but also a centrosymmetry relative to the center of the tile, as evidenced in Fig. 3A and fig. S22. Even within the same type of triangular tiles, we observed "two side lobes" with different orientations near the tile center. This observation suggested a deviation from the expected threefold rotational symmetry within the polar field of the triangular tiles, which could be caused by small lateral displacements in the superimposed atomic columns.

The corresponding grid, atomic structures, and experimental images (Fig. 3A) provide a detailed depiction of each Stampfli tile. In the rhombus tiles, despite the variations in the two subtiles created by different stacking sequences, they recurred in an alternating pattern around a 12-fold center. The STEM-ADF images and the associated polar fields of these rhombic subtiles were identical (Fig. 3B). Ignoring the stacking order, the structure appeared the same when viewed in projection. For the square tiles, the columns of atoms overlapping near the centers consisted of Mo/S_2, S_2/S_2, S_2/Mo, and Mo/Mo. These configurations gave rise to a self-affine square tiling pattern, which follows the deflation and inflation rule with a scaling factor of $\sqrt{3+\sqrt{2}}$ (Fig. 3A, dotted square) (41).

We emphasize that vorticity was only observed in areas with small rotational displacements where atomic columns were nearly

aligned. It disappeared in regions characterized by well-arranged BA/AB and AA stackings. Furthermore, the detailed STEM images did not align perfectly with the Stampfli mapping, possibly because of drifting and distortion during the 4D-STEM data acquisition process. To maintain the integrity of the original data, we did not apply affine transformations to correct these imperfections during data processing. Instead, we chose to showcase the quasiperiodic order through the SAED results (fig. S12B). There were two side lobes forming a clock pattern in the polar field, which correspond to a mesoscale local stacking periodicity of ~2 nm perpendicular to the rhombic tile boundary. This configuration aligned well with the joint line connecting the vorticity inversion lobes (fig. S13).

In situ TEM manipulation of bilayer quasicrystal MoS$_2$

We performed in situ TEM experiments to further manipulate the Stampfli tile patterns of bilayer quasicrystal MoS$_2$. The pristine 30° twisted-bilayer MoS$_2$ grown using the CVD method was free of strain, and we initially introduced a crack into the twist sample. The crack mediated by the electron beam propagated away from the beam-exposed area, causing substantial shear and strain in the vicinity of the bilayer MoS$_2$ crack. No beam-induced defects were observed in these areas. The different crystal orientations of the individual layers caused the crack to propagate along various directions, which resulted in the bifurcation of the crack tip (fig. S23A). This approach allowed us to observe and document the relative tiny interlayer displacement at the atomic scale under interlayer shear strain in the bilayer quasicrystal MoS$_2$. Such interlayer motion directly affected the local tessellation pattern of the quasicrystal, as depicted in Fig. 4A, suggesting that these clock polar patterns were closely associated with the vertices of the tiles.

To visualize and analyze the atomic displacements, we plotted false-colored Mo atomic displacements derived from sequential STEM-ADF images using the atom-tracking method (fig. S23). Arrows are used to signify the direction of the displacements, tracking the transition from the primary (initial) to the secondary (subsequent) image frames (fig. S23C). By subtracting displacements between two successive STEM-ADF image frames (fig. S23, D and F), specifically between the individual layers, we mapped the local interlayer shift of the bilayer quasicrystal (Fig. 4B). As a result, this specific sample area exhibited a maximum lateral interlayer displacement of 80 pm.

These results demonstrated a correlation between the quasicrystal tiling configurations and the relative interlayer displacements. Notable modifications in the quasicrystal tessellation were readily identified in highly distorted regions, which are highlighted by white grid overlays in the STEM-ADF images (Fig. 4, C and D). These modifications were in good agreement with the patterns of atomic displacements shown in Fig. 4B. In essence, our observations substantiated the modulation of both the tiling and polar domain structures through subtle interlayer displacement in bilayer MoS$_2$. The examination of node components (fig. S24) indicated a consistent distribution of node elements among various groups, despite tiling patterns undergoing considerable lateral shift (1117 pm; Fig. 4D, red arrow) after the manipulation. Therefore, in the bilayer quasicrystal MoS$_2$, the tiling pattern could amplify the interlayer displacement by >10 times and provide a versatile approach to manipulating the atomic and polar structures in these twisted bilayers. It should be noted that the amplification factor is anisotropic to the atomic movement, which results in varied magnitudes and differs from the predicted moiré period (fig. S25).

Conclusions

Our results successfully demonstrate the emergence of an in-plane polar vortex in twisted-bilayer MoS$_2$, highlighting the universal nature of topological chiral polar structures in 2D moiré systems. By combining experimental 4D-STEM measurements with theoretical simulations, we not only clarified the intricate angle-dependent polar structures in twisted-bilayer MoS$_2$, but also uncovered the potential to manipulate the vortex polar domains through interlayer displacement. The presence of correlated in-plane polar vortices and local atomic stacking orders provides opportunities for manipulating these chiral polar vortices through external electric field or interlayer sliding and twisting. These discoveries provide useful insights into the complex behavior of polar structures in twisted 2D bilayers, paving the way toward tuning emergent quantum properties at atomic scale and thus enabling promising prospects for high-density information storage and processing.

REFERENCES AND NOTES

1. S. Carr et al., Phys. Rev. B **98**, 224102 (2018).
2. Z. Song, X. Sun, L. Wang, J. Phys. Chem. Lett. **11**, 9224–9229 (2020).
3. Y. Cao et al., Nature **556**, 43–50 (2018).
4. A. Uri et al., Nature **620**, 762–767 (2023).
5. X. Wang et al., Nat. Nanotechnol. **17**, 367–371 (2022).
6. C. R. Woods et al., Nat. Phys. **10**, 451–456 (2014).
7. A. Weston et al., Nat. Nanotechnol. **15**, 592–597 (2020).
8. H. Yoo et al., Nat. Mater. **18**, 448–453 (2019).
9. S. Susarla et al., Science **378**, 1235–1239 (2022).
10. Y. Liu, Y. Huang, X. Duan, Nature **567**, 323–333 (2019).
11. K. S. Novoselov, A. Mishchenko, A. Carvalho, A. H. Castro Neto, Science **353**, aac9439 (2016).
12. K. Kang et al., Nature **550**, 229–233 (2017).
13. L. Du et al., Nat. Rev. Phys. **3**, 193–206 (2021).
14. G. Li et al., Nat. Phys. **6**, 109–113 (2010).
15. K. Ko et al., Nat. Mater. **22**, 992–998 (2023).
16. S. J. Ahn et al., Science **361**, 782–786 (2018).
17. W. Yao et al., Proc. Natl. Acad. Sci. U.S.A. **115**, 6928–6933 (2018).
18. Y. Li et al., Nature **625**, 494–499 (2024).
19. J. Íñiguez, P. Zubko, I. Luk'yanchuk, A. Cano, Nat. Rev. Mater. **4**, 243–256 (2019).
20. V. Govinden et al., Nat. Mater. **22**, 553–561 (2023).
21. A. K. Yadav et al., Nature **530**, 198–201 (2016).
22. Z. Hong, L.-Q. Chen, Acta Mater. **152**, 155–161 (2018).
23. S. Das et al., Nature **568**, 368–372 (2019).
24. Y. J. Wang et al., Nat. Mater. **19**, 881–886 (2020).
25. Y. Sun et al., Sci. Adv. **5**, eaav4355 (2019).
26. L. Li, M. Wu, ACS Nano **11**, 6382–6388 (2017).
27. K. Yasuda, X. Wang, K. Watanabe, T. Taniguchi, P. Jarillo-Herrero, Science **372**, 1458–1462 (2021).
28. J. Shi et al., Adv. Mater. **29**, 1701486 (2017).
29. D. Bennett, G. Chaudhary, R.-J. Slager, E. Bousquet, P. Ghosez, Nat. Commun. **14**, 1629 (2023).
30. G. Sánchez-Santolino et al., Nature **626**, 529–534 (2024).
31. G. Varnavides et al., Iterative phase retrieval algorithms for scanning transmission electron microscopy. arXiv:2309.05250 [cond-mat.mtrl-sci] (2023).
32. M. W. Tate et al., Microsc. Microanal. **22**, 237–249 (2016).
33. I. Lazić, E. G. T. Bosch, S. Lazar, Ultramicroscopy **160**, 265–280 (2016).
34. K. Müller-Caspary et al., Phys. Rev. B **98**, 121408 (2018).
35. W. Gao et al., Nature **575**, 480–484 (2019).
36. Y. Nahas et al., Nat. Commun. **6**, 8542 (2015).
37. N. D. Mermin, Rev. Mod. Phys. **51**, 591–648 (1979).
38. Y.-T. Shao et al., Nat. Commun. **14**, 1355 (2023).
39. M. Vizner Stern et al., Science **372**, 1462–1466 (2021).
40. S. H. Sung et al., Nat. Commun. **13**, 7826 (2022).
41. J. Barthel, Ultramicroscopy **193**, 1–11 (2018).
42. J. Ni et al., Science **374**, eabj0039 (2021).
43. P. Stampfli, Helv. Phys. Acta **59**, 1260 (1986).
44. C. S. Tsang et al., Data for: Polar and quasicrystal vortex observed in twisted bilayer molybdenum disulfide, Zenodo (2024); https://doi.org/10.5281/zenodo.13340734.

SEED DISPERSAL

Evidence of a European seed dispersal crisis

Sara Beatriz Mendes[1]*, Jens Mogens Olesen[2], Jane Memmott[3], José Miguel Costa[1], Sérgio Timóteo[1], Ana Laura Dengucho[1], Leonardo Craveiro[1], Ruben Heleno[1]*

Seed dispersal is crucial for ecosystem persistence, especially in fragmented landscapes, such as those common in Europe. Ongoing defaunation might compromise effective seed dispersal, but the conservation status of pairwise interactions remains unknown. With a literature review, we reconstructed the first European-wide seed dispersal network and evaluated the conservation status of interactions by assessing each interacting partner's IUCN (International Union for Conservation of Nature) conservation status and population trends. We found that a third of the disperser species and interactions face potential extinction and that 30% of the plant species have most of their dispersers threatened or declining. Our study reveals a developing seed dispersal crisis in Europe and highlights large knowledge gaps regarding the dispersers and conservation status of zoochorous plants, urging further scrutiny and action to conserve the seed dispersal service.

The movement of genetic diversity across space is a fundamental cornerstone of long-term ecosystem functioning and resilience (1). For most plants, the dispersal of seeds by either biotic (i.e., animals) or abiotic (e.g., wind) vectors is crucial because it is the only stage of their life cycle when individuals can move (2). Seed dispersal is an important ecological function and a key regulating service (3), allowing plant species to alter their distributional range, avoid high competition and natural enemies near the mother plant, recover from perturbations, and maintain gene flow between fragmented populations (4–6). Animal dispersers are thus important ecosystem service providers (7), particularly for plants with traits that promote biotic seed dispersal (8).

The European landscape has undergone severe transformations in recent millennia, becoming the most fragmented continent on Earth (9). Habitat fragmentation due to transportation networks, urbanization, and agriculture poses multiple physical barriers to plant dispersal. These constraints affect gene flow and population dynamics (10), being particularly worrying given plants' need to track favorable climatic envelopes because of climate change (11). Furthermore, the increasing frequency and intensification of wildfires (12) magnifies the importance of seed dispersal for the recolonization of burnt areas (13).

In this context, the decline and local extinction of animals—many of which disperse viable seeds—can have profound implications for European plant recruitment (14, 15). It is estimated that 2248 native European plant species (~23% of the European flora) have

seeds with specific adaptations for biotic dispersal, growing to 66% for woody species (8). Accordingly, the decline of seed dispersers could transform European landscapes into "empty forests" (16), creating a plant extinction debt (17).

Surprisingly, despite considerable efforts to conserve European species (18, 19), the conservation of the biotic interactions that hold those species together has received comparatively little attention (20). This is unfortunate, as the extinction of interactions often anticipates that of species (21). By simultaneously considering the species and the interactions that support them, ecological network analysis provides a critical framework for evaluating how the loss of species and interactions affect community functioning and resilience (22). However, and despite the urge to conserve biotic interactions and the ecological functions they provide, we still lack a comprehensive framework to assess the conservation status of biotic interactions (23).

Defaunation has raised concerns over animals' capacity to effectively disperse seeds, but generalizations are hindered by a predominance of single-guild studies (24, 25). Consequently, the general conservation status of seed dispersal services remains unevaluated [but see (11)]. We address this knowledge gap by assembling a European-wide seed dispersal network. We conducted a systematic literature review in Google Scholar and Web of Science to extract all records of frugivory and seed dispersal interactions by all vertebrates and invertebrates on the European continent. To minimize geographical and taxonomical biases, we performed direct searches for each animal species in English, Portuguese, Spanish, French, German, and Russian. This enabled us to (i) evaluate the conservation status of seed dispersal services in Europe, (ii) identify specific plants and regions particularly vulnerable to dispersal failure, and (iii) develop a general framework to assess the conservation status of biotic interactions.

European frugivory and seed dispersal network

Our literature searches retrieved more than 50,000 publications, including scientific papers, books, grey literature, curated datasets, reports, and personal communications. Information was compiled from 1843 published references (1660–2023). Most (65%) of these publications were in English, with the remaining being written in 25 other languages, namely Russian (10%), Spanish (7%), German (7%), French (4%), Hungarian (2%), and 20 other European languages (5%). Interaction data originated from 38 European countries, including the nine European biomes (Fig. 1). The European frugivory and seed dispersal network had 15,229 distinct plant-animal interactions among 2154 plant species (138 families) and 516 animal species (103 families) (Fig. 2A). Naturally, owing to the broad spatial scale of the study, not all species co-occur.

Eighteen percent of all interactions correspond to records of confirmed seed dispersal (with viable seeds), 57% to records of potential seed dispersal (entire seeds dispersed but viability untested), and 25% to frugivory records without information about seed fate. Frugivory was discarded from further analysis, while the former two were combined and considered hereafter as seed dispersal interactions. Seed predation interactions were not included in this study. Thus, the European seed dispersal network (Fig. 2B) included 11,414 interactions between 1902 plant species and 455 disperser species. The latter included 283 bird species, 85 arthropods, 69 mammals, 11 reptiles, four mollusks, two fish, and one annelid. Most plant and disperser species were native to Europe; however, ~11% of the plants (n = 203) and 8% of the dispersers (n = 35) were introduced, and 0.7% (n = 3) had uncertain origin (Fig. 2B). Introduced plants were excluded from further analyses.

Only 35% of the 1699 native plants included in the network had specific traits to promote biotic seed dispersal. Thirty percent had adaptations to promote abiotic dispersal, and 42% had unspecialized seeds (Fig. 2C). It is widely recognized that even plants without adaptations for biotic dispersal benefit from animal dispersal and that losing dispersers can reduce their dispersal distances (26, 27). However, because these plants do not strictly depend on animals for dispersal, hereafter we focused exclusively on native plants with adaptations for animal dispersal (n = 592). These were dispersed by 398 disperser species through 5030 interactions (Fig. 2D).

Each animal species dispersed on average 13 plant species (median = 4; range 1 to 119). The number of plant species dispersed by each disperser (animal degree) and the importance of each animal species for the plant community (measured as the sum of the dependencies of

[1]Centre for Functional Ecology, Associate Laboratory TERRA, Department of Life Sciences, University of Coimbra, 3000-456 Coimbra, Portugal. [2]Department of Biology, Aarhus University, 8000 Aarhus C, Denmark. [3]School of Biological Sciences, University of Bristol, Bristol, UK.
*Corresponding author. Email: sarabmendes@gmail.com (S.B.M.); rheleno@uc.pt (R.H.)
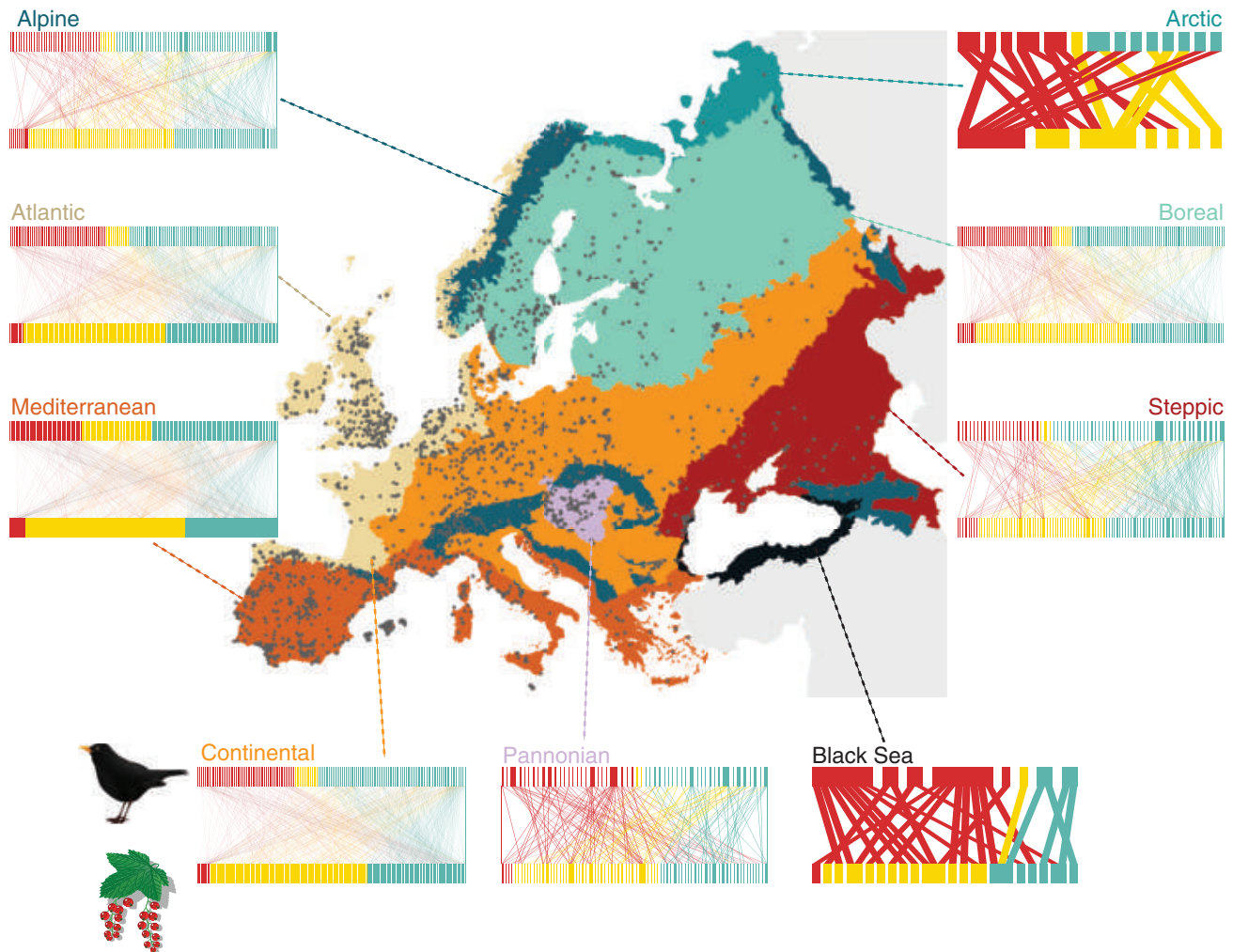
Fig. 1. Seed dispersal interactions across the European biomes. Points represent the location of frugivory and seed dispersal interactions. Networks represent the pairwise interactions between native and introduced seed dispersers (top boxes) and native plant species with biotic seed dispersal syndromes (bottom boxes). Network nodes and interactions are colored according to their conservation status: High Concern (red), Not Evaluated (yellow), and Low Concern (teal).

plant species on each focal disperser, i.e., animal species strength) (*28*) were highly correlated (Kendall correlation test, n = 398, τ = 0.696, P < 0.001), suggesting that this service is predominantly provided by generalist dispersers, which have a high species strength mainly because they disperse many species. The most important dispersers include abundant and widespread mammals such as red deer (*Cervus elaphus*, 119 plants), wild boar (*Sus scrofa*, 115 plants), red fox (*Vulpes vulpes*, 100 plants), sheep (*Ovis aries*, 99 plants), and birds: blackbird (*Turdus merula*, 105 plants), Eurasian magpie (*Pica pica*, 96 plants), and Eurasian blackcap (*Sylvia atricapilla*, 95 plants).

Each plant species had on average nine dispersers (median = 4; range 1 to 90). The plant species with the most dispersers were elder (*Sambucus nigra*, 90 dispersers), European blueberry (*Vaccinium myrtillus*, 88 dispersers), rowan (*Sorbus aucuparia*, 81 dispersers), sweet cherry (*Prunus avium*, 70 dispersers),

crowberry (*Empetrum nigrum*, 69 dispersers), common grape vine (*Vitis vinifera*, 69 dispersers), and red raspberry (*Rubus idaeus*, 65 dispersers). Most plants (63%) had 5 or fewer dispersers, and although plants' dependence on dispersers is highly variable and hard to estimate (*29*), losing any of them could hinder future plant recruitment (*7*). Dispersers' diversity is critical to secure effective seed dispersal services either because animal species provide complementary services (*30, 31*) or because they promote community redundancy (*32*). Plants with a highly diverse disperser assemblage are then expected to have an advantage in tracking rapidly changing climate, putting additional pressure on plants with reduced dispersal capacity (*33*). Moreover, new plant introductions, often facilitated by climate change, might divert seed dispersal services away from natives (*33*).

Fifty-seven percent of the native plants relied exclusively on native dispersers, 36% were

dispersed by native and non-native dispersers, and 7% were dispersed only by non-native animals, whereby the latter can ameliorate but unlikely fully compensate the loss of native dispersers (*34, 35*).

We found that the European biotic network was highly modular [modularity (M) = 0.41; P < 0.0001] and that dispersers were clustered into six distinct modules of densely interacting species (*36*). Eight of the dispersers acted as network hubs (species that are highly connected both within and across modules); 12 acted as module hubs (particularly well connected within their module); 30 acted as connectors (species that have many links with species on other modules); and the remaining animal species were peripherals (species with few interactions).

To estimate the sample coverage of our dataset, we built incidence matrices using individual bibliographic references as sampling units. Although no network is ever fully sampled (*37*), the biotic seed dispersal network
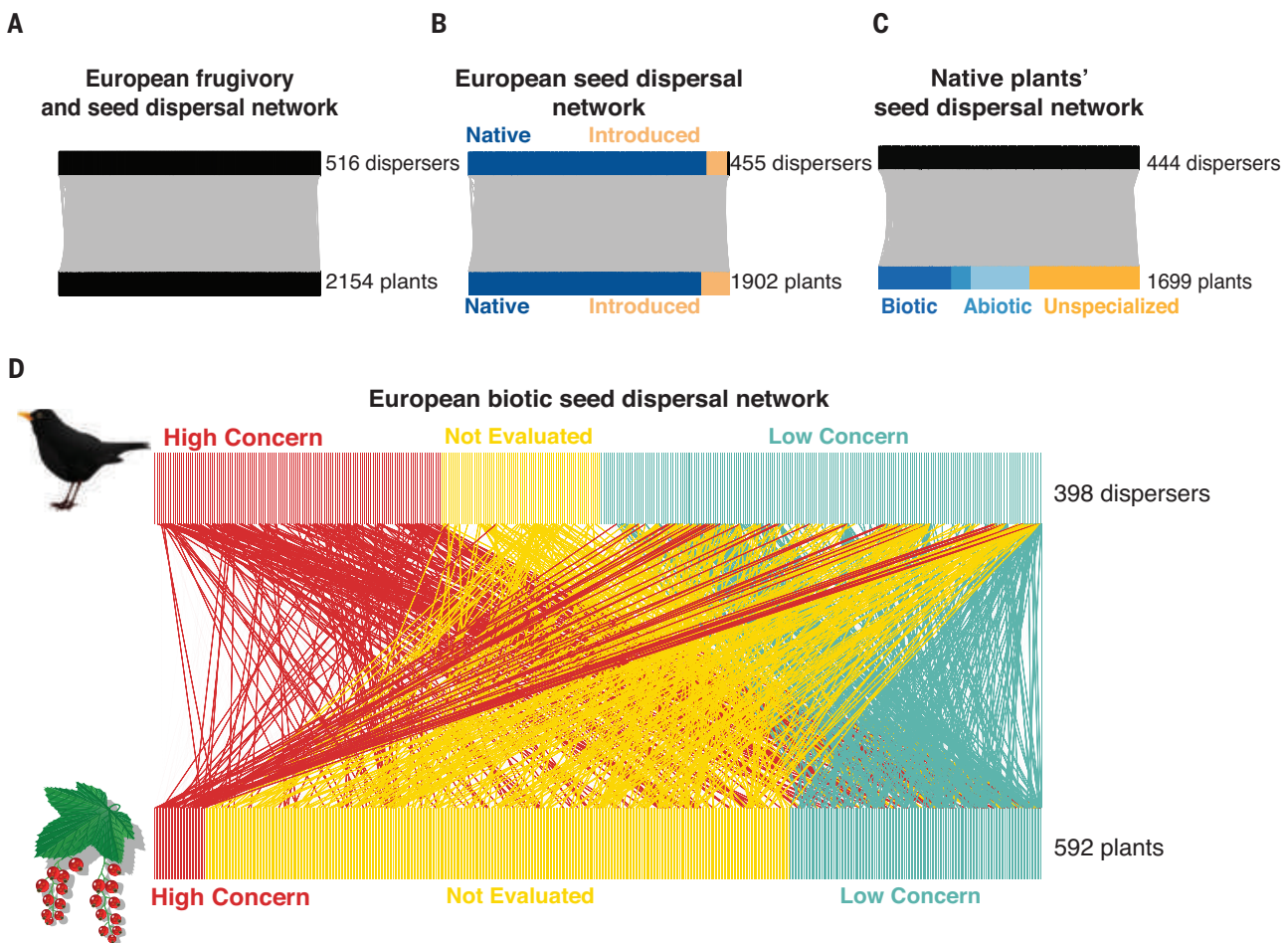
Fig. 2. Visualization of the aggregated European frugivory and seed dispersal networks. Top boxes represent seed dispersers, and bottom boxes represent plant species. (**A**) European frugivory and seed dispersal network, including all records of frugivory and potential seed dispersal independently from the level of confirmation of seed viability (see supplementary materials, materials and methods), but excluding those records where seed predation has been confirmed. (**B**) European seed dispersal network, including only interactions where seed viability has been experimentally confirmed or assumed based on the presence of intact seeds in fecal samples. Nodes with unknown origin are shown in black (top right). (**C**) European native plants' seed dispersal network, representing pairwise interactions between native and introduced seed dispersers and native plant species. Each plant is colored according to the presence of dispersal syndromes, and species with unspecialized diaspores are shown in orange. (**D**) European biotic seed dispersal network, including pairwise interactions between native and introduced dispersers and native plants with biotic dispersal syndromes. Network nodes and interactions are colored according to their conservation status.

had a very high sample coverage for dispersers (98%), plants (97%), and interactions (83%), which indicates that our review captured the vast majority of the published information on seed dispersal interactions. Sample coverage estimates for each biome varied between 57 to 95% for dispersers, 73 to 95% for plants, and 27 to 75% for interactions (table S1). These estimates exclude only the Arctic and Black Sea biomes because of their small extent and scarcity of studies (table S1). This level of sample coverage was only possible because of the directed literature searches for each animal species, which revealed many diet studies in which seeds are reported (38) but not included in specific seed dispersal studies.

Despite such high sample coverage, data are available for only 26% of the ~2248 native plant species with zoochorous traits, highlighting important knowledge gaps in seed dispersal research in Europe. To confirm this, we performed searches for 100 European plant species with zoochorous traits not present in our network (table S2) and found no new interactions, further suggesting that the dispersers of many of these rare and localized species remain unknown. Part of this gap could be explained by the predominant animal-centered nature of seed dispersal research, which should be complemented by plant-centered studies. To check whether this bias could be affecting our capacity to find information, we performed another set of targeted searches for 100 European plant species with zoochorous traits (table S3) already present in our network. Again, we detected no new interactions, which suggests that such bias is intrinsic to the literature but not affecting our capacity to find relevant published information. Therefore, identifying the seed dispersers for many understudied plant species—many of them with restricted and threatened populations—is key, even in one of the most studied continents.

A seed dispersal crisis in Europe

To assess the health of the seed dispersal services in Europe, we first obtained the conservation status and population trends from the International Union for Conservation of Nature (IUCN) Red List (23) for all species. Eleven percent of the dispersers were Threatened, 64% Not Threatened, and 25% Not Evaluated (table S4); 43% were Declining (undergoing population declines), 35% were Stable, and 22% were Increasing (tables S4 and S5). We then sorted species into three broader conservation categories by combining their conservation status and population trends: (i) "High Concern" for species listed as Near Threatened, Threatened, or Declining; (ii) "Not Evaluated"

Fig. 3. Distribution of seed disperser species according to their network role regarding the modular structure.
Each point represents a disperser species colored according to its conservation status: High Concern (red), Not Evaluated (yellow), and Low Concern (teal).
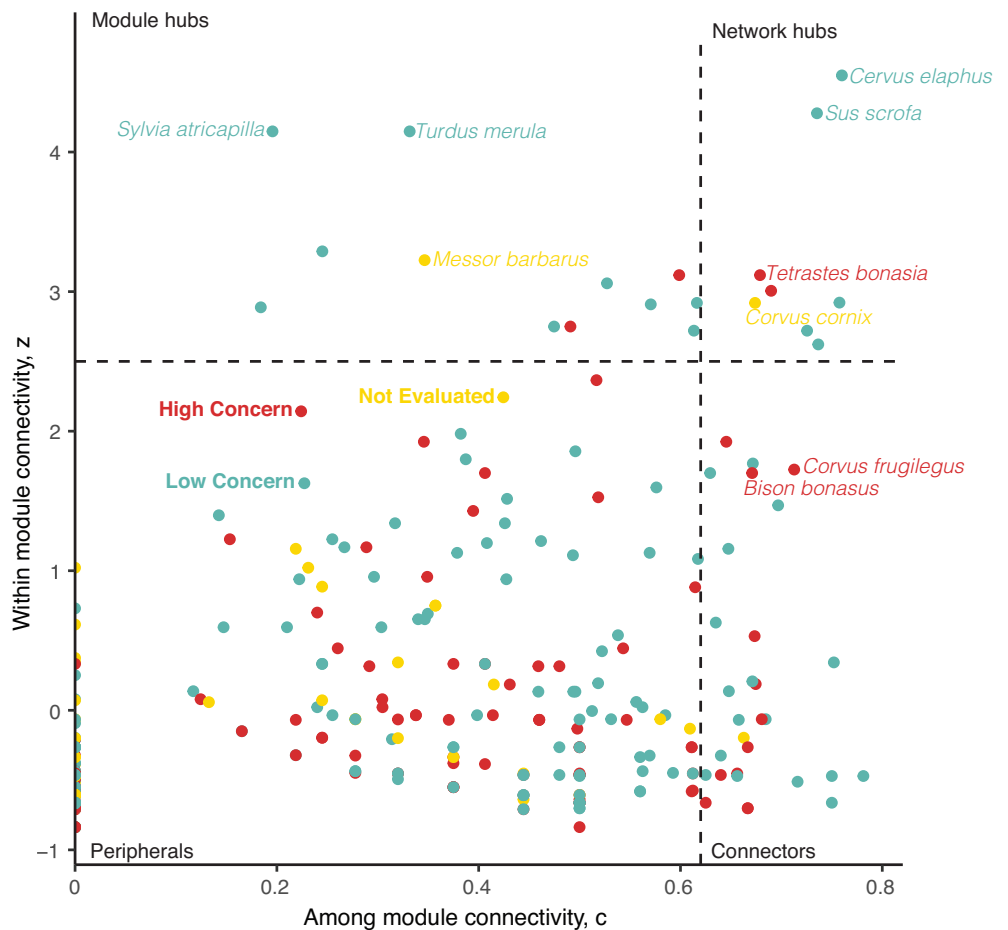
for Not Evaluated and data-deficient species with unknown population trends; and (iii) "Low Concern" for Not Threatened species with Stable, Increasing, or Unknown population trends. We found that for native plants, 33% of the dispersers were of High Concern, 18% were Not Evaluated, and 49% were of Low Concern (Fig. 2d).

The most important dispersers, in terms of degree (number of partner species) and species strength (a measure of the dependence of mutualistic partners on the focal species), included species of High Concern from several animal groups, including birds (garden warbler *Sylvia borin*, rook *Corvus frugilegus*, and redwing *Turdus iliacus*), mammals (European bison *Bison bonasus*, reindeer *Rangifer tarandus*, and European rabbit *Oryctolagus cuniculus*), reptiles (Lilford's wall lizard *Podarcis lilfordi*), and ants (European red wood ant *Formica polyctena* and Southern wood ant *F. rufa*). Furthermore, all European biomes had at least one-third of their dispersers classified as High Concern (table S6). Lastly, among network hubs, module hubs, and connector species, 32% were of High Concern (Fig. 3). The loss of High Concern dispersers that play essential network roles might accelerate the

fragmentation of the network structure (*36*). In turn, dispersal failure due to the loss of dispersers may disrupt plant recruitment, constrain gene flow, reduce genetic and phenotypic plant diversity, and truncate seed dispersal distances, which is particularly worrying given plants' need to track climate change (*7, 11*).

Only 1% of the biotically dispersed plants were evaluated as Threatened, while 32% were evaluated as Not Threatened, and most (67%) lacked IUCN assessment (table S4). Regarding population trends, 20% of the plant species were Declining, 75% were Stable, and only 5% were Increasing (tables S4 and S5). Therefore, 6% were of High Concern (e.g., English oak *Quercus robur*, French rose *Rosa gallica*, and Lapland buttercup *Ranunculus lapponicus*), 67% were Not Evaluated, and 27% were of Low Concern (Fig. 2D).

Our data shows that most native plant species (*n* = 357; 60%) had at least one disperser of High Concern and one-third (*n* = 190; 32%) had at least half of their evaluated dispersers in this category. These are important warning signs for the future of European plants, particularly considering that the seed dispersal service does not stop abruptly when the last disperser goes extinct but gradually erodes as the dispersers'

populations decline (*16, 39*). Indeed, animal species can become functionally extinct well before their actual extinction (*40, 41*). Although our dataset reveals no documented disperser extinction in the last four centuries, many High Concern species may no longer be effectively delivering the dispersal services compiled here.

We found that Low Concern animals dispersed more native plant species than did Not Evaluated (Tukey test, z = 5.26, P < 0.001) and High Concern dispersers (Tukey test, z = 3.84, P < 0.001) (fig. S1A and table S7). Similarly, Low Concern plants also had more dispersers than did High Concern plants (Tukey test, z = 2.25, P = 0.06) (fig. S1B and table S7). However, we found no differences in species strength between High Concern, Low Concern, and Not Evaluated dispersers (Kruskal-Wallis test, $\chi2$ = 0.99, df = 2, P = 0.611) (fig. S1C), suggesting that High Concern dispersers, many of which with small populations and reduced distributions, are just as important as those of Low Concern, despite the latter being typically more widespread and abundant.

Interaction conservation status

To assess interaction conservation status, we extended the framework developed for species
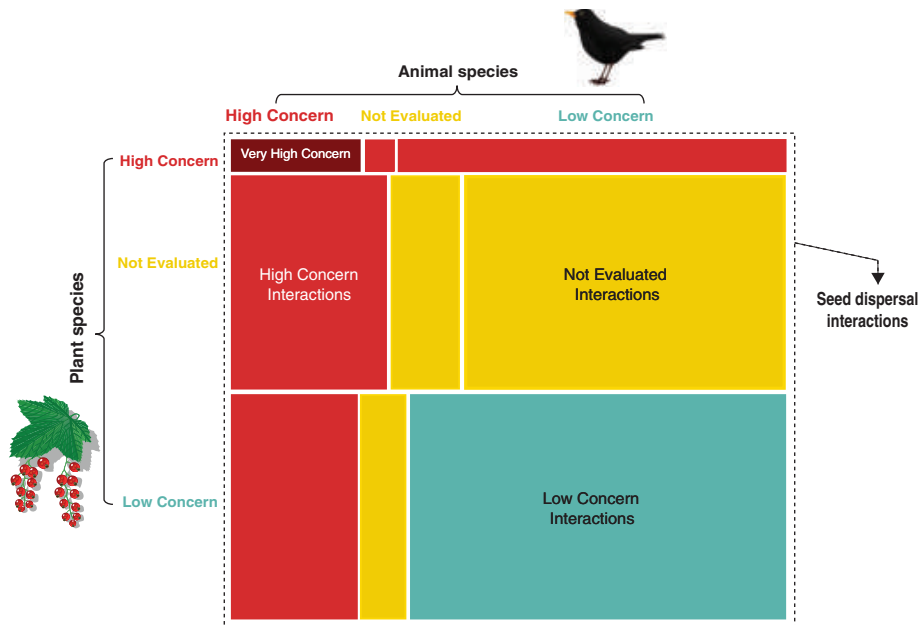
Fig. 4. Proportion of seed dispersal interactions per interaction conservation status. Mosaic plot depicting the proportion of seed dispersal interactions according to their conservation status as defined in supplementary materials, materials and methods: Very High Concern (dark red), High Concern (red), Not Evaluated (yellow), and Low Concern (teal) interactions.

conservation categories to classify every plant-disperser species interaction as: (i) "Very High Concern," if both interacting species were of High Concern; (ii) High Concern, if at least one of the partner species was Threatened or Declining; (iii) Not Evaluated, if at least one partner species was Not Evaluated and the other partner was not of High Concern; and (iv) Low Concern, if both interacting partners were of Low Concern. We found that 2% of the interactions were classified as Very High Concern, 29% as High Concern, 36% as Not Evaluated, and 33% as Low Concern (Fig. 4 and table S8). Our results showed that the proportion of High and Very High Concern interactions (31%) far exceeded that of High Concern species (17%) (χ^2 = 78.40, df = 1, P < 0.001), suggesting that the conservation status of interactions is likely to be a more sensitive and comprehensive indicator of community health than the conservation status of species alone. Indeed, conservation efforts should first concentrate on Very High Concern interactions, which are likely to be disrupted soon. This framework may provide a useful metric for guiding conservation and restoration strategies targeting ecosystem functioning (*42*), while bolstering regional collaboration among ecologists, conservation practitioners, and policy-makers.

Interaction conservation statuses were not homogeneously distributed across the European biomes (χ^2 = 163.39, df = 12, P < 0.001) (table S6). However, all biomes had a sizeable proportion of High Concern interactions, ranging from 25% in the Mediterranean to 54%

in the Pannonian, 72% in the Arctic, and 81% in the Black Sea biomes (Fig. 1 and table S6). The high prevalence of threatened interactions in some of the more geographically restricted, understudied, and degraded biomes in Europe (*43, 44*) clearly highlight the need to understand and protect the interaction networks that support their distinctive species assemblages.

A likely underestimated picture

Despite Europe's strong tradition in natural history, ecological research, and well-structured scientific and conservation communities, there are still critical knowledge gaps limiting our assessment of the health of seed dispersal services. Noticeably, the conservation status of 67% of the zoochorous plant species has not been assessed by the IUCN, and 73% have unknown European-level population trends. However, there is evidence that many of these species are declining, at least regionally (*14, 45*). For example, 70% of the plant species in Germany have declined in the last 60 years (*46*). This assessment bias is even more evident for invertebrate seed dispersers (mostly ants), 85% of which have still not been assessed by IUCN. This is particularly concerning because recent evidence shows that unassessed species are not a random subset of all species but tend to have restricted ranges and smaller populations, which are likely associated with unfavorable conservation status (*47*). Accordingly, the fraction of unevaluated species likely hides a much larger conservation threat for Euro-

pean dispersers, plants, and seed dispersal interactions (*48*). Additionally, our data revealed a significant difference in the conservation status of the dispersers of native and introduced plants with zoochorous traits (χ^2 = 14.096, df = 2, P < 0.001). Indeed, introduced plants have a higher proportion of Low Concern dispersers than do native plants (table S9), potentially facilitating their spread and exacerbating the competition for resources and seed dispersers for native plants (*33*).

Although our literature search was planned to minimize taxonomical biases, we identified a significant correlation between the number of source references from which their interactions were retrieved (i.e., number of papers with interactions of species i) and both the number of plant species dispersed by each animal and the number of dispersers per plant species (Kendall correlation test, n = 398, τ = 0.674, P < 0.001; n = 592, τ =0. 818, P < 0.001, respectively).

This bias reflects a research bias toward species with socioeconomic importance, broad distributions, high abundance, and large body size, which might blur the real magnitude of the seed dispersal crisis rather than inflating it. This bias is reflected in the high proportion of studies focused on Low Concern dispersers (75%), whereas High Concern dispersers were only reported in 38% of the studies, and Not Evaluated dispersers in 7% (some studies focused on more than one species, whereby the total percentage exceeds 100%). Because data are only available for 26% of the European zoochorous plant species, we performed bootstrap analysis to explore how the estimated proportion of High Concern interactions would vary on the basis of variable subsamples (ranging from 10 to 80%) of our original dataset. This analysis showed that even estimates based on as little as 10% of our full dataset would provide practically similar estimates (fig. S2), supporting the reliability of our estimates.

Conclusions

Despite the consensus about the importance of biotic interactions in general—and seed dispersal in particular—for ecosystem functioning (*1, 5*), ecologists and land managers still have incomplete knowledge about seed dispersal interactions and their conservation status. We showed evidence of a seed dispersal crisis in Europe that might have started with the Pleistocene megafauna extinctions (*49*) and that presents an uncertain future. This crisis is particularly worrying given that plants need to track rapidly shifting climatic envelopes in a continent strongly affected by habitat fragmentation (*9*). Our study provides compelling evidence for the need to prioritize the study and conservation of seed dispersal interactions that plants— and therefore people—depend upon.

REFERENCES AND NOTES

1. A. Perino et al., Science 364, eaav5570 (2019).
2. H. N. Ridley, The Dispersal of Plants Throughout the World (Reeve & Company, Limited, ed. 1, 1930).
3. IPBES, "Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services" (IPBES secretariat, 2019); https://doi.org/10.5281/zenodo.3831673.
4. J. M. Olesen, S. K. Jain, in Conservation Genetics, V. Loeschcke, S. K. Jain, J. Tomiuk, Eds. (Birkhäuser Basel, 1994), pp. 417–426.
5. A. Traveset, R. Heleno, M. Nogales, in Seeds: The Ecology of Regeneration in Plant Communities, M. Fenner, Ed. (CABI, ed. 3, 2014), pp. 62–93.
6. H. F. Howe, J. Smallwood, Annu. Rev. Ecol. Syst. 13, 201–228 (1982).
7. H. S. Rogers, I. Donoso, A. Traveset, E. C. Fricke, Annu. Rev. Ecol. Evol. Syst. 52, 641–666 (2021).
8. P. Vargas, R. Heleno, J. M. Costa, Biodivers. Data J. 11, e104079 (2023).
9. C. Estreguil, G. Caudullo, D. de Rigo, J. San Miguel, "Forest Landscape in Europe: Pattern, Fragmentation and Connectivity" (European Commission - Institute for Environment and Sustainability, 2012); https://publications.jrc.ec.europa.eu/repository/handle/JRC77295.
10. K. R. McConkey et al., Biol. Conserv. 146, 1–13 (2012).
11. E. C. Fricke, A. Ordonez, H. S. Rogers, J.-C. Svenning, Science 375, 210–214 (2022).
12. J. Dupuy et al., Ann. For. Sci. 77, 35 (2020).
13. J. Benedicto-Royuela et al., Conserv. Lett. 17, 12990 (2023).
14. W. A. Ozinga et al., Ecol. Lett. 12, 66–74 (2009).
15. C. J. Gardner, J. E. Bicknell, W. Baldwin-Cantello, M. J. Struebig, Z. G. Davies, Nat. Commun. 10, 4590 (2019).
16. K. H. Redford, Bioscience 42, 412–422 (1992).
17. J. P. González-Varo, R. G. Albaladejo, M. A. Aizen, J. Arroyo, A. Aparicio, J. Appl. Ecol. 52, 580–589 (2015).
18. European Commission, "European Red List of Birds 2021" (European Commission, 2022); https://doi.org/10.2779/959320.
19. M. Bilz, S. P. Kell, N. Maxted, R. V. Lansdown, European Red List of Vascular Plants (Publications Office of the European Union, 2011).
20. J. M. Tylianakis, E. Laliberté, A. Nielsen, J. Bascompte, Biol. Conserv. 143, 2270–2279 (2010).
21. A. Valiente-Banuet et al., Funct. Ecol. 29, 299–307 (2015).
22. R. Heleno, W. J. Ripple, A. Traveset, Web Ecol. 20, 1–10 (2020).
23. IUCN, "The IUCN Red List of Threatened Species," Version 2024-1 (2024); https://www.iucnredlist.org.
24. J. P. González-Varo, J. M. Arroyo, P. Jordano, Methods Ecol. Evol. 5, 806–814 (2014).
25. G. Escribano-Avila, C. Lara-Romero, R. Heleno, A. Traveset, Ecological Networks in the Tropics (Springer International Publishing, 2018).
26. R. Heleno, P. Vargas, Glob. Ecol. Biogeogr. 24, 518–526 (2015).
27. A. J. Green, C. Baltzinger, Á. Lovas-Kiss, Oikos 2022, oik.08327 (2022).
28. J. Bascompte, P. Jordano, J. M. Olesen, Science 312, 431–433 (2006).
29. E. C. Fricke, J. J. Tewksbury, E. M. Wandrag, H. S. Rogers, Proc. Biol. Sci. 284, 20162302 (2017).
30. P. D. Moore, Nature 414, 406–407 (2001).
31. E. W. Schupp, P. Jordano, J. M. Gómez, New Phytol. 188, 333–353 (2010).
32. J. P. González-Varo et al., Proc. Natl. Acad. Sci. U.S.A. 120, e2302440120 (2023).
33. R. H. Heleno, in Plant Invasions: The Role of Biotic Interactions, A. Traveset, D. M. Richardson, Eds. (CABI, 2020), pp. 256–269.
34. J. M. Heinen et al., Nat. Commun. 14, 1019 (2023).
35. R. H. Heleno et al., Oikos 2022, oik.08279 (2022).
36. J. M. Olesen, J. Bascompte, Y. L. Dupont, P. Jordano, Proc. Natl. Acad. Sci. U.S.A. 104, 19891–19896 (2007).
37. P. Jordano, Funct. Ecol. 30, 1883–1893 (2016).
38. J. B. Kiss, J. Rékasi, Sesiunea Muzeului Banatului At: Timişoara 1, 133–140 (1982).
39. B. Rumeu et al., Funct. Ecol. 31, 1910–1920 (2017).
40. T. Säterberg, S. Sellman, B. Ebenman, Nature 499, 468–470 (2013).
41. K. R. McConkey, D. R. Drake, Ecology 87, 271–276 (2006).
42. M. A. Palmer, J. B. Zedler, D. A. Falk, in Foundations of Restoration Ecology, M. A. Palmer, J. B. Zedler, D. A. Falk, Eds. (Island Press/Center for Resource Economics, 2016), pp. 3–26.
43. K. Sundseth, "Natura 2000 in the Steppic Region" (Office for Official Publications of the European Communities, 2009); https://doi.org/10.2779/7833.
44. K. Sundseth, "Natura 2000 in the Pannonian region" (Office for Official Publications of the European Communities, 2009); https://doi.org/10.2779/79432.
45. G. Niedrist, E. Tasser, C. Lüth, J. Dalla Via, U. Tappeiner, Plant Ecol. 202, 195–210 (2009).
46. D. Eichenberg et al., Glob. Change Biol. 27, 1097–1110 (2021).
47. J. Borgelt, M. Dorber, M. A. Høiberg, F. Verones, Commun. Biol. 5, 679 (2022).
48. C. Finn, F. Grattarola, D. Pincheira-Donoso, Biol. Rev. Camb. Philos. Soc. 98, 1732–1748 (2023).
49. M. Davoli et al., Glob. Ecol. Biogeogr. 33, 34–47 (2024).
50. S. B. Mendes, R. Heleno, Supporting data and code for Mendes et al. 2024 Evidence of a European seed dispersal crisis, version 2, Figshare (2024); https://doi.org/10.6084/m9.figshare.25901920.

ECOLOGICAL SPECIATION

Ecological speciation in Darwin's finches: Ghosts of finches future

Jeffrey Podos[1,2]* and Katie M. Schroeder[2]

The theory of ecological speciation posits that adaptive divergence among incipient species raises incidental barriers to reproduction, thus catalyzing the emergence of new species. In this study, we conducted an experimental test of this theory in Galápagos finches, a clade in which beaks and mating songs are mechanistically linked. We forecasted the acoustic structure of songs for a set of possible evolutionary futures (successive droughts spurring increasingly large beaks) and, in a field assay, presented resulting song simulations to territorial males. We found that responses to songs dropped off after six simulated drought events, to degrees roughly comparable to drops in response to songs that diverged through cultural drift and acoustic adaptation. Our results support, in Darwin's finches, the feasibility and mechanistic bases of an ecological speciation hypothesis.

S peciation, the process by which new species emerge, often has an ecological component (1–3). In one widely recognized mechanism of "ecological speciation," divergent trajectories of morphological adaptation among incipient species enhance probabilities of subsequent reproductive isolation by driving incidental changes to cues or signals that animals use for mate recognition (4–9) (Fig. 1A). Cases of ecological speciation are often recognized retrospectively, in correlations among adaptive and signaling traits, after speciation is already complete (5, 6). By contrast, we lack predictive data regarding how and when, as speciation proceeds, ecological divergence and its impacts on mating traits actually elevate species recognition barriers.

One group of animals for which ecological factors have been long recognized as foundational to speciation are Darwin's finches of the Galápagos Islands, Ecuador. Divergence in this adaptive radiation has centered largely on beak form and function, which we now know adapt through natural selection to local food resources and interspecific competitors (10–16). In theory, divergence in Darwin's finch beaks could enhance, as a secondary impact, species discrimination and reproductive isolation among diverging incipient species (16, 17). One reason for this is that beaks are not only the primary loci of finch adaptive evolution, they also shape the production of these birds' vocal mating signals. While singing, Darwin's finches (like many songbirds) modulate their beak gapes in synchrony with vocal source (syrinx) activity, presumably to maintain effective resonance filtering and tonality across varying source frequencies (18–20). Accordingly, evolutionary changes in finch beaks and seed-crushing

[1]Department of Biology, University of Massachusetts Amherst, Amherst, MA 01003, USA. [2]Graduate Program in Organismic & Evolutionary Biology, University of Massachusetts Amherst, Amherst, MA 01003, USA.
*Corresponding author. Email: jpodos@umass.edu
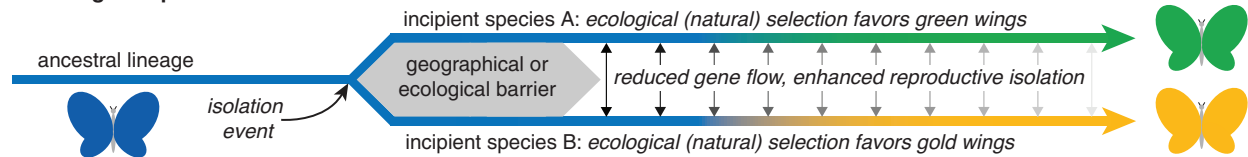
Fig. 1. The process of ecological speciation. (**A**) Ecological speciation begins when an ancestral lineage (blue line, evolving from left to right, illustrated for a hypothetical lineage of butterflies) experiences an isolation event, such as when encountering a new geographic or ecological barrier (gray shading) that splits the ancestral lineage into multiple descendant lineages. These lineages, now incipient species, may experience distinct ecological environments and thus divergent trajectories of natural selection and phenotypic evolution. In the example shown, natural selection favors wing colors that enhance survival in the butterflies' new habitats, for example, by minimizing detection by predators. Ecological speciation ensues if diverging adaptive traits or genetically correlated (pleiotropic) traits also mediate mate choice, for example, if female butterflies prefer to mate with like-colored males (green with green or gold with gold). In such cases, diverging lineages should experience increasingly reduced gene flow on secondary contact, thus enhancing interlineage reproductive isolation and fostering speciation. (**B**) Ecological speciation in Darwin's finches is expected because a primary locus of natural selection, beak form and function, can invoke knock-on divergence in acoustic mating signals, given that beak movements mediate not just feeding but also song production. In one of numerous possible routes for ecological speciation in the finches, an ancestral line splits owing to interisland dispersal. The original population, now incipient species A, experiences stabilizing natural selection and maintains its prevailing beak size, whereas the dispersed population, now incipient species B, encounters drought conditions and evolves corresponding adaptations, larger beaks and stronger jaws, for husking larger, tougher seeds. Accordingly, birds in incipient species B are predicted to experience linked evolutionary reductions in performance-related vocal traits (such as trill rate and frequency bandwidth), which could, on secondary contact, reduce interlineage acoustic species recognition and thus enhance reproductive isolation. Finch sketches were reproduced from Darwin (*54*).

capacities should influence, as a secondary consequence, the structure of some song traits, especially those requiring rapid or pronounced beak gape modulations [(*21–23*); see (*24–27*) for parallel examples in other Neotropical bird clades]. Given beak-song linkages and the central role of mating songs in Darwin's finch species recognition (*16*), ecological divergence among incipient species could, in principle, enhance finches' ability to distinguish diverging lineages by song and could consequently foster speciation (Fig. 1B). This proposed mechanism, however, has not yet been subject to direct empirical tests (*28, 29*). We do not yet know how much ecological change and matched evolution of beaks and songs would be required to elevate barriers to reproductive isolation. We also have been unable to assess the impacts on species recognition of beak-related song divergence, as per the ecological speciation hypothesis, relative to other factors, especially cultural drift and acoustic adaptation, that can also drive song divergence and reproductive isolation.

In this work, we start to fill these gaps for a population of Darwin's medium ground finches (*Geospiza fortis*) at El Garrapatero, Santa Cruz Island, Galápagos. This population features discrete small and large morphs that have diverged markedly in body and beak dimensions (*30*). These morphs are favored over intermediate-sized birds through disruptive selection (*31, 32*), specialize on partially distinct sets of food resources (*33*), sing songs that match their divergent beak morphologies (*22*), and mate assortatively by morph (*34*). In a prior study (*35*) with this population, territorial males responded more strongly to playback of songs of their own morph, consistent with a hypothesis of ecological speciation (*21, 23, 34*). However, that study did not isolate the functional (communicative) relevance of song variation linked specifically to beak divergence, as the playback songs in that study (same or different morph) differed acoustically not just because of beak divergence but also because of cultural drift, i.e., the accumulation of song-copying inaccuracies or innovations across generations (*36*), which, with sufficient time and isolation, can also cause songs to diverge and thus enhance species discrimination. Songs of songbirds in isolated lineages can also diverge through independent trajectories of acoustic adaptation, which is selection for efficient transmission through distinct acoustic environments that the incipient species occupy, for example, through the evolution of low-frequency, slowly repeated notes in woodland habitats (*37–39*). Beak divergence, cultural drift, and acoustic adaptation can thus be regarded as alternative drivers of vocal reproductive isolation (*40–42*), and sorting among them is necessary for assessing the relative weight in the Darwin's finch radiation of the proposed ecological speciation mechanism.

Projecting beak-related impacts on song structure

As a first step in our study, we constructed song stimuli to be used for playback trials that incorporated acoustic changes that we anticipated would occur along one possible course of future evolution in beak morphology. Toward this end, we selected a sample of songs in our study population recorded previously (*22*) from birds of known beak sizes, and then for each, we generated three digitally modified songs, "ghosts of finches future," that projected expected impacts on song structure that would in principle be caused by one, three, or six cumulative future drought events (Fig. 2). Our process was guided by three observations: (i) A classic long-term study of *G. fortis* from Daphne

Fig. 2. Logic and method for constructing song playback stimuli to test the ecological speciation hypothesis, illustrated here for 1 of our 12 stimulus sets. (Left) The gray dots and gray line represent vocal deviation (a measure of vocal performance) as a function of beak depth for *G. fortis* at our study site. The plot was constructed from data collected and analyzed during a prior study (*22*). Higher vocal deviation values correspond to lower vocal performance

(*21*). The orange square marks one male finch's beak depth (11.61 mm) and the corresponding vocal deviation value for his song; digital recordings of his song provided the basis for constructing this stimulus set. Blue squares, increasingly dark in hue, mark projected beak depths and corresponding vocal deviations for this male and his song, following one, three, and six cumulative simulated drought events [magnitudes of beak evolution follow (*43*)]. Larger beaks, under positive selection in drought years to crack tough seeds, are associated with reduced vocal performance and manifest here in larger vocal deviations. (Right) Sketches of beaks (scaled to indicate simulated changes in relative beak depth) plus spectrograms of playback songs for control (top row) and simulated songs (remaining rows) for birds with increasingly large beaks and corresponding projected reductions in vocal performance. Eleven other stimulus sets were constructed from the songs of 11 other finches with different beak sizes. Beak sketches were adapted from drawings by Bowman (*11*).
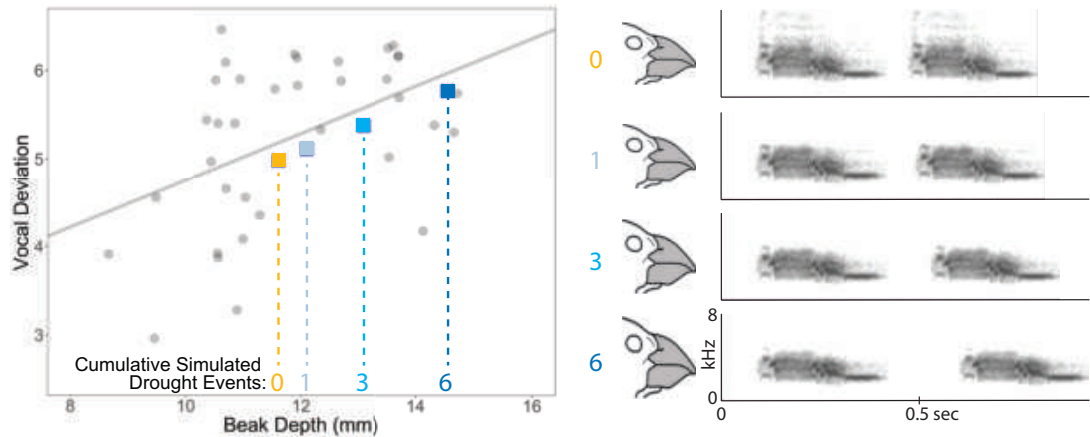
Major showed that droughts tend to spur natural selection for larger beaks (*12*, *14*); (ii) the evolution of larger beaks and associated increases in bite force capacity (*23*) should diminish birds' vocal performance and, in particular, birds' abilities to match source frequency changes with vocal tract modulations required for rapid trills with broad frequency bandwidths [vocal parameters that can be captured in a single performance metric, "vocal deviation" (*21*, *22*)]; and (iii) as morphological impacts of droughts accumulate, songs should evolve toward increasingly diminished trill rates and frequency bandwidths, diverging incrementally from present-day songs. The average change in beak depth after six simulated drought events, following values reported by Grant and Grant (*43*), approximates actual degrees of separation in beak depth at our study site between both (a) the small and large *G. fortis* morphs and (b) the large *G. fortis* morph and *Geospiza magnirostris* (*30*).

Field test of the ecological speciation hypothesis

We presented song stimuli to 12 *G. fortis* males using a simulated territorial intrusion assay. Each male was presented with playback of a natural song from another male of its own size morph plus that song's three simulated versions (Fig. 2, right). The four song treatments (control, one-drought event, three-drought events, and six-drought events) were presented to each bird over the course of a week; birds heard one treatment per test day in 3-min blocks, with each test day separated by a nontest day to limit carry-over effects. Treatment order was

counterbalanced across the experiment, and all birds were tested with a unique stimulus set. As in three prior studies with this species on this island (*35*, *44*, *45*), stronger aggressive responses to playback are interpreted as reflecting greater stimulus salience. Guided by those studies, we measured response strength primarily with respect to birds' flight behavior (latency to fly off the starting perch, flight rate, and proximity to the playback speaker), as birds appeared to search for the simulated intruder. We also measured latencies and rates of vocal responses (songs), even though the three prior studies noted above generally did not detect effects of playback treatment on finches' vocal activity. Our main prediction was that beak-related changes in song structure, as a stand-alone factor, would cause detectable and increasingly severe decrements in birds' responses to playback because such songs should be less likely to be recognized and categorized by subjects as belonging to their own species. A major alternative potential outcome was that the acoustic changes implemented would not impact vocal species recognition. Such an outcome would imply that finches base their assessment of species identity on acoustic features not linked to beak evolution.

Our first main finding was that birds responded less strongly to the most extreme stimulus manipulation (six-drought songs) relative to control (zero-drought) songs (Fig. 3 and table S1). In terms of flight latencies (Fig. 3A and table S1), birds hearing control songs left their starting perches much earlier than when they heard six-drought songs (estimated mar-

ginal means, 16.2 versus 63.4 s; z-score = 2.383, P = 0.02). In flight rates, birds flew 31% less often in response to six-drought songs than to control songs (Fig. 3B) (z-score = −2.094, P = 0.04). In approach distances, birds challenged with control songs approached the speaker more closely, by an average of 1.4 m for perched distances (Fig. 3C) (z-score = 1.864, P = 0.06) and 2.2 m when also accounting for overflights of the playback speaker (Fig. 3D) (z-score = 2.066, P = 0.04). The flight and approach results support our prediction that beak-related changes to song would lead to diminished responses. By contrast and consistent with prior studies with this species on this island (*35*, *44*, *45*), no strong effects of treatment were found for singing latencies or rates (fig. S1). The overall stronger response to control songs versus six-drought songs was corroborated in a principal components analysis that included all response variables (table S2) (principal component (PC) 1, z-score = −2.068, P = 0.04). These results together support the feasibility of an ecological mechanism in driving species recognition among diverging finch species, at least when beak divergence is substantial, here approaching natural levels of divergence among related ground finch species or morphs. It is also clear that our study subjects did recognize the six-drought simulated songs as salient, even though their responses were diminished. This implies that the acoustic modifications we introduced to songs did not completely obstruct cross-lineage recognition.

Next, we considered the intensity of finches' responses to one- and three-drought playback
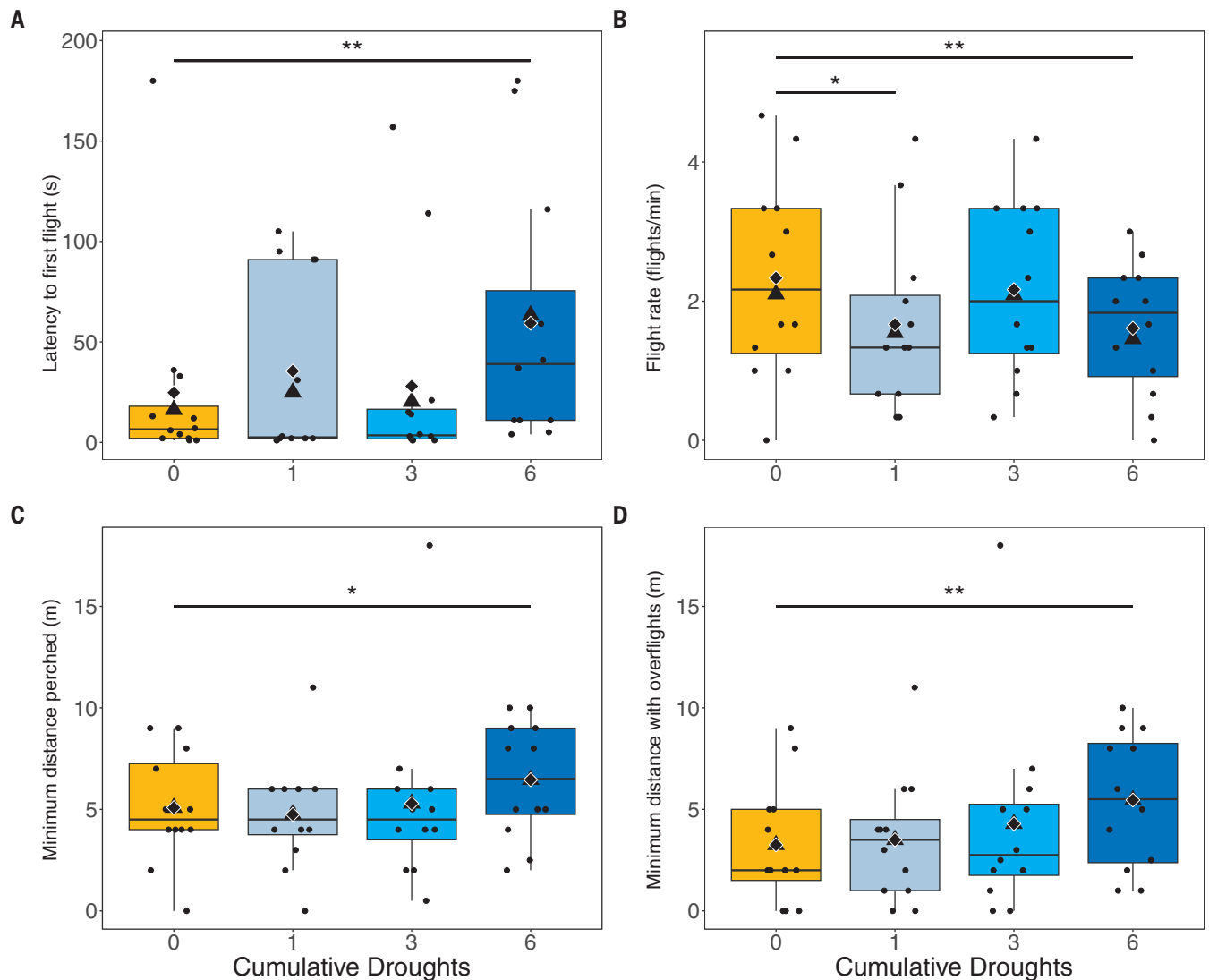
Fig. 3. Flight responses of territorial male *G. fortis* to playback of control songs (zero-drought treatment) and simulated songs (one-, three-, or six-drought treatment). Flight responses include latency to first flight (**A**), flight rate (**B**), minimum distance perched (**C**), and minimum distance with overflights (**D**). Treatment colors correspond to those in Fig. 2. Dots refer to individual data points; hinges of the box and whisker plots refer to 25th and 75th percentiles; middle lines, triangles, and diamonds, respectively, describe medians, estimated marginal means, and arithmetic means; and whiskers denote data ranges, excluding outliers >1.5× the interquartile range from the hinge. $N = 12$ *G. fortis* males. *$P < 0.1$; **$P < 0.05$.

songs relative to control and six-drought songs. It was possible that decrements in response would accumulate in roughly linear fashion, in concordance with the graded nature of the acoustic changes that we introduced across drought treatments. Another possible outcome was that birds' responses would drop off immediately for one- and three-drought songs, relative to their responses to control songs. Such an outcome would suggest a particularly strong (and evolutionarily early) impact of beak-related vocal evolution on species recognition. Yet another potential outcome, that birds would respond to one- and three-drought songs with degrees of vigor similar to that for control songs, would imply that song retains lineage-specific signatures even

during mild to moderate changes in beaks and associated song attributes. Such an outcome would imply that functional impacts of beak-song linkages on species recognition emerge only toward later stages of incipient species divergence. Our results support the third interpretation. Virtually all response levels did not differ between playback of control songs and their one- or three-drought song counterparts (Fig. 3 and table S1). One weak exception to this pattern concerned flight rates (Fig. 3B), for which birds flew 26% less often in response to one-drought songs than to control songs (z-score = –1.784, $P = 0.07$). This exception was also captured by the analysis of PC3 (z-score = –2.250, $P = 0.02$), which loaded heavily for numbers of flights (table

S2); by contrast, no differences across zero-, one-, and three-drought treatments were detected in analyses of the remaining PCs (1, 2, 4, and 5). Our results together suggest that beak-driven changes to song on their own would require a substantial degree of divergence, corresponding to between three and six drought events, to enable vocal discrimination among diverging finch species, at least as indicated here in the responses of territorial males.

Contribution of beak-related song divergence to vocal species recognition

Our next goal was to weigh the impact on species recognition of song evolution related to beak divergence relative to impacts related to
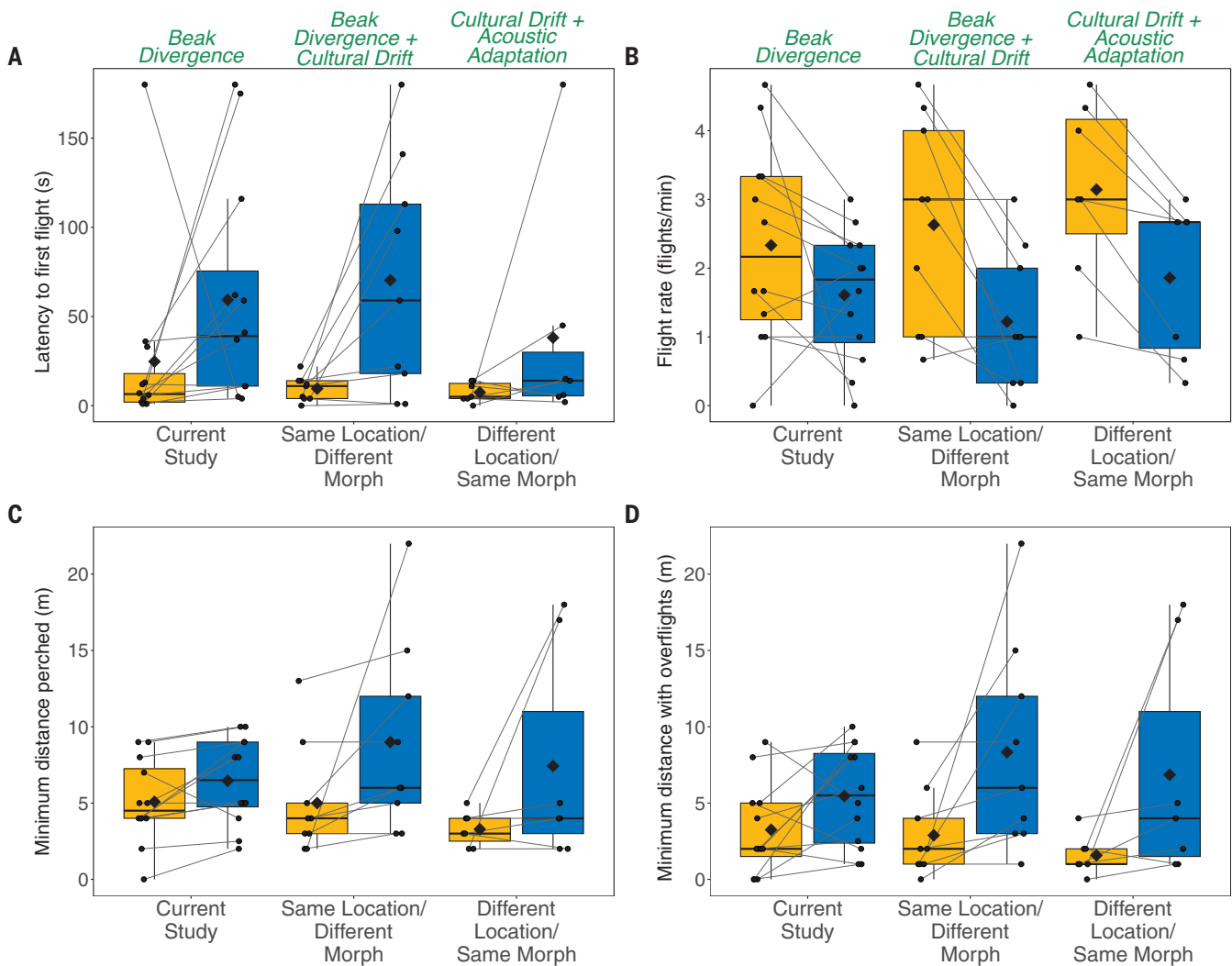
Fig. 4. Flight responses of territorial male *G. fortis* at El Garrapatero to playback of control songs and to three sets of probe songs, compiled across studies. The control song treatments (same location/same morph) are shown in orange, and the probe song treatments are shown in blue. For the current study, control and probe songs differed owing to simulated beak divergence only (data also shown in Fig. 3; zero versus six); for the same location/different morph comparison [data from (*35*)], control and probe songs differed owing to beak divergence plus cultural drift; and for the different location/same morph comparison [data from (*35*)], control and probe songs differed owing to cultural drift plus acoustic adaptation. Weaker responses to probe songs versus control songs are evident across the board in lower values for flight rate (**B**) but in higher values for the remaining response parameters (**A**, **C**, and **D**). Greater contrasts in responses to control versus probe songs imply more severe loss of recognition of the probe songs. Dots refer to individual data points with gray lines connecting trials on the same individual; hinges of the box and whisker plots refer to the 25th and 75th percentiles; middle lines and diamonds describe, respectively, medians and arithmetic means; and whiskers describe the range of data, excluding outliers >1.5× the interquartile range from the hinge.

cultural drift and acoustic adaptation. Toward this end, we evaluated outcomes from the current study (control versus six-drought songs; Fig. 4) relative to outcomes generated in a prior playback study (*35*) conducted with the same population, playback protocol, and repeated measures design. In that study, birds' responses to control songs were compared to their responses to probe songs from the same location but of the other morph (same location/different morph; Fig. 4) or from a different location (Borrero Bay, Santa Cruz Island) but of the same morph (*35*) (different location/same morph; Fig. 4).

For the same location/different morph comparison, we can presume that control and probe playback songs differed from each other acoustically not just owing to beak-related song divergence but also owing to distinct morph-specific trajectories of cultural drift, given that (i) young birds likely only learn and introduce copy errors to songs from adults of their own morphs and that (ii) females tend to prefer mates of their own type, thus favoring true-to-morph song learning (*34*). By contrast, for this comparison, we can negate any impact of differential acoustic adaptation, as the two morphs at El Garrapatero overlap in habitat, some-times singing from the same perches. In this comparison, we observed elevated contrasts of responses to probe versus control song categories for all parameters relative to the contrasts reported in the current study. These elevated contrasts, also revealed in a comparison of effect sizes (table S4, compare columns 1 and 2), delineate an extra gain in vocal discrimination capacities enabled by the interaction of beak divergence and cultural drift in driving song divergence.

For the different location/same morph comparison, we can presume that control and probe songs differed from each other owing to both

cultural drift and acoustic adaptation to distinct habitats, yet without any impact of beak-related song divergence. We can make this latter presumption for two reasons: all birds recorded at Borrero Bay were small-morphed, and we restricted our consideration of response data to only small-morphed subjects. Figure 4 and table S4 (compare columns 1 and 3) illustrate that the interaction of cultural drift and acoustic adaptation bears strongly on species recognition, markedly more so than beak-related song divergence alone yet marginally less so than the interaction of beak-related song divergence and cultural drift. We also evaluated response data for a fourth control versus probe song comparison from a study in which territorial males at El Garrapatero were presented with control songs and probe songs from yet another locality, Academy Bay (44). In that study, however, beak size morphs were not accounted for in either the stimuli presented or the study subject tested, which limits the value of cross-study comparisons. Nevertheless, a notable pattern emerged (fig. S2): the contrast in responses to control versus probe songs was more pronounced than in each of the other three comparisons, consistent with the fact that probe songs differed from control songs not just because of cultural drift and acoustic adaptation but also, to some extent, their interactions with beak divergence. As more factors contribute to song divergence, more acoustic cues for song-mediated species recognition presumably emerge, which should further enhance cross-species vocal recognition.

Ecological speciation in Darwin's finches

Speciation in many animal clades, including Darwin's finches, is catalyzed by the evolutionary divergence of communication cues or signals that, on secondary contact, can facilitate species recognition, assortative mating, and reproductive isolation (1–9, 40–42). Understanding probabilities and dynamics of speciation thus requires attention to factors that drive cue or signal divergence and to when and how animals discriminate among diverging cue or signal variants. For Darwin's finches, we can narrow our focus down to vocal signals given a series of observational and experimental studies that have demonstrated a central role for song and its divergence in species recognition and assortative mating (16, 46–49). Thus, the central pending question for testing an ecological speciation hypothesis in Darwin's finches is this: As incipient finch species diverge, what types and magnitudes of divergence in song will enable species discrimination and, accordingly, assortative mating and reproductive isolation?

Most broadly, our results support the feasibility of an ecological speciation hypothesis for *G. fortis*. Cumulative effects on song of three to six 1977-level drought events, predicted

to drive increases in beak depth of about 1.5 to 3.0 mm and corresponding changes in vocal trill rate and frequency bandwidth, appear sufficient to disrupt vocal recognition among diverging lineages. We are not aware of prior studies that have linked anticipated evolutionary adaptive responses to dynamic environmental conditions, corresponding expected changes in the structure of cues or signals used for mate recognition, and the potential for species recognition and behavioral isolation among lineages inhabiting those environments. Moreover, our comparisons of response patterns and effect sizes across studies (Fig. 4 and table S4) indicate how beak-related vocal divergence works alongside cultural drift and acoustic adaptation as drivers of species recognition.

These inferences might be regarded as conservative for at least three reasons. First, our study focused solely on responses to playback of males, whereas the most important decisions about breeding and mate choice in Darwin's finches are made by females. One might assume that patterns of vocal discrimination in males should apply directly to females, which is a case that has already been made for Darwin's finches (49). Yet, in general, females might evolve to be more discriminating in their assessment of vocal signals, as the main fitness costs of errors made in species recognition are generally more severe for females (cross-species breeding) than for males (momentarily defending their territories against heterospecific males) (50–52). If so, we might predict that female Darwin's finches' responses to song will be dampened earlier in the course of beak and song structure divergence. A second reason that our inferences might be conservative is that, as beaks diverge, songs will likely diverge not just in trill rate and frequency bandwidth, as manipulated here, but also in more fine-scale performance-related phonological dimensions (53) that we did not attempt to represent in our simulated songs. A third reason is that finches might also assess beak divergence and thus species identities visually. Available evidence suggests that in the absence of vocal signals, finches indeed modulate their responses to simulated territorial intruders in accordance with intruders' beak sizes (17). Redundancy of visual and vocal signals should further enhance accurate species recognition and thus foster speciation earlier in divergence than implied by our song-only data (29). Ultimately, quantifying how ecological divergence elevates reproductive barriers requires experiments such as ours that test how the animals themselves respond to ecologically selected cue and signal variants.

REFERENCES AND NOTES

1. D. Schluter, *The Ecology of Adaptive Radiation* (Oxford Univ. Press, 2000).
2. J. A. Coyne, H. A. Orr, *Speciation* (Sinauer Associates, Inc, 2004).
3. T. Price, *Speciation in Birds* (Roberts & Company Publishers, 2008).
4. S. Gavrilets, *Fitness Landscapes and the Origin of Species* (Princeton Univ. Press, 2004).
5. D. Schluter, *Science* 323, 737–741 (2009).
6. P. Nosil, *Ecological Speciation*, Oxford Series in Ecology and Evolution (Oxford Univ. Press, 2012).
7. C. D. Jiggins, R. E. Naisbit, R. L. Coe, J. Mallet, *Nature* 411, 302–305 (2001).
8. D. W. Pfennig, K. S. Pfennig, *Evolution's Wedge: Competition and the Origins of Diversity* (Univ. of California Press, 2012).
9. A. P. Hendry, *Eco-Evolutionary Dynamics* (Princeton Univ. Press, 2017).
10. D. Lack, *Darwin's finches* (Cambridge Univ. Press, 1947).
11. R. I. Bowman, *Morphological differentiation and adaptation in the Galapagos finches* (Univ. of California Publications in Zoology, vol. 58, 1961).
12. P. T. Boag, P. R. Grant, *Science* 214, 82–85 (1981).
13. D. Schluter, P. R. Grant, *Am. Nat.* 123, 175–196 (1984).
14. P. R. Grant, B. R. Grant, *Science* 313, 224–226 (2006).
15. M.-O. Beausoleil *et al.*, *Evolution* 77, 2533–2546 (2023).
16. P. R. Grant, B. R. Grant, *How and Why Species Multiply: The Radiation of Darwin's Finches* (Princeton Univ. Press, 2008).
17. L. M. Ratcliffe, P. R. Grant, *Anim. Behav.* 31, 1139–1153 (1983).
18. S. Nowicki, *Nature* 325, 53–55 (1987).
19. J. Podos, J. A. Southall, M. R. Rossi-Santos, *J. Exp. Biol.* 207, 607–619 (2004).
20. T. Riede, R. A. Suthers, N. H. Fletcher, W. E. Blevins, *Proc. Natl. Acad. Sci. U.S.A.* 103, 5543–5548 (2006).
21. J. Podos, *Nature* 409, 185–188 (2001).
22. S. K. Huber, J. Podos, *Biol. J. Linn. Soc. Lond.* 88, 489–498 (2006).
23. A. Herrel, J. Podos, B. Vanhooydonck, A. P. Hendry, *Funct. Ecol.* 23, 119–125 (2009).
24. N. Seddon, *Evolution* 59, 200–215 (2005).
25. E. P. Derryberry *et al.*, *Evolution* 66, 2784–2797 (2012).
26. N. A. Mason, K. J. Burns, *Biol. J. Linn. Soc. Lond.* 114, 538–551 (2015).
27. N. C. Garcia, P. L. Tubaro, *Auk* 135, 262–275 (2018).
28. M. J. Ryan, *Nature* 409, 139–140 (2001).
29. J. Podos, S. Nowicki, *Bioscience* 54, 501–510 (2004).
30. A. P. Hendry *et al.*, *Proc. Biol. Sci.* 273, 1887–1894 (2006).
31. A. P. Hendry, S. K. Huber, L. F. De León, A. Herrel, J. Podos, *Proc. Biol. Sci.* 276, 753–759 (2009).
32. M.-O. Beausoleil *et al.*, *Proc. R. Soc. B Biol. Sci.* 286, 20192290 (2019).
33. L. F. De León, G. Rolshausen, E. Bermingham, J. Podos, A. P. Hendry, *Evol. Ecol. Res.* 14, 365–380 (2012).
34. S. K. Huber, L. F. De León, A. P. Hendry, E. Bermingham, J. Podos, *Proc. Biol. Sci.* 274, 1709–1714 (2007).
35. J. Podos, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 1031–1039 (2010).
36. P. J. B. Slater, *Trends Ecol. Evol.* 1, 94–97 (1986).
37. H. Brumm, M. Naguib, *Adv. Stud. Behav.* 40, 1–33 (2009).
38. J. A. Tobias *et al.*, *Evolution* 64, 2820–2839 (2010).
39. R. H. Wiley, *Noise Matters: The Evolution of Communication* (Harvard Univ. Press, 2015).
40. J. Podos, S. K. Huber, B. Taft, *Annu. Rev. Ecol. Evol. Syst.* 35, 55–87 (2004).
41. M. R. Wilkins, N. Seddon, R. J. Safran, *Trends Ecol. Evol.* 28, 156–166 (2013).
42. E. P. Derryberry *et al.*, *Ecol. Evol.* 8, 1890–1905 (2018).
43. P. R. Grant, B. R. Grant, *Evolution* 49, 241–251 (1995).
44. J. Podos, *Anim. Behav.* 73, 833–844 (2007).
45. J. Podos, R. Dybboe, M. O. Jensen, *Curr. Zool.* 59, 8–19 (2013).
46. L. M. Ratcliffe, P. R. Grant, *Anim. Behav.* 33, 290–307 (1985).
47. B. R. Grant, P. R. Grant, *Evolution* 50, 2471–2487 (1996).
48. P. R. Grant, B. R. Grant, *Am. Nat.* 149, 1–28 (1997).
49. B. R. Grant, P. R. Grant, *Biol. J. Linn. Soc. Lond.* 76, 545–556 (2002).
50. W. A. Searcy, E. A. Brenowitz, *Nature* 332, 152–154 (1988).
51. W. A. Searcy, S. Nowicki, M. Hughes, S. Peters, *Am. Nat.* 159, 221–230 (2002).

52. J. E. Danner et al., Am. Nat. 178, 53–63 (2011).
53. D. M. Logue, J. A. Sheppard, B. Walton, B. E. Brinkman, O. J. Medina, Bioacoustics 29, 709–730 (2019).
54. C. Darwin, Journal of Researches into the Natural History and Geology of the Countries Visited during the Voyage of H.M.S. Beagle Round the World, under the Command of Capt. Fitz Roy, R.N (John Murray, ed. 2, 1845).
55. J. Podos, K. M. Schroeder, Ecological Speciation in Darwin's Finches: Ghosts of Finches Future, Dryad (2024); https://doi.org/10.5061/dryad.k3j9kd5dh.
56. J. Podos, K. M. Schroeder, DFDrought_FullScript.R, Zenodo (2024); https://doi.org/10.5281/zenodo.13348739.

NEUROPSYCHIATRY

Somatic mosaicism in schizophrenia brains reveals prenatal mutational processes

Eduardo A. Maury[1,2,3]†, Attila Jones[4,5]†, Vladimir Seplyarskiy[6,7]†, Thanh Thanh L. Nguyen[8,9], Chaggai Rosenbluh[4], Taejong Bae[10], Yifan Wang[10], Alexej Abyzov[10], Sattar Khoshkhoo[1,3,11], Yasmine Chahine[1,3], Sijing Zhao[1], Sanan Venkatesh[5], Elise Root[8], Georgios Voloudakis[12], Panagiotis Roussos[12], Brain Somatic Mosaicism Network‡, Peter J. Park[6], Schahram Akbarian[13,14], Kristen Brennand[8,9], Steven Reilly[8], Eunjung A. Lee[1,3], Shamil R. Sunyaev[6,7]*, Christopher A. Walsh[1,3,15,16]*, Andrew Chess[4,5,14]*

Germline mutations modulate the risk of developing schizophrenia (SCZ). Much less is known about the role of mosaic somatic mutations in the context of SCZ. Deep (239×) whole-genome sequencing (WGS) of brain neurons from 61 SCZ cases and 25 controls postmortem identified mutations occurring during prenatal neurogenesis. SCZ cases showed increased somatic variants in open chromatin, with increased mosaic CpG transversions (CpG>GpG) and T>G mutations at transcription factor binding sites (TFBSs) overlapping open chromatin, a result not seen in controls. Some of these variants alter gene expression, including SCZ risk genes and genes involved in neurodevelopment. Although these mutational processes can reflect a difference in factors indirectly involved in disease, increased somatic mutations at developmental TFBSs could also potentially contribute to SCZ.

Schizophrenia (SCZ) has a substantial genetic component, with common variants (minor allele frequency >1%) of individually small effect, as well as rare copy number variants (CNVs) and single-nucleotide variants (SNVs) with larger effects, all contributing to genetic risk (1). Somatic variants, which occur throughout development and hence are present in a fraction of cells in the body (2, 3), are familiar drivers of cancer but are increasingly recognized as contributing to neurodevelopmental conditions including focal epilepsy (4, 5) and autism spectrum disorders (ASDs) (6, 7). Recent work implicates somatic CNVs (sCNVs) in a fraction of SCZ cases (8), whereas the contribution of somatic SNVs (sSNVs) remain largely unexplored.

Study design and variant discovery

We analyzed somatic variants directly from postmortem brain, using deep WGS of DNA extracted from NeuN+ neurons of dorsal lateral prefrontal cortex (DLPFC) from 61 individuals with a diagnosis of SCZ and 25 neurotypical controls (Fig. 1A and table S2; materials and methods) to specifically capture mutations occurring during early prenatal development. Because neocortical neurons are all postmitotic by ~30 gestational weeks (9), somatic mutations clonally shared by neurons occur in progenitor cells prior to 30 weeks and are not confounded by postgestational clonal mutations.

Individuals were determined to be of European and African ancestry on the basis of principal component analysis (fig. S1A). Polygenic risk score (PRS) normalized by ancestry revealed that, as expected, individuals with SCZ had a higher PRS for disease than controls (Kolmogorov-Smirnov test $P = 0.0086$, fig. S1B). Brain tissue was homogenized, and nuclei were stained for NeuN and subjected to fluorescence-activated nuclei sorting (FANS) using standard methods (10). DNA extracted from 500,000 to 1,000,000 nuclei was sequenced without amplification (materials and methods). Median genome coverage of ~239× showed no significant difference in coverage between cases and controls (Wilcox rank-sum test, $P = 0.38$, Fig. 1B).

Somatic SNVs were identified using best practices of the Brain Somatic Mosaicism Network (BSMN), which offers high sensitivity (11, 12). The final call-set of 3286 sSNVs (2424 in SCZ and 862 in controls, table S3) showed variant allele fractions (VAFs) from 0.92 to 39.7%. We randomly selected 111 variants to validate with 96 having enough coverage for orthogonal amplicon-based sequencing (materials and methods). Ninety out of 96 sSNVs validated (94%) with VAFs that were highly correlated with WGS estimates ($R^2 = 0.87$, Fig. 1C), and there were no differences in validation between cases (62 validated, 5 not) and controls (28 validated, 1 not) (Fisher exact test $P = 0.66$). One outlier SCZ sample showed 188 mutations without technical anomalies or unconventional nucleotide substitution patterns (13); because this high mutational burden could dominate downstream statistics, the sample was excluded from all counts and further analyses.

[1]Division of Genetics and Genomics, Manton Center for Orphan Disease, Boston Children's Hospital, Boston, MA 02115, USA. [2]Bioinformatics and Integrative Genomics Program and Harvard/MIT MD-PHD Program, Harvard Medical School, Boston, MA 02115, USA. [3]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [4]Department of Cell, Developmental and Regenerative Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. [5]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. [6]Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA. [7]Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA. [8]Department of Genetics, Yale School of Medicine, New Haven, CT 06520, USA. [9]Department of Psychiatry, Yale School of Medicine, New Haven, CT 06520, USA. [10]Department of Quantitative Health Sciences, Center for Individualized Medicine, Mayo Clinic, Rochester, MN 55905, USA. [11]Department of Neurology, Brigham and Women's Hospital, Boston, MA 02115, USA. [12]Center for Disease Neurogenomics, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. [13]Department of Psychiatry and Neuroscience, Friedman Brain Institute, Mount Sinai, New York, NY 10029, USA. [14]Department of Neuroscience, Friedman Brain Institute, Mount Sinai, New York, NY 10029, USA. [15]Departments of Pediatrics and Neurology, Harvard Medical School, Boston, MA 02115, USA [16]Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA 02115, USA

*Corresponding author. Email: andrew.chess@mssm.edu (A.C.); christopher.walsh@childrens.harvard.edu (C.A.W.); ssunyaev@hms.harvard.edu (S.R.S.)
†These authors contributed equally to this work.
‡Brain Somatic Mosaicism Network authors and their affiliations are listed in table S1.
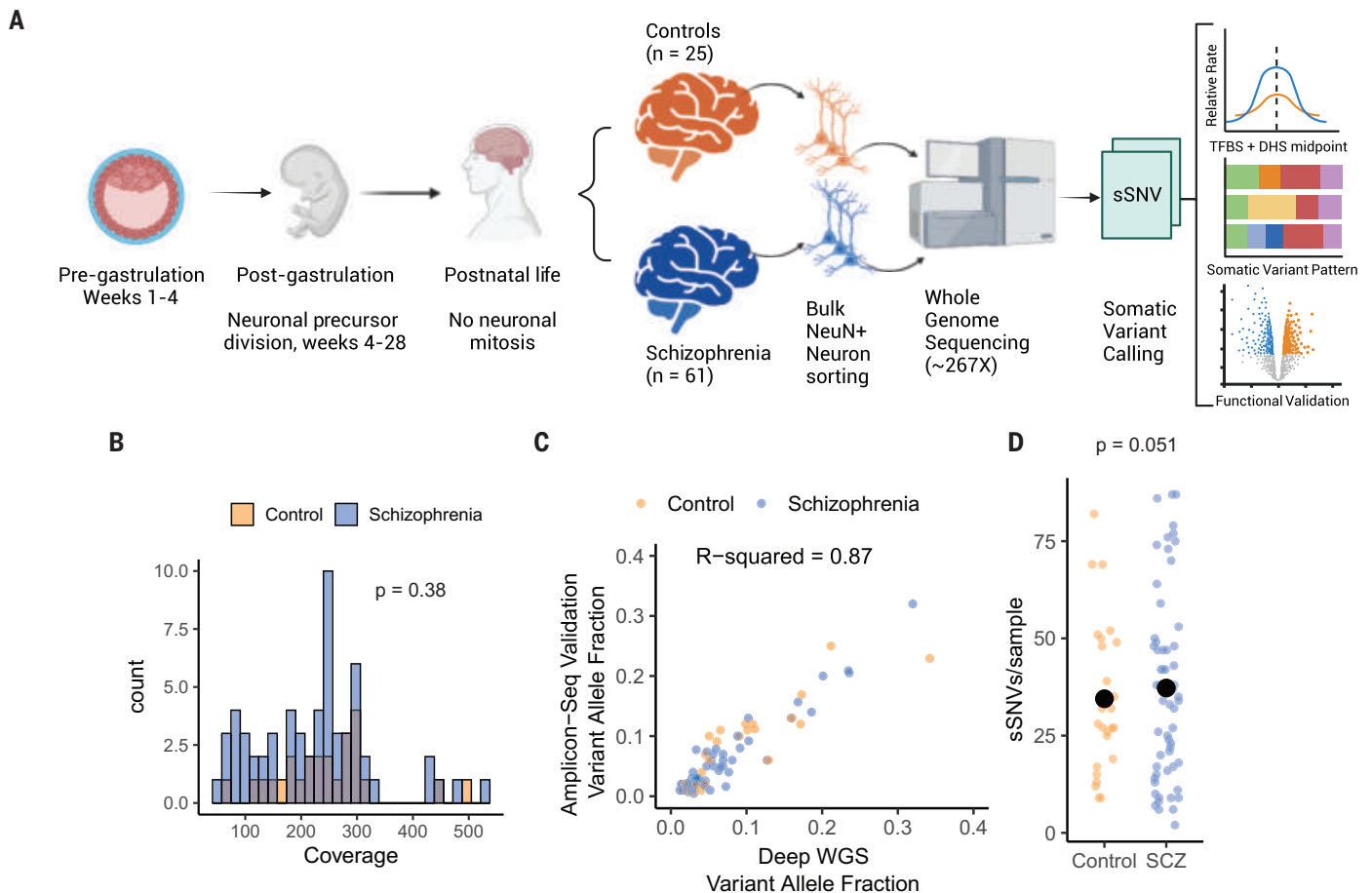
Fig. 1. Experimental design and orthogonal validation. (**A**) Schematic of experimental and analysis design. Neuronal clonal somatic mutations that are shared across neurons originate during prenatal brain development, occurring either before organogenesis (pregastrulation) or during neuronal proliferation during neurogenesis, resulting in somatic variants present in cells across multiple tissues. Mutations occurring postnatally in neurons are not clonal and hence are undetectable with this method. (**B**) Histogram of average sequencing coverage for schizophrenia cases and control samples. (**C**) Scatter plot of deep WGS VAF for variant submitted for validation and the VAF from the validation amplicon sequencing from SCZ and controls samples. R-squared value was computed from ordinary linear regression model. (**D**) Scatter plot of number of sSNVs per sample for schizophrenia cases and control after removal of an outlier SCZ case with 188 sSNVs. Large black points represent the sample medians. The P value was calculated using permutation-based negative binomial stepwise regression (materials and methods).

Genome-wide sSNV burden in cases and controls

After exclusion of the outlier SCZ sample, genome-wide sSNV counts in remaining SCZ cases averaged 37.3 per sample compared to 34.5 in controls, which did not achieve statistical significance using permutation-based negative binomial regression ($P = 0.051$, Fig. 1D; materials and methods). For each permutation, we randomly shuffled diagnosis labels and ran a forward negative binomial stepwise regression model to account for ancestry principal components, sex assigned at birth, and technical covariates (sequencing facility, coverage, year of autopsy, age of death, cause of death, postmortem interval, and institution where diagnosed). Regression analysis was performed on 45 SCZ cases and 19 controls with information across all covariates, including ancestry principal components (table S4). As expected, age was not associated with higher sSNV per sample ($P > 0.05$, fig. S1C), emphasiz-

ing that identified clonal variants occurred prenatally in neuronal precursors, remaining static after birth. Although permutation provides uninflated P values, power analysis suggests that with mutation rates increased <1.7-fold in SCZ versus controls, as observed here, this test provides low power to detect significant differences (fig. S1, D and E).

One SCZ case showed a sCNV overlapping intron 1 and potentially exon 2 of *SORCS2* (fig. S2, A and B), which is implicated in attention-deficit hyperactive disorder (ADHD), and bipolar disorder (*14*, *15*), though roles of *SORCS2* in SCZ are not established. sSNVs were not enriched in genome-wide association study (GWAS) loci associated with SCZ (binomial regression, $P = 0.936$). Exonic sSNVs (87 total, 2.6%, including the outlier sample) were equally common in cases (1.02 per individual) and controls (1.00 per individual, $P = 1$, Fisher's exact test), and we did not detect somatic stop-gain, splice-site altering, or missense variants

at genes implicated in SCZ in germline de novo or rare variant studies (*16*) in our small sample (table S3). However, we did find a stop-gain T>G sSNV on exon 1 of *STX12/13* (chr1: 28099835, T>G, p.L6*), a highly constrained gene [probability of heterozygous loss intolerance (pLI) of 0.96 (*17*)] that encodes an endosomal synaptic transport protein.

Higher sSNV rate at active transcription factor binding sites (TFBSs) in SCZ

Analysis of sSNV distribution across the genome, using fetal brain tracks from Roadmap Epigenomes (*18*), revealed increased sSNVs in open chromatin regions in SCZ compared to controls. Previous comparison of ASD to controls showed enrichment of sSNVs at open chromatin regions (*7*, *12*). We found higher sSNV rates in SCZ versus controls at fetal brain deoxyribonuclease (DNase) hypersensitive sites (DHSs), which are indicative of open chromatin (binomial regression, $P = 0.0015$,
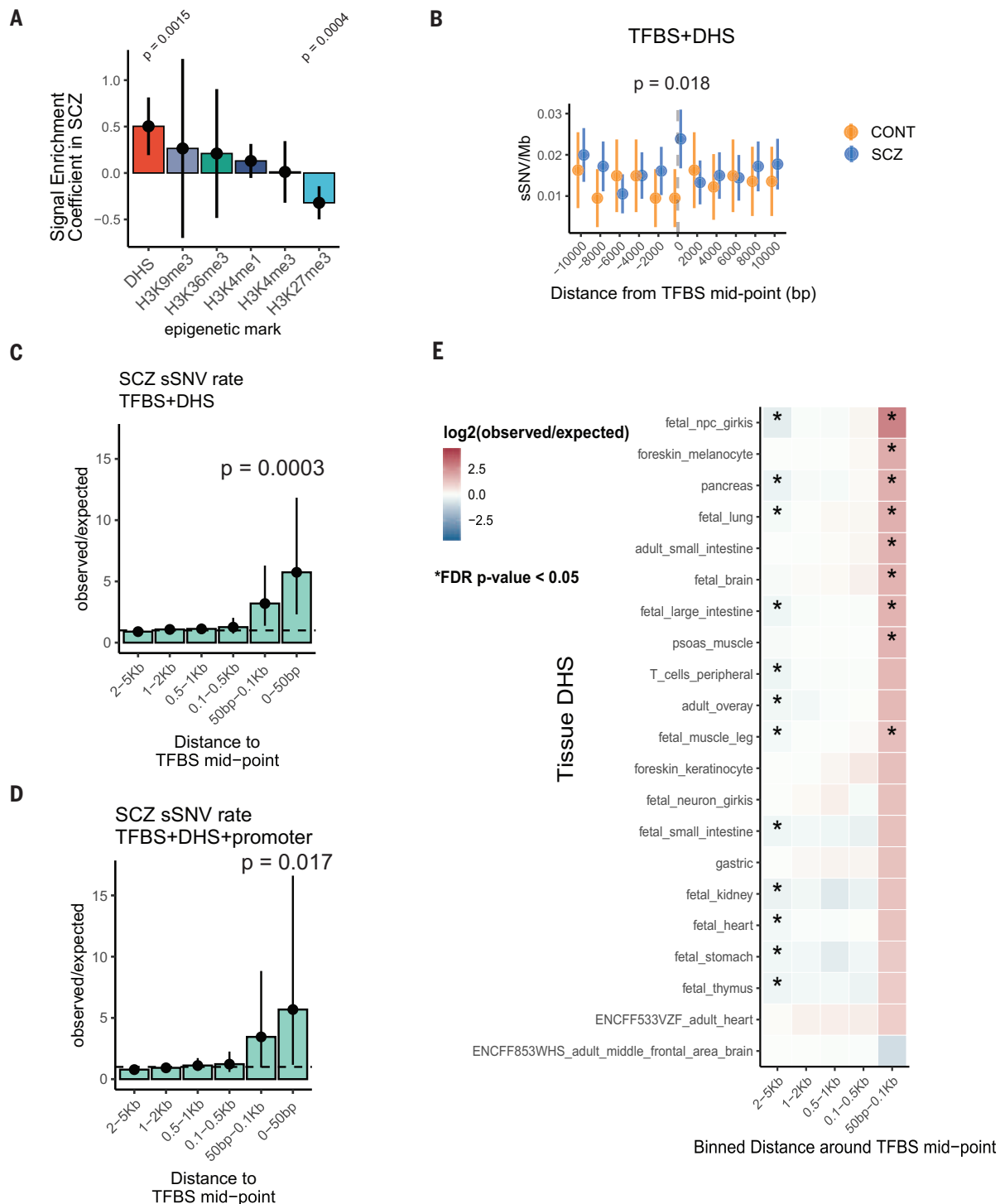
Fig. 2. Increased sSNV rate at developmentally active transcription factor binding sites (TFBSs in SCZ. (**A**) Bar plot of binomial regression interaction term between epigenomic tracks and disease status. Positive values indicate enrichment in SCZ and negative values indicate depletion. Line ranges indicate 95% CIs from binomial regression. (**B**) Somatic SNV rate at ±10-kb region from active TFBSs in fetal brain (TFBS+DHS) in SCZ and controls. (**C** and **D**) Bar plot of observed over expected mutation rate at binned regions around TFBSs in SCZ. (**E**) Heatmap of rate ratios in SCZ at TFBSs using different DHS tracks. For (B), (C), (D), and (E), *P* values and CIs were calculated using Poisson tests. For (E), stars indicate statistical significance at the false discovery rate adjusted *P* < 0.05 level.

Fig. 2A). Conversely, we found lower sSNV rates in SCZ at trimethylated histone H3 lysine 27 (H3K27me3) regions, which are associated with down-regulation of genes and closed chromatin (*19*) (binomial regression, *P* = 0.0004,

Fig. 2A). To ensure that the genome-wide overdispersion of sSNVs did not inflate these statistics, we obtained a null *P*-value distribution by permuting diagnosis labels. This empiric null *P*-value distribution was very close to the

expected null, suggesting robustness to overdispersion (fig. S3A). We did not detect casecontrol differences in sSNV rate at regions of increased fetal brain gene expression, nor a systemic transcriptional strand bias (fig. S3B).

We also did not find significant association between sSNV rate and replication timing or replication fork direction (fig. S3C).

Previous studies in cancers observed enriched sSNVs at active TFBSs overlapping DHSs, owing to hindrance of DNA repair by bound transcription factors (TFs) [20–22]. To test whether a similar phenomenon could explain the local increase in sSNVs at DHS regions in SCZ, we calculated sSNV rates near the midpoint of TFBSs, accounting for the number of genomes and sites sampled in each SCZ case (materials and methods). We aggregated the hg19 TFBS BED (browser extensible data) files from Vorontsov et al. [23] using human TF tracks with highest reliability and reproducibility (A tracks). These tracks aggregate across experimental designs and tissues, so that they are not tissue specific. We used the top 10% of DHS intensity from fetal brain tracks of Roadmap Epigenomes [18] to obtain likely active TFBSs. We observed increased sSNVs near (±1 kb) the midpoint of active TFBSs in SCZ cases compared to controls (Poisson test, rate ratio RR = 2.51 [1.12:6.62], P = 0.018, Fig. 2B). Results were robust to DHS intensity threshold (fig. S4, A and B). No individual TF achieved statistical significance after multiple hypothesis correction.

Further genome-wide analysis of SCZ sSNVs, comparing rates near active TFBSs to expected genome-wide rates after accounting for trinucleotide context, revealed 5.74-fold enrichment within 50 base pairs (bp) from the TFBS mid-point (P = 0.0003, Fig. 2C), and 5.68-fold enrichment near promoters (P = 0.017, Fig. 2D), with effects fading >100 bp from the TFBS mid-point, suggesting highly localized mutational processes. Similar enrichment was observed across DHS intensity cut-offs, with increasing effect sizes with higher DHS signal (fig. S4, C and D). This rate comparison is across genomes of SCZ only, excluding effects of sequencing or hereditary differences. We observed enrichment of sSNV at TFBSs across DHSs from multiple tissues, developmental stages, and embryonic germ layers (including 10 fetal and 10 adult; table S5 and Fig. 2E), suggesting a pattern that is developmental, but not tissue specific. No similar enrichment was observed in controls.

Specific sSNV patterns at TFBSs in SCZ

Two specific base substitution patterns were observed in SCZ cases but not in controls. Somatic SNVs at CpG sites showed 24.0-fold enrichment of CpG>GpG substitutions at active TFBSs at promoters compared to the expected C>G genome-wide rate accounting for trinucleotide context (2 observed, 0.083 expected) [95% confidence interval (CI), 2.90 to 86.5, P = 0.0047, Fig. 3A and fig. S5A]. C>G and C>A transversions at CpG contexts characterize a known mutational process (component 11, Fig. 3B) [24] reflecting enzymatic demethylation,

which involves resection of oxidated methylcytosine, creating an abasic site [25] (Fig. 3C). Replication of the abasic site before repair creates CpG transversions [26]. One CpG>GpG variant in SCZ was near the promoter of GRN, encoding the essential, dosage-sensitive protein progranulin (Fig. 3D). GRN haploinsufficiency causes frontotemporal dementia in adults and pediatric neuronal degeneration [27] and has been reported in SCZ [28].

In addition to enrichment of CpG transversions at active TFBSs in SCZ (observed/expected 14.5, 95% CI [1.76:52.57], P = 0.0086, Fig. 3E), we observed a similar trend in bulk brain DNA from ASD cases analyzed previously [7] (observed/expected 6.92, 95% CI [0.18:38.6], P = 0.13, Fig. 3E). By contrast, we did not find CpG transversions ≤100 bp from the mid-point of TFBSs in (i) our control samples, (ii) control samples from the WGS ASD cohort [7], or (iii) a recent nondiseased twin study (Fig. 3E) [29], so that CpG transversions at TFBSs were increased in SCZ versus aggregated controls (P = 0.013, binomial test). By contrast, somatic CpG transversions at CpG islands showed similar rates in SCZ, ASD, and controls (Fig. 3F), suggesting that CpG transversions in CpG islands are not disease associated.

Relative rates of sSNVs across base changes at non-CpG sites in SCZ samples, although accounting for trinucleotide context, showed a highly localized increase in T>G substitutions within 100 bp from the TFBS midpoint versus genome-wide expectation (observed = 3, expected = 0.09, observed/expected = 34.3, 95% CI [7.08:100.3], P = 1.04 × 10^{-4}, fig. S5B), which was further enhanced near promoters (observed = 2, expected = 0.024, observed/expected = 82.6, 95% CI [10.0:298.4], P = 2.88 × 10^{-4}, Fig. 4A). T>G sSNVs in active TFBSs showed significant enrichment in cases versus the aggregated-control sample above (P = 0.0032, binomial test). Genome-wide T>G mutations also represented a higher proportion of sSNVs in SCZ versus control (permutation Fisher's exact test, odds ratio OR = 2.23, P < 0.0001, fig. S6).

Of note, three unrelated pairs of SCZ cases showed the exact same T>G substitution at the exact same genomic position (Fig. 4B), which we call same variant same site (SVSS) recurrence. We saw no somatic T>G SVSS recurrence in controls or other deep WGS samples, including for ASD [7]. We confidently validated two out of three pairs of T>G variants through orthogonal amplicon sequencing; for the third pair, we only had DNA from one individual, which showed positive validation (Fig. 4B). The exceedingly low probability (Poisson test, P = 1.22 × 10^{-11}) of observing three recurrent sSNVs by chance suggests that mutational processes driving these T>G mutations are highly localized, with T>G mutational hotspots showing an estimated sSNV

rate ~1.44 × 10^5 times the expected genome-wide rate (materials and methods).

Analysis of T>G mutations across the Pan Cancer Analysis of Whole Genomes [PCAWG [30]] suggested potential mechanisms for T>G mutagenesis at TFBSs. We found similarly high rates of T>G mutations at TFBSs in a subset of liver and bladder cancer samples, with six liver and two bladder cancer samples showing strong T>G mutation enrichment at TFBSs (>5-fold; Fig. 4C red circles). Similar to SCZ, these liver and bladder cancer samples showed SVSS recurrence, which was enriched in samples with a high sSNV rate at TFBSs versus those with a low sSNV rate at TFBSs (~100- versus ~10-fold, respectively; Fig. 4, D and E). Three of the six liver and both bladder samples showing SVSS recurrence carried somatic missense mutations in XPD, a key DNA repair gene, in line with observations that XPD dysfunction can increase sSNVs at TFBSs [31]. Cancer samples with XPD mutations were also enriched in T>G mutations at TFBSs versus noncarriers (Wilcoxon rank-sum test, P = 6.3 × 10^{-6}, Fig. 4D).

The trinucleotide mutational spectra at TFBSs of liver and bladder cancers showing high TFBS mutation rates and XPD deficiency were very similar to that of SCZ at active TFBSs (Fig. 4F). Despite notably converging patterns between liver/bladder cancer and SCZ, we did not find XPD somatic or germline mutations in SCZ, nor T>G mutations in common, though XPF (ERCC4), which encodes another core nucleotide excision repair gene, is a reproducible SCZ GWAS hit [32, 33]. Thus, mosaic SCZ mutations may be driven by factors mimicking XPD dysfunction, such as other factors inhibiting DNA repair, though perhaps the similar mutation spectra are coincidental.

Functional interrogation of somatic variants

We used massively parallel reporter assays (MPRAs) in a human neuroblastoma cell line (SK-N-SH; materials and methods) to assess gene regulatory impacts of the full set of sSNVs identified in cases and controls (Fig. 5A and tables S3 and S6). MPRA regulatory activity measurements were highly reproducible (five replicates' pairwise Pearson's r ≥ 0.99, fig. S7) and recapitulated known SK-N-SH positive and negative controls (fig. S8) [34, 35]. The rate of somatic mutations causing significant expression modulation (emVars) did not differ by diagnosis (permutation Fisher's exact test P = 0.49). Variants were equally likely to up- or down-regulate expression (fig. S9A). T>G transitions were nominally more likely to be emVars in SCZ versus controls (permuted Fisher's exact test P = 0.03, OR = 2.6; fig. S9, B to E), but no mutation type was enriched after multiple hypothesis correction. Some emVars were located at TFBSs within DHSs near neurodevelopmental genes. For
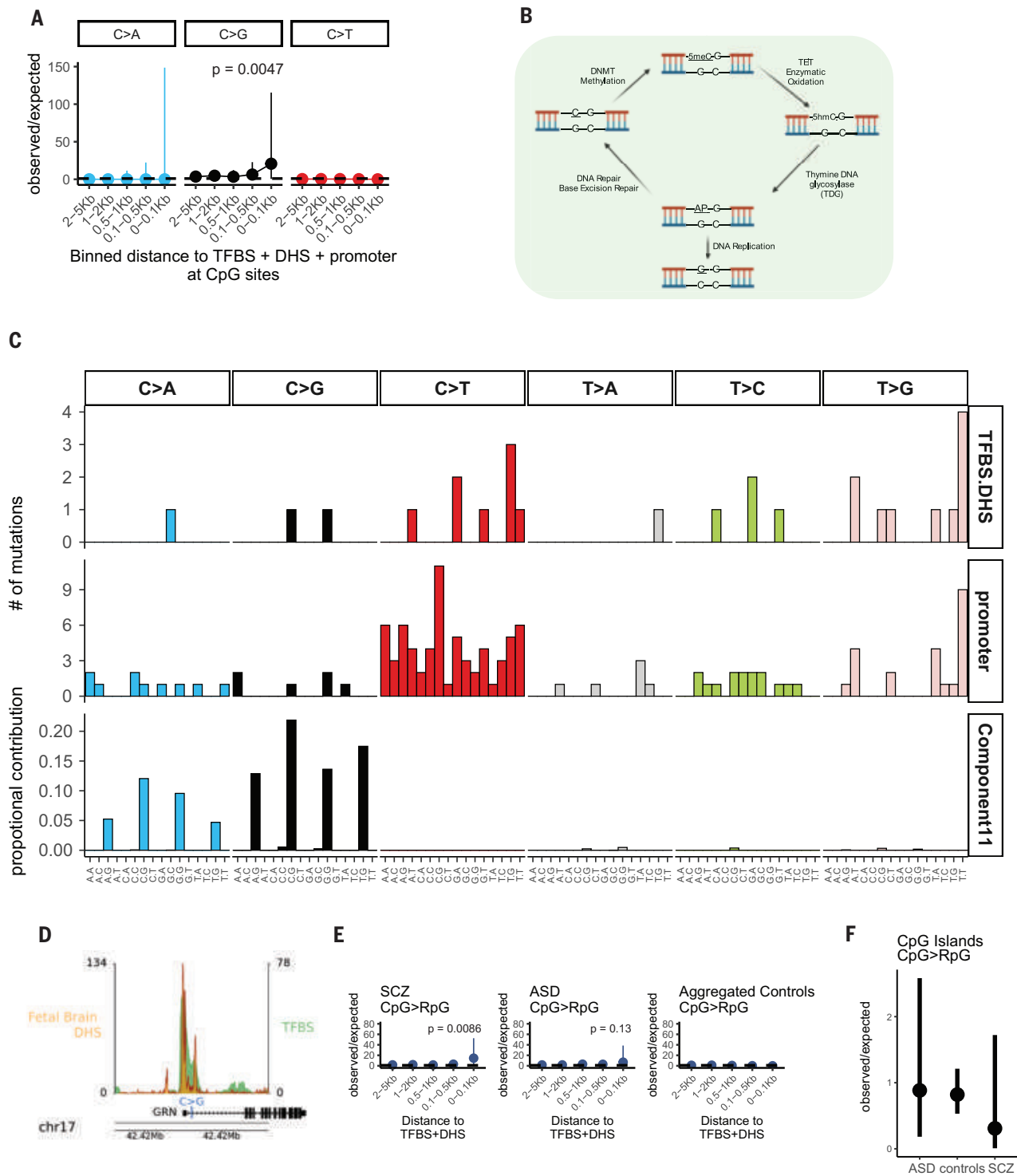
Fig. 3. Increased somatic CpG transversions at active transcription factor binding sites in SCZ. (**A**) Forest plots of rate ratios in SCZ of different base changes in active TFBSs at CpG sites. (**B**) Schematic of enzymatic demethylation mechanism resulting in CpG transversions. Abbreviations: 5meC, 5-methyl-cytosine; 5hmC, 5-hydroxymethyl-cytosine; AP abasic site. (**C**) Trinucleotide context plot of sSNVs in schizophrenia at active TFBSs and promoter sites, and CpG transversion signature component 11 (*24*).

(**D**) Illustration of promoter CpG>GpG mutation of *GRN* gene with DHS and TFBS tracks. (**E**) Forest plots of observed versus expected CpG transversions at active TFBSs in promoter regions from schizophrenia (SCZ), autism spectrum disorder (ASD), and aggregated controls. (**F**) Forest plot of the relative observed versus expected CpG transversions at CpG islands across diagnostic categories. For (A), (E), and (F), *P* values and 95% CIs were computed using a Poisson test.

Fig. 4. Increased somatic T>G substitutions at active TFBSs in SCZ and cancer samples. (**A**) Forest plots of rate ratios in SCZ of different base changes in active TFBSs at promoter regions at non-CpG sites. *P* values and 95% CIs were computed using a Poisson test. (**B**) List of T>G variants occurring at the same genomic position. (**C**) T>G sSNVs observed versus expected mutation rate at TFBSs across various cancer types. Samples on the *x* axis are sorted on the basis of observed/expected ratios for each cancer category. Pink data points indicate samples with enriched T>G burden at TFBSs. Triangles indicate samples with *XPD* mutation. (**D**) Observed over expected ratio of T>G sSNVs at TFBSs in cancer samples carrying *XPD* mutations, versus noncarriers. (**E**) Forest plot of sSNV rate in liver and bladder cancers stratified by enrichment of T>G mutations at TFBSs [pink data points from (C)]. (**F**) Ninety-six trinucleotide context of SCZ sSNVs at active TFBSs (TFBS+DHS) and at promoter regions, along with liver and bladder cancer sSNVs from samples with *XPD* dysfunction at active TFBSs. The corresponding tissue DHS track for each cancer type was obtained from the ENCODE database (table S5).
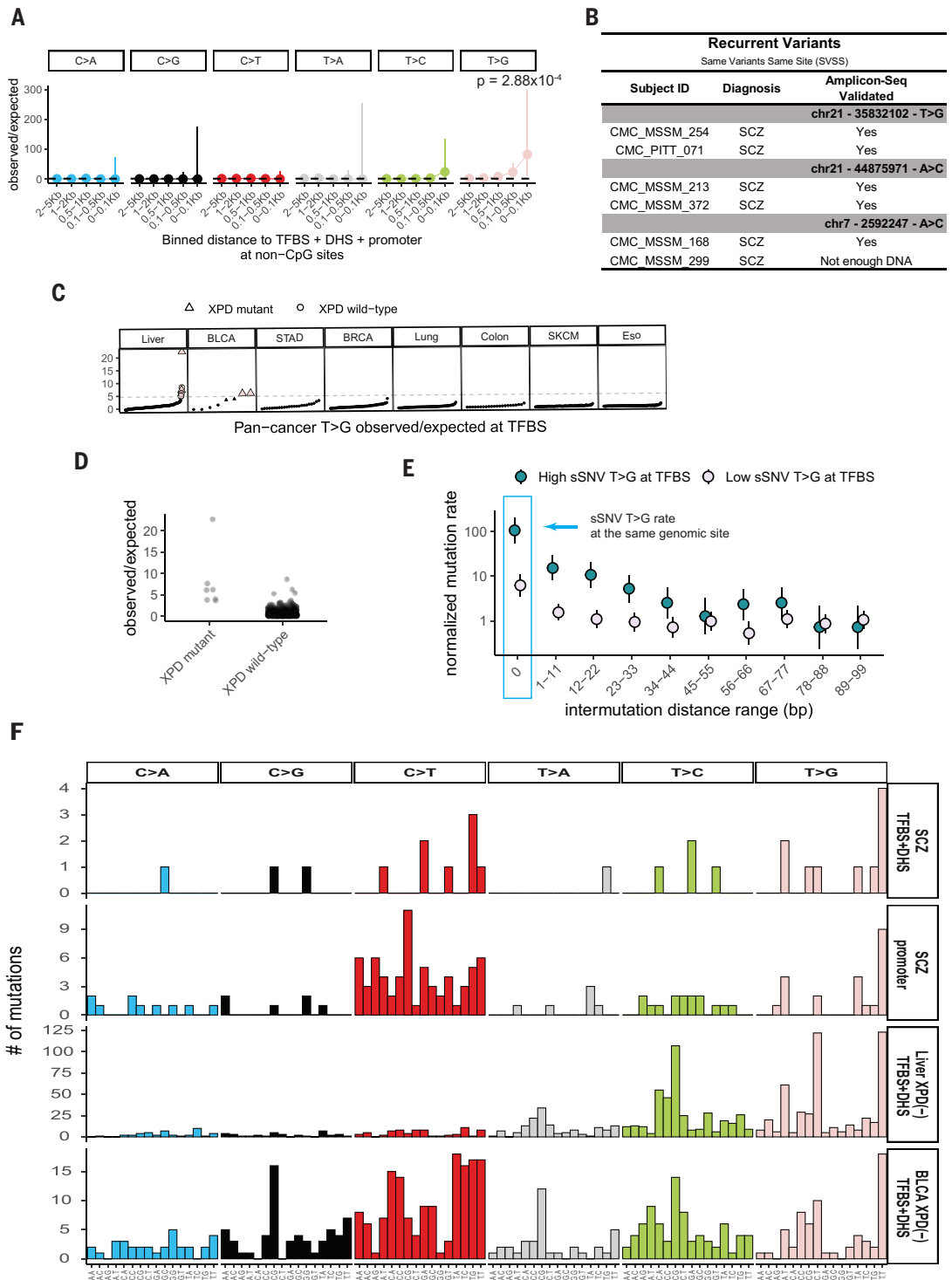
example, emVar chr19:13166346 T>G decreases regulatory activity [Benjamini-Hochberg (BH)–corrected Wald's test *P* < 0.0001, log2 fold change (log2FC) = −1.36] and is near *NFIX* (Fig. 5B and fig. S9D), in which heterozygous loss-of-function mutations cause Malan syndrome, which is characterized by brain overgrowth and behavioral abnormalities (*36*). Another emVar, chr19:11593076 A>C (BH-corrected

Wald's test *P*<0.0001, log2FC= −0.57) is near *ELAVL3*, a neuron-specific RNA-binding protein that regulates glutamate neurotransmission and neuronal excitability (*37*, *38*) (Fig. 5C and fig. S9D).

We predicted brain-specific genes that are targeted by regulatory elements harboring somatic emVars using gene-enhancer linkage maps (table S9 and materials and methods)

(*39*, *40*), linking 88 emVars to 247 candidate target genes, with some sSNVs linking to multiple genes (range: 2 to 13 targets). Two somatic emVars target seven genes overlapping SCZ risk loci (Fig. 5, D and E; materials and methods). These variants had the same direction of effect in all tested contexts in MPRAs and caused regulatory disruption across most windows (fig. S9D). In particular, emVar chr6:26533434
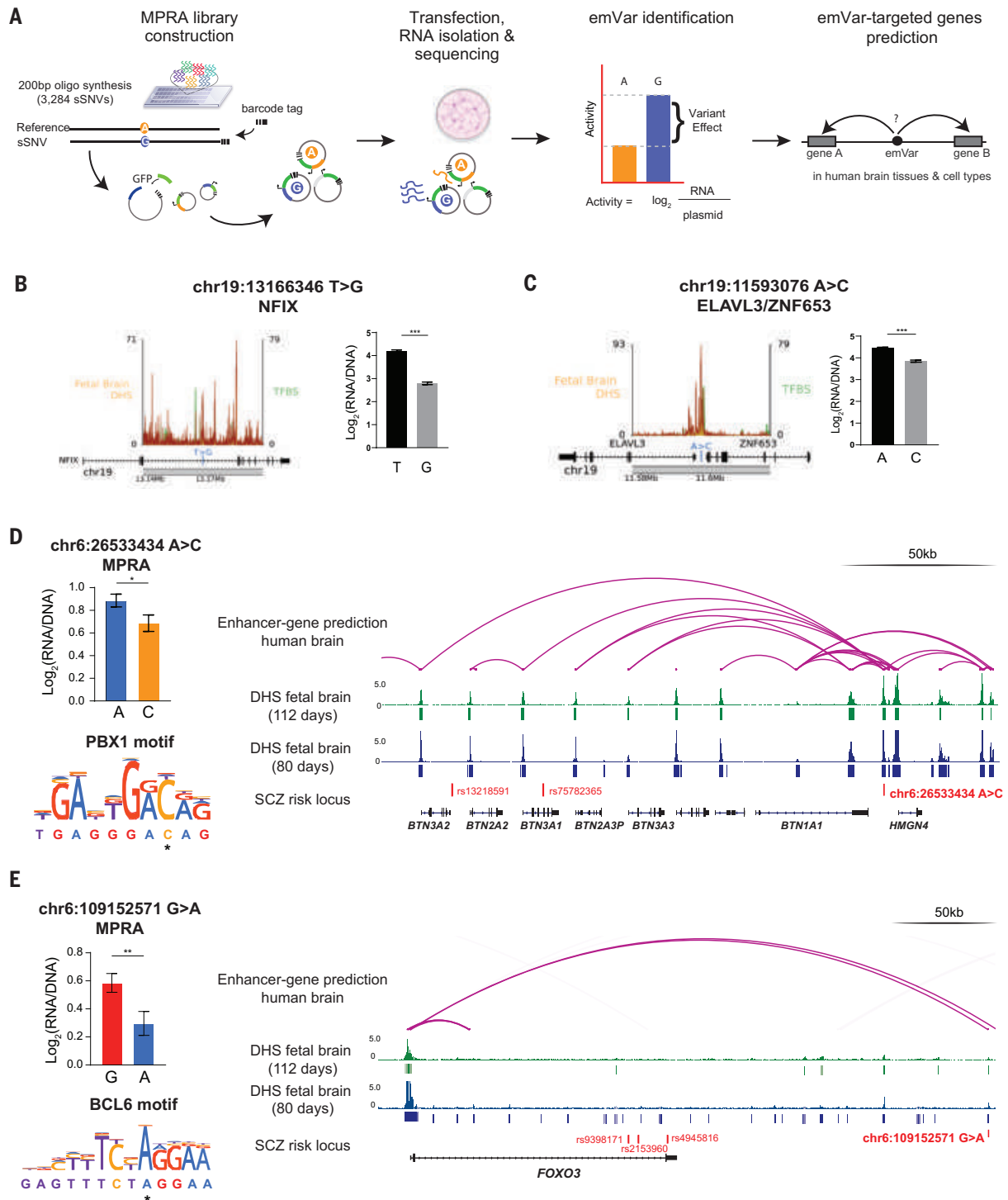
Fig. 5. Transcriptional impact of early developmental somatic variants in SCZ and control individuals. (**A**) Schematic of MPRA experimental design. (**B** and **C**) Schematic of T>G sSNVs occurring near developmental genes *NFIX* and *ELAVL3/ZNF63*, with DHS and TFBS tracks. MPRA bar plots represent expression amounts from each allele in MPRA. values represent BH−corrected Wald's test between the log ratios of the reference and alternative alleles.

(**D** and **E**) MPRA results, motif break prediction, and integrative genomic viewer of enhancer-gene linkage map for somatic-emVars targeting known SCZ risk genes. MPRA bar plots represent expression amounts from each allele in MPRA, and *P* values represent BH-corrected Wald's test between the log ratios of the reference and alternative alleles. DHS tracks for human fetal brain tissues at different stages are from the ENCODE portal.

A>C, a T>G that down-regulates activity (BH-corrected Wald's test $P < 0.05$, log2FC = −0.18), creates a predicted binding site for PBX1 ($P < 0.0001$, allele difference = 0.99), a repressive

regulator of neuron development (*41*) (Fig. 5D). The variant maps its gene-enhancer activity to the genes *BTN1A1*, *BTN2A3P*, *BTN3A1*, *BTN3A2*, *BTN3A3*, and *HMGN4* within the major histo-

compatibility complex class I region (Fig. 5D), a locus reproducibly associated with SCZ (*42*, *43*). The emVar chr6:109152571 G>A also decreases transcription (BH-corrected Wald's test $P < 0.001$,

log2FC = −0.27) and is predicted to loop to the *FOXO3* promoter (Fig. 5E and fig. S9D), which is associated with schizophrenia. The variant creates a predicted binding site for BCL6 (*P* < 0.0001, allele difference = 1.52), a direct repressor of *FOXO3* (*44*). Together, *BCL6* and *FOXO3* reciprocally regulate neural stem cell proliferation and differentiation (*45*).

Discussion

Although our data are limited by sample size, they suggest mutational models that could explain distinctive sSNV patterns in SCZ. CpG transversions make up ~2.4% of all mosaic mutations in brain tissue, potentially originating in the early zygote shortly after fertilization, when global DNA demethylation of the paternal and maternal genomes restores totipotency at the maternal-to-zygotic transition (*24, 25, 46*). Alterations in this process, either endogenous or exogenous, would predispose to somatic CpG transversions. The high VAF of CpG transversions at TFBSs (average VAF = 13%) is consistent with this very early occurrence. We speculate that the last step of demethylation could be obstructed by TF binding, analogous to interference between TF binding and DNA repair in cancer (*20–22*), where enrichment of sSNVs at active TFBSs has been attributed to steric interference of TFs with the repair apparatus. Comparison of mosaic mutations between SCZ and controls is notable because the overall burden in CpG transversions is higher than in the germ line for both cases and controls (*24*), but effects of TF binding are specific to neuropsychiatric disease.

Somatic T>G mutations may reflect an *XPD* dysfunction–like mechanism, as suggested by the similarity in mutational patterns at TFBSs between SCZ and cancers deficient in *XPD*. This mechanism could produce preferential sSNV accumulation at active TFBSs owing to hindrance of DNA repair by TFs bound to damaged DNA (*21*). However, although we did not find deleterious somatic or germline mutations in *XPD* in any SCZ samples, we cannot exclude altered nucleotide excision repair or *XPD* expression by other mechanisms. The root cause of T>G mutations, even in cancer, is unclear. They have been proposed to reflect oxidative damage to deoxyribonucleotides in rapidly dividing cells (*47, 48*), which could reflect stressors during development. For example, maternal infection and maternal immune activation (MIA) have been implicated in SCZ by epidemiology and animal models (*49*), but whether MIA causes somatic mutations, and if so in what pattern, are unknown.

The relationship between mutational processes observed here and SCZ might reflect several models. Developmental sSNVs may exert direct effects on disease liability analogous to germline mutations: Even though only some neurons harbor these variants, the affected population may be large enough to produce phenotypic manifestations. Alternatively, differences in detected somatic variants between SCZ and controls may reflect differences in clonal structures of progenitor populations, producing differential detection sensitivity, akin to focal cortical dysplasia (*4*). Lastly, various factors involved in SCZ etiology might be mutagenic independent of direct effects on SCZ, manifesting, for example, as enhanced CpG demethylation, inhibition of DNA repair, or decreased removal of mutated cells.

Our data show enrichment of somatic mutagenic processes that were previously characterized in other contexts, rather than enrichment of functional classes of mutations known to influence SCZ risk in the germ line (such as coding mutations). Some somatic variants at TFBSs can alter expression of neurodevelopmental genes, favoring models that somatic mutations increase disease liability. Somatic SNVs at TFBSs active in development are thus simultaneously products of mutagenesis hotspots and ideal candidates to create risk for developmental brain dysfunction, increasing the probability that variants disrupt transcriptional regulation crucial to neuronal function. They may synergize with germline SCZ risk alleles that typically control gene dosage (*1, 16, 17*). Finally, the highly recurrent sites affected suggest that nonspecific mutagenic processes can be channeled by specific TF binding to create recurrent patterns of mutation and potentially increased liability for behaviorally complex phenotypes.

REFERENCES AND NOTES

1. M. J. Owen, S. E. Legge, E. Rees, J. T. R. Walters, M. C. O'Donovan, *Mol. Psychiatry* **28**, 3638–3647 (2023).
2. T. Bae *et al.*, *Science* **359**, 550–555 (2018).
3. S. Bizzotto *et al.*, *Science* **371**, 1249–1253 (2021).
4. E. L. Heinzen, *Curr. Opin. Genet. Dev.* **65**, 1–7 (2020).
5. S. Khoshkhoo *et al.*, *JAMA Neurol.* **80**, 578–587 (2023).
6. Y. Dou *et al.*, *Hum. Mutat.* **38**, 1002–1013 (2017).
7. R. E. Rodin *et al.*, *Nat. Neurosci.* **24**, 176–185 (2021).
8. E. A. Maury *et al.*, *Cell Genomics* **3**, 100356 (2023).
9. M. Marin-Padilla, *J. Craniofac. Genet. Dev. Biol.* **10**, 137–146 (1990).
10. A. Matevossian, S. Akbarian, *J. Vis. Exp.* (20): 914 (2008).
11. Y. Wang *et al.*, *Genome Biol.* **22**, 92 (2021).
12. T. Bae *et al.*, *Science* **377**, 511–517 (2022).
13. L. B. Alexandrov *et al.*, *Nature* **578**, 94–101 (2020).
14. K. P. Lesch *et al.*, *J. Neural Transm.* **115**, 1573–1585 (2008).
15. H. M. Ollila *et al.*, *Mol. Psychiatry* **14**, 351–353 (2009).
16. D. Liu *et al.*, *Nat. Genet.* **55**, 369–376 (2023).
17. K. J. Karczewski *et al.*, *Nature* **581**, 434–443 (2020).
18. A. Kundaje *et al.*, *Nature* **518**, 317–330 (2015).
19. Y. Cai *et al.*, *Nat. Commun.* **12**, 719 (2021).
20. D. Perera *et al.*, *Nature* **532**, 259–263 (2016).
21. R. Sabarinathan, L. Mularoni, J. Deu-Pons, A. Gonzalez-Perez, N. López-Bigas, *Nature* **532**, 264–267 (2016).
22. R. Katainen *et al.*, *Nat. Genet.* **47**, 818–821 (2015).
23. I. E. Vorontsov *et al.*, *BMC Res. Notes* **11**, 756 (2018).
24. V. B. Seplyarskiy *et al.*, *Science* **373**, 1030–1035 (2021).
25. X. Wu, Y. Zhang, *Nat. Rev. Genet.* **18**, 517–534 (2017).
26. K. Chan, M. A. Resnick, D. A. Gordenin, *DNA Repair* **12**, 878–889 (2013).
27. M. E. Ward *et al.*, *Sci. Transl. Med.* **9**, eaah5642 (2017).
28. P. Momeni *et al.*, *Neurocase* **16**, 273–279 (2010).
29. H. Jonsson *et al.*, *Nat. Genet.* **53**, 27–34 (2021).
30. L. A. Aaltonen *et al.*, *Nature* **578**, 82–93 (2020).
31. J. A. Barbour *et al.*, Global and local redistribution of somatic mutations enable the prediction of functional XPD mutations in bladder cancer. BioRxiv 2022.01.21.477237 [Preprint] (2024); https://doi.org/10.1101/2022.01.21.477237.
32. Z. Li *et al.*, *Nat. Genet.* **49**, 1576–1583 (2017).
33. V. Trubetskoy *et al.*, *Nature* **604**, 502–508 (2022).
34. J. R. Xue *et al.*, *Science* **380**, eabn2253 (2023).
35. R. Tewhey *et al.*, *Cell* **172**, 1132–1134 (2018).
36. V. Malan *et al.*, *Am. J. Hum. Genet.* **87**, 189–198 (2010).
37. M. R. Mulligan, L. S. Bicknell, *Eur. J. Hum. Genet.* **31**, 1209–1217 (2023).
38. G. Ince-Dunn *et al.*, *Neuron* **75**, 1067–1080 (2012).
39. J. Nasser *et al.*, *Nature* **593**, 238–243 (2021).
40. Y. E. Li *et al.*, *Science* **382**, eadf7044 (2023).
41. O. Golonzhka *et al.*, *Neuron* **88**, 1192–1207 (2015).
42. S. M. Purcell *et al.*, *Nature* **460**, 748–752 (2009).
43. H. Stefansson *et al.*, *Nature* **460**, 744–747 (2009).
44. W. R. Wu *et al.*, *Theranostics* **10**, 707–724 (2020).
45. V. M. Renault *et al.*, *Cell Stem Cell* **5**, 527–539 (2009).
46. M. A. Eckersley-Maslin, C. Alda-Catalinas, W. Reik, *Nat. Rev. Mol. Cell Biol.* **19**, 436–450 (2018).
47. K. Satou, K. Kawai, H. Kasai, H. Harashima, H. Kamiya, *Free Radic. Biol. Med.* **42**, 1552–1560 (2007).
48. Y. A. Guo *et al.*, *Nat. Commun.* **9**, 1520 (2018).
49. M. L. Estes, A. K. McAllister, *Science* **353**, 772–777 (2016).

ORGANIC CHEMISTRY

Catalytic asymmetric fragmentation of cyclopropanes

Ravindra Krushnaji Raut[1], Satoshi Matsutani[2], Fuxing Shi[3], Shuta Kataoka[2], Margareta Poje[3], Benjamin Mitschke[3], Satoshi Maeda[1,4,5,6], Nobuya Tsuji[1]*, Benjamin List[1,3]*

The stereoselective activation of alkanes constitutes a long-standing and grand challenge for chemistry. Although metal-containing enzymes oxidize alkanes with remarkable ease and selectivity, chemical approaches have largely been limited to transition metal–based catalytic carbon–hydrogen functionalizations. Alkanes can be protonated to form pentacoordinated carbonium ions and fragmented into smaller hydrocarbons in the presence of strong Brønsted acids. However, catalytic stereocontrol over such reactions has not previously been accomplished. We show here that strong and confined acids catalyze highly enantioselective fragmentations of a variety of cyclopropanes into the corresponding alkenes, expanding the boundaries of catalytic selective alkane activation. Computational studies suggest the involvement of the long-debated cycloproponium ions.

Alkanes can still be considered as the primary feedstock of the global economy. However, until today, the catalytic and stereoselective activation of alkanes defines one of the most difficult problems of chemistry. Traditionally, the stereoselective activation of alkanes has been limited to C–H functionalizations. For example, natural and engineered enzymes activate alkanes to produce secondary alcohols, with high enantioselectivities (1). Current chemical strategies for alkane activation primarily revolve around the utilization of transition metal complexes, specifically metal-oxo (2), metal-carbenoid (3, 4), or combined metal-photoredox systems (5, 6), facilitating enantioselective C–H functionalizations (7). By contrast, the acidic principle of the technically used alkane-cracking zeolite catalyst is the proton, which initiates the fragmentation of C–C σ bonds of long-chain alkanes, providing smaller alkene and alkane hydrocarbons. Unlike conventional trivalent carbenium ions, protonated alkanes feature pentacoordinated carbonium ions, having three-center two-electron bonds as defined by Olah et al. (Fig. 1A) (8). Such a bonding scenario has also been involved in high-temperature, entropy-driven catalytic cracking (Fig. 1B). In general, zeolite catalysis is neither site selective nor stereoselective, even though it can exhibit selectivity for certain substrates and products

(9). However, introducing stereoselectivity, particularly enantioselectivity, remains unaddressed.

Furthermore, cyclopropanes possess distinctive orbital properties due to relatively large ring strains (27.5 kcal/mol compared with cyclohexane) (10, 11), as explained by Coulson and Moffitt's bent orbitals and Walsh's orbital model (12, 13), rendering them prone toward protonation by strong acids. Although the structure of the protonated cyclopropane had been discussed for a long time (14), computational studies suggest an unsymmetrically bridging and methyl-eclipsed structure, which also involves some degree of carbonium character (Fig. 1C) (15–18). Despite such academic interest in the protonated cyclopropanes (19), chemical approaches to control the stereochemistry of their fragmentation have been strictly limited to engineered substrates, which possess basic functional groups that can be activated to gain reactivities by increasing polarization of the cyclopropane σ bonds and to control enantioselectivity by forming well-defined substrate-catalyst complexes (20–28). By contrast, alkyl-substituted cyclopropanes have been posited as biochemical substrates to generate carbonium ions that undergo subsequent rearrangements or fragmentations as described by Djerassi and colleagues (29, 30). We envisaged to stereochemically control the reactions of such cyclopropanium ions within the well-defined catalytic active site of a confined chiral counteranion (31). Recently, our group developed imidodiphosphorimidates (IDPi) (32), which are highly confined and acidic organocatalysts, facilitating previously challenging asymmetric transformations including the asymmetric activation of simple alkenes (33–37). The extraordinary ability of this class of catalysts to control carbocationic intermediates (38) was also demonstrated with a nonclassical 2-norbornyl carbocation (39). Although nortricyclene can be used as its precursor, the scope was strictly limited to a single example: an addition

of 1,3,5-trimethoxybenzene to the 2-norbornyl cation. Therefore, a general stereoselective activation of nonstabilized cyclopropanes has never been achieved. In this study, we present the highly enantioselective fragmentation of fully saturated cyclopropanes by means of IDPi Brønsted acid catalysis, a process that can be described as asymmetric "cracking" of cyclopropanes, akin to petroleum chemistry (Fig. 1D).

Reaction development

Our initial study commenced with 3-substituted bicyclo[3,1,0]hexane **1a**, with the expectation that desymmetrization of the molecule upon asymmetric protonation followed by deprotonation and isomerization would provide alkenes (40). Conventional weaker Brønsted acid catalysts, such as phosphoric acids (41–43), imidodiphosphates (IDP) (44), and iminoimidodiphosphates (iIDP) (45), as expected, offered no to poor reactivity toward this aliphatic cyclopropane (fig. S1). By contrast, IDPi **7a** exhibited good reactivity, yielding a mixture of alkenes **2a** to **6a** (Fig. 2). After a meticulous analysis, the predominant components were identified as trisubstituted alkenes **2a** and **3a**, present in an almost racemic form, alongside minor amounts of **4a**, **5a**, and **6a**. We hypothesized that, in addition to conventional steric effects of the substituents, enhancing noncovalent interactions between the substrate and the catalyst anion is crucial for controlling the regioselectivity and enantioselectivity of such aliphatic molecules. Taking advantage of the tunability of IDPi catalysts, we performed a systematic screening of the catalyst. At this point, we hypothesized that the 3,3′-positions of 1,1′-Bi-2-naphthol (BINOL) substituents define the shape of the microenvironment, aided by London dispersion, and that the nitrogen substituents define the basicity of the corresponding anion, thereby controlling the electrostatic interactions of heteroatoms with substrates. Consistent with our expectations, IDPi catalyst **7b**, featuring BINOL substituents with aliphatic *tert*-butyl groups to induce steric hindrance and dispersion interactions, demonstrated promising enantioselectivities and regioselectivities toward cyclopentenes **2a** to **5a** over cyclohexenes. To further improve the enantioselectivity, we then replaced trifluoromethanesulfonyl groups of the catalytic active site with arylsulfonyl groups, anticipating an increase in basicity of the catalytic active site and enhanced conformer rigidity by indirect steric effects and π-π stacking with the BINOL backbone. Indeed, IDPi **7c** delivered the desired products with improved enantioselectivities, particularly for substituted cyclopentenes **2** to **5**. After additional optimization of the catalyst and conditions, IDPi **7d** was found to produce the desired alkenes in a satisfactory combined yield with excellent enantioselectivities, subsequent to further optimization

[1]Institute for Chemical Reaction Design and Discovery (WPI-ICReDD), Hokkaido University, Sapporo 001-0021, Japan. [2]Graduate School of Chemical Sciences and Engineering, Hokkaido University, Sapporo, Hokkaido 060-8628, Japan. [3]Max-Planck-Institut für Kohlenforschung, D-45470 Mülheim an der Ruhr, Germany. [4]Department of Chemistry, Graduate School of Science, Hokkaido University, Sapporo, Hokkaido 060-0810, Japan. [5]ERATO Maeda Artificial Intelligence for Chemical Reaction Design and Discovery Project, Hokkaido University, Sapporo, Hokkaido 060-0810, Japan. [6]Research and Services Division of Materials Data and Integrated System (MaDIS), National Institute for Materials Science (NIMS), Tsukuba, Ibaraki 305-0044, Japan.
*Corresponding authors. Email: tsuji@icredd.hokudai.ac.jp (N.T.); list@mpi-muelheim.mpg.de (B.L.)
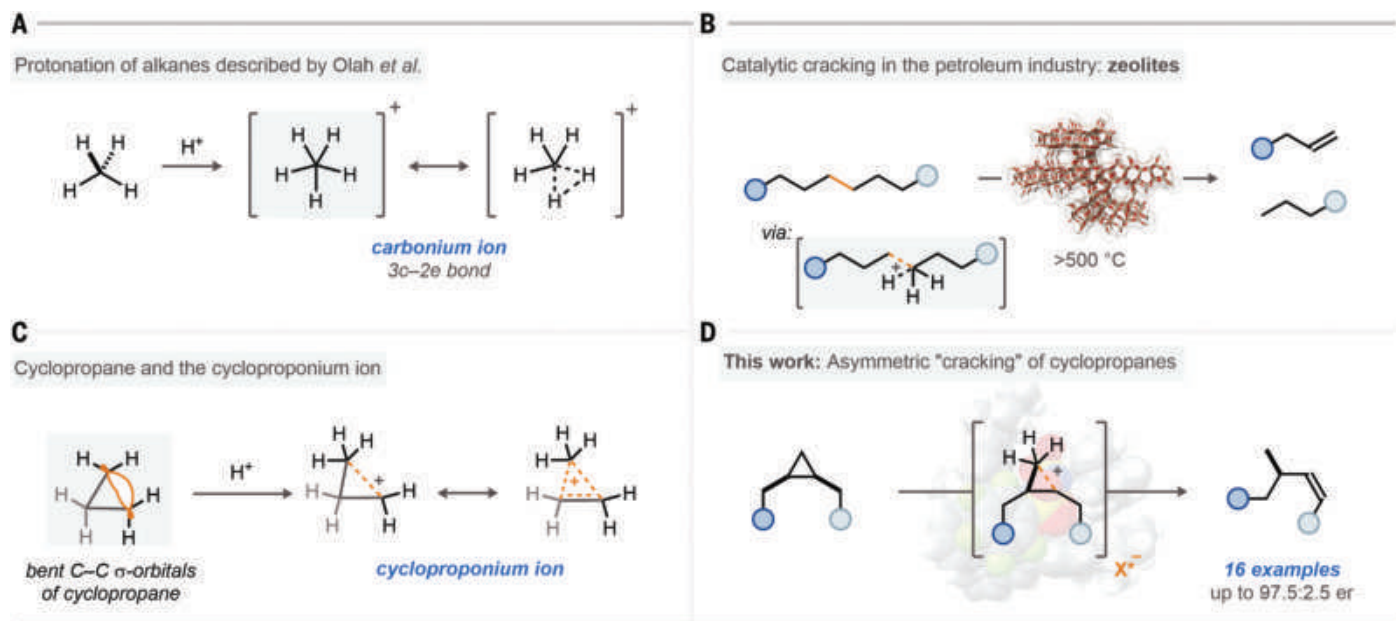
Fig. 1. Background. (**A**) Protonation of alkanes by carbonium ion (8). (**B**) Catalytic cracking process in petrol chemistry. (**C**) Physical properties of cyclopropane and proposed structures of protonated cyclopropanes. (**D**) Organocatalytic asymmetric "cracking" of cyclopropanes.
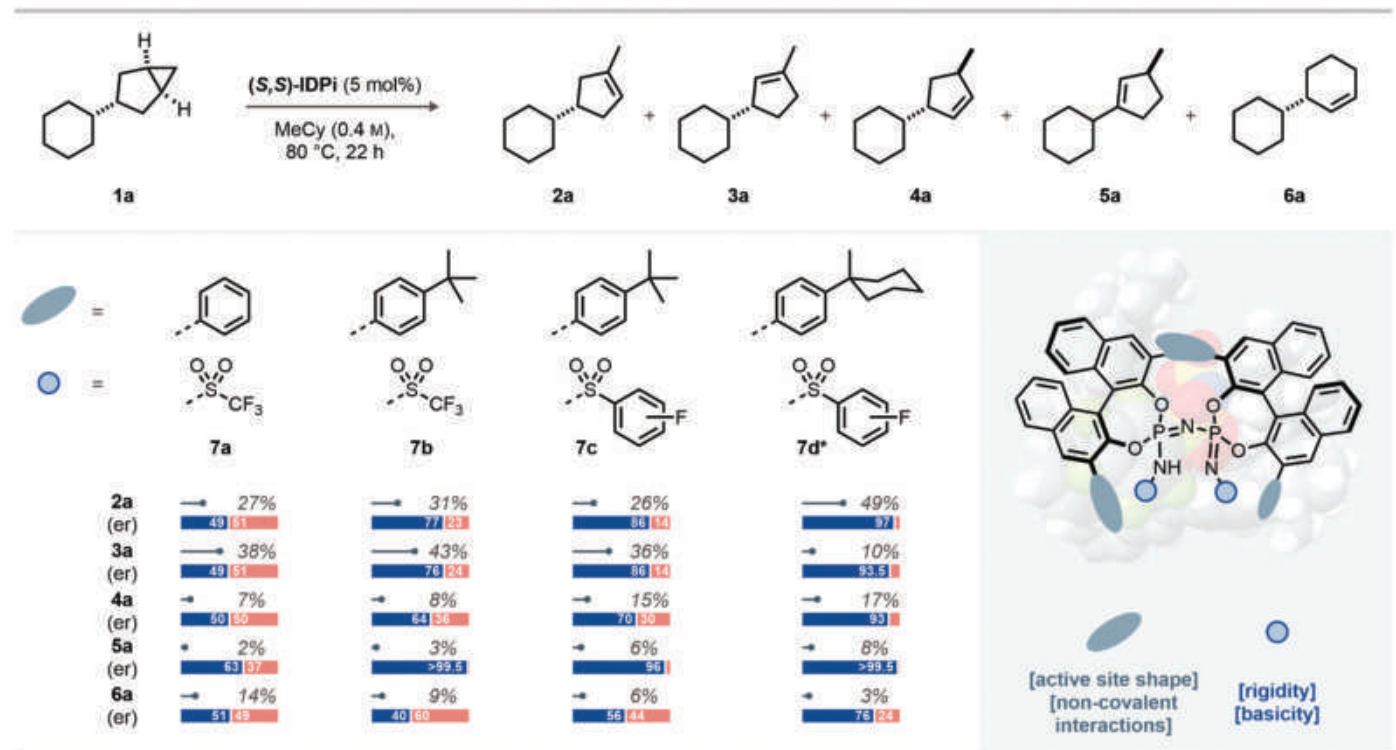
Fig. 2. Summary of the optimization of catalysts and reaction conditions. Yields and conversions were determined on the basis of ^{1}H NMR, using mesitylene as an internal standard. Enantiomeric ratio was determined by gas chromatography. *Reaction was performed in 1.0 M solution at 15℃ for 7 days. MeCy, methylcyclohexane.

of the reaction temperature and the reaction time. The major product, assigned as cyclopentene **2a**, was obtained in 40% yield with a 97:3 enantiomeric ratio. Additionally, three minor alkene isomers **3a**, **4a**, and **5a**, were obtained in 6%, 13%, and 7% yield, respectively, with

excellent enantioselectivities. The enantioselectivity of **5a** was nearly perfect, with only a single enantiomer observed by gas chromatography (GC). A minor amount of cyclohexene **6a** was also detected with moderate enantioselectivity.

Substrate scope exploration

With the optimized conditions in hand, the substrate scope was investigated (Fig. 3). In addition to cyclic (**2a** and **2b**) and acyclic (**2c**) secondary alkyl substituents, a sterically congested tertiary substituent was tolerated,
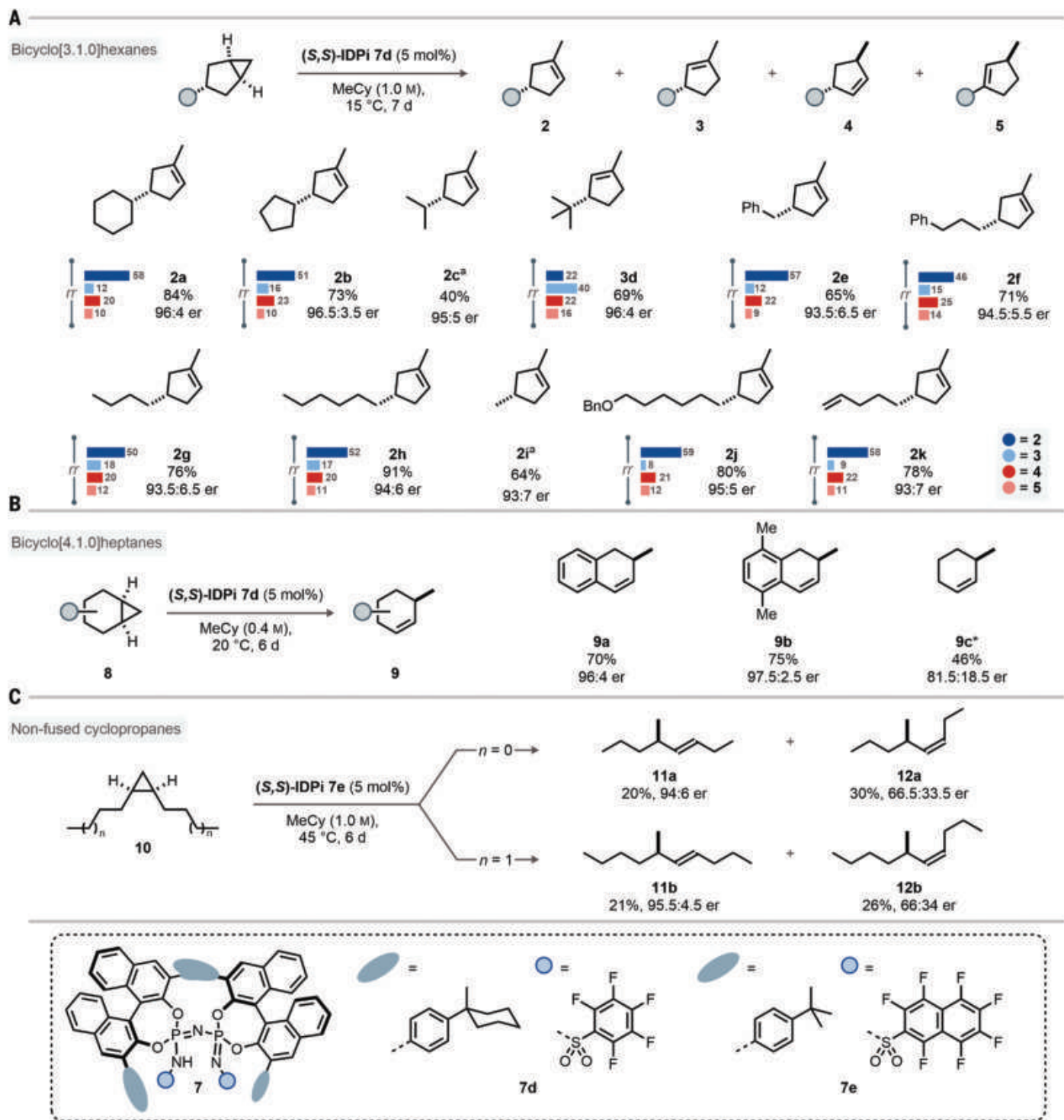
Fig. 3. Substrate scope. Yields and conversions were determined on the basis of gas chromatography (GC), using an internal standard. Enantiomeric ratio was also determined by GC. (**A**) Scope for bicyclo[3.1.0]hexanes. The major regioisomers are depicted. Unless otherwise noted, combined yields of **2**, **3**, **4**, and **5** and enantiomeric ratios based on the product distribution are presented (see supplementary materials for details). The ratio between **2:3:4:5** is presented as regioisomeric ratio (rr). [a]Reaction was done in *n*-pentane (0.4 M) instead of methylcyclohexane (MeCy), at room temperature. Due to the analytical problem of other isomers, yield and enantioselectivity of **2** was presented. (**B**) Scope for bicyclo[4.1.0]heptanes. *The reaction was performed in MeCy (1.0 M) at 25°C for 5 days. (**C**) Scope for nonfused cyclopropanes. er, enantiomeric ratio.

affording products **3d** in excellent yield and enantioselectivity. A series of primary alkyl-substituted substrates, such as benzyl (**2e**), 3-phenylpropyl (**2f**), *n*-butyl (**2 g**), and *n*-hexyl

(**2h**) groups, were compatible without erosion of enantioselectivities. Oxidative cleavage of alkene **3e** afforded the corresponding literature-known aldehyde, suggesting the abso-

lute configuration of the product as (*R*)-**3e** (fig. S9) (*46*). To our surprise, a substrate with an even smaller methyl substituent afforded the corresponding product **2i**, in 64% yield
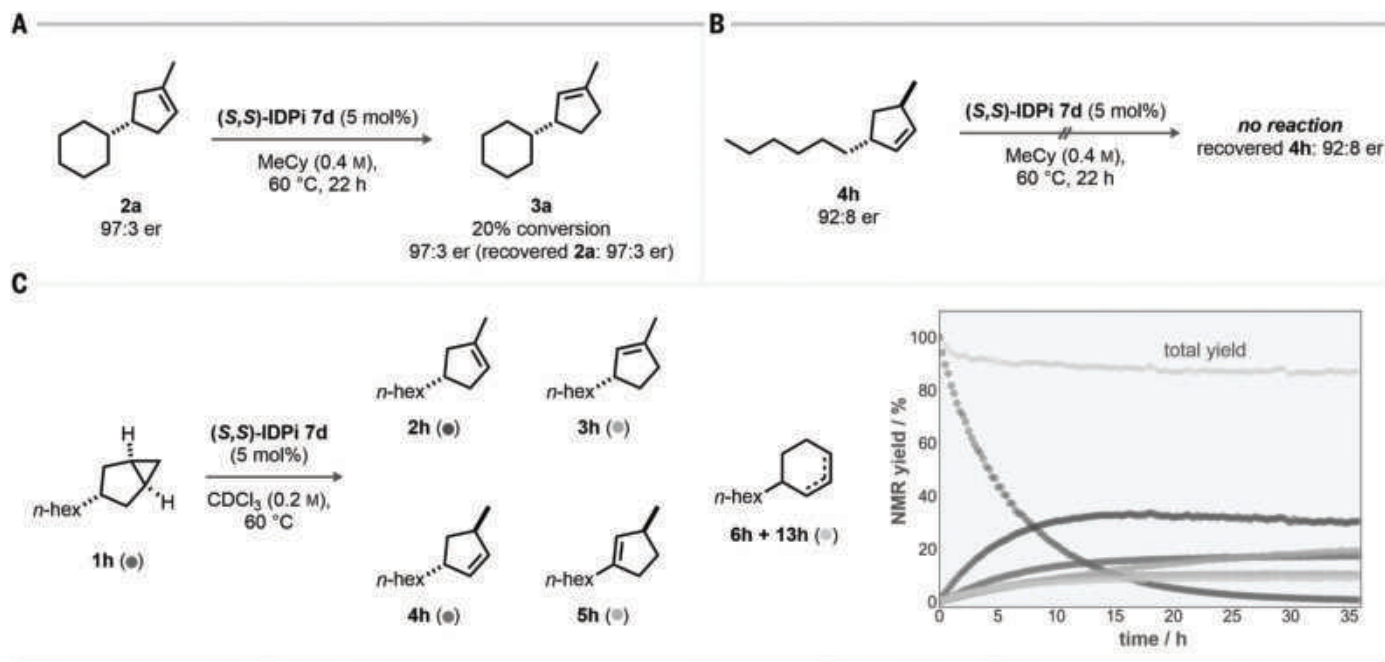
Fig. 4. Mechanistic studies. (**A**) Isomerization study of **2a** under the reaction conditions at a higher temperature. (**B**) Isomerization study of **4h** under the reaction conditions at a higher temperature. (**C**) Reaction profile monitored by ^{1}H NMR, using **1h** as a substrate. er, enantiomeric ratio.

and an enantiomeric ratio of 93:7, demonstrating a superior ability of IDPi catalyst to control stereochemistry of such small and simple alkanes. Substrates with a benzyloxy group (**2j**) and terminal alkene (**2k**) also furnished the desired products in excellent yields and enantioselectivities, demonstrating the functional group compatibility of the described method. The obtained mixture of regioisomers was further converted into functionalized compounds by hydrogenation, the Ritter reaction, the Friedel–Crafts reaction, and methylation by cyclopropanation followed by hydrogenolysis. These products were obtained without erosion of enantioselectivities, albeit with moderate diastereoselectivities (fig. S6). Moreover, the developed method can also be applied to bicyclo[4,1,0]heptanes **8**. Presumably owing to the additional stabilization of corresponding benzylic carbocations, substituted benzocyclohexenes **9a** and **9b** were obtained in good yield and in excellent enantioselectivities. Even very simple bicyclo[4,1,0]heptanes underwent the desired transformation to provide the corresponding cyclohexene **9c** in reasonable yield and enantioselectivity. Encouraged by these results, we investigated the reactivity of nonfused cyclopropanes **10**, which we expected to be even more challenging because of their highly flexible nature. Indeed, IDPi catalyst **7e** was able to activate 1,2-bis-*n*-propyl cyclopropane (**10a**) at an elevated temperature, affording a mixture of chiral disubstituted olefins (**11a**, **12a**). Whereas (*Z*)-alkene **12a** was obtained in only moderate enantioselectivity, the corresponding (*E*)-alkene **11a** was

obtained in excellent enantioselectivity, albeit with the yield remaining moderate. 1,2-Bis-*n*-butyl cyclopropane (**10b**) was also compatible and gave increased enantioselectivities. The absolute configuration was determined as indicated by comparing **11b** and **12b** with an enantiomerically enriched authentic sample (*47*).

Experimental and computational mechanistic studies

We next turned our attention toward the reaction mechanism. A main question became whether the enantioselectivity was governed by the designed asymmetric protonation process or in unintended kinetic resolutions driving alkene isomerization of the products. When trisubstituted alkene **2a** was treated with catalyst **7d** at 60°C, 20% isomerization to **3a** was observed, yet the enantiomeric ratio of both **2a** and **3a** remained unchanged (Fig. 4A). Because the enantiomeric ratio of **3a** obtained from the reaction is lower than that of **2a**, the isomerization of **2a** is less likely to be the major pathway leading to **3a** in the reaction mixture. Furthermore, disubstituted olefin **4h** remained completely unreactive in the presence of catalyst **7d** even at elevated temperature, suggesting that neither **2h** nor **5h** were obtained through isomerization of **4h** under the reaction conditions (Fig. 4B). These results were further supported by a kinetic study using proton nuclear magnetic resonance (^{1}H NMR) spectroscopy with an external standard. The products **4h**, **5h**, **6h**, and **13h** reached a plateau after one day at 60°C in CDCl$_3$, and only olefin **3h** kept increasing with consuming **2h** (Fig. 4C).

To investigate the detailed reaction mechanisms and origin of the enantioselectivities, density functional theory (DFT) studies of the reaction pathway were performed using catalyst **7d** and substrate **1a**. The calculated energy profile for the generation of product **2a** at CPCM(cyclohexane)-ωB97M-V/(ma)-def2-TZVPP//r^{2}SCAN-3c level of theory (*48–50*) is depicted (Fig. 5A). Initially, the substrate embeds itself inside the IDPi catalyst cleft, resulting in the formation of complex **I**. After forming this complex, the substrate is then protonated by the IDPi catalyst through four possible pathways; a protonation through **TS1A** results in carbonium intermediate **II**, whereas the protonation through **TS1B** generates another nonclassical carbocation **III**, and the protonation on the other side through **TS1A'** and **TS1B'** would furnish **II'** and **III'**, leading to the minor enantiomer of the product **2a** (Fig. 5B). The carbonium characters of **II** and **III** are evident from the bond lengths and angles visualized in Fig. 5A. Further analysis by natural population analysis and Wiberg bond index also agreed with this observation, and it seems that these carbonium ions are stabilized in the microenvironment offered by the catalyst (figs. S17 and S18). The intermediates **II** and **II'** could undergo a facile 1,3-H-shift to furnish **III** and **III'**, respectively. These calculated energy barriers suggest that the reaction pathway could be different for each enantiomer; whereas the major enantiomer would preferentially be protonated on the more substituted carbon through **TS1A** followed by 1,3-H-shift, the minor enantiomer favors the protonation on
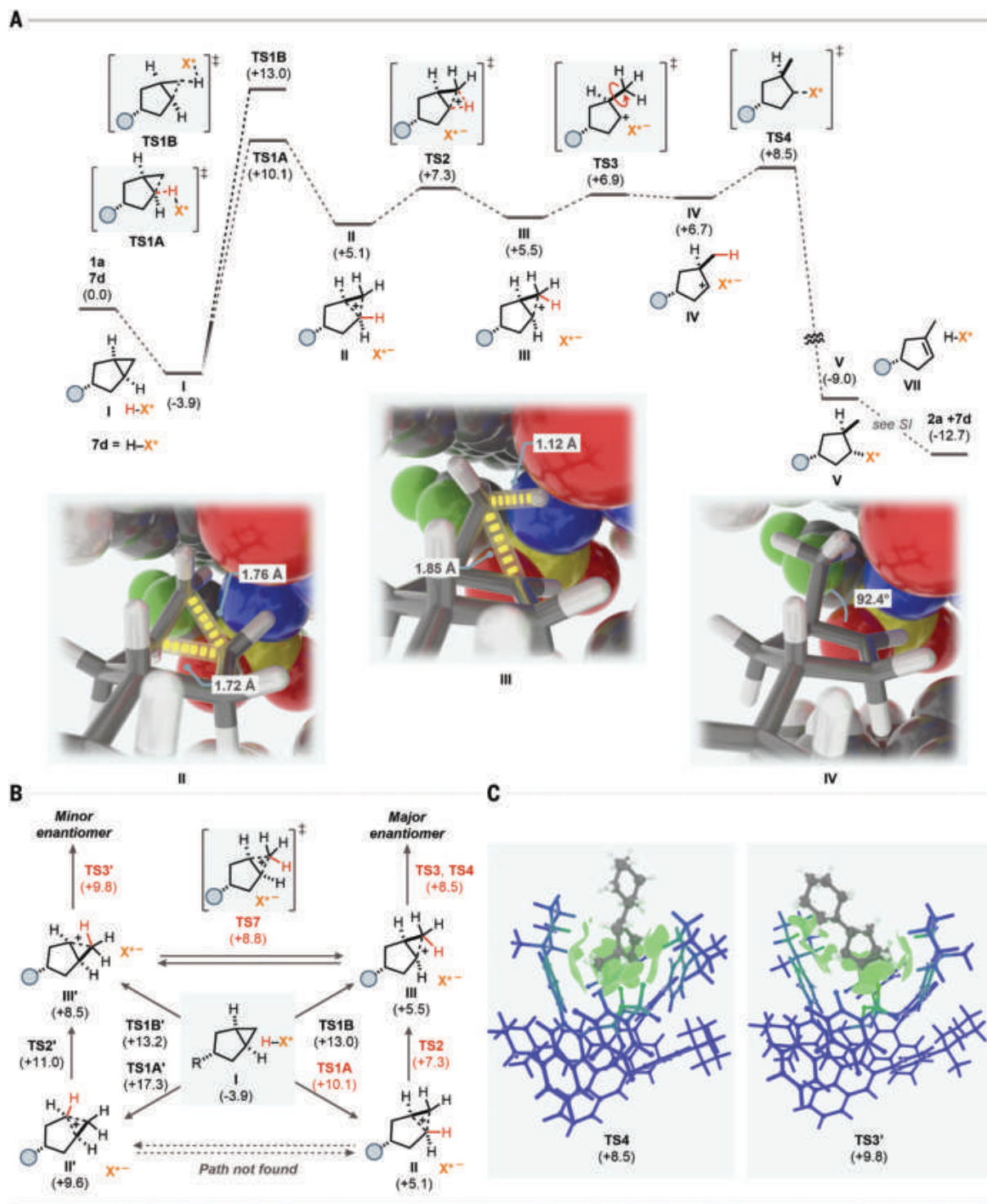
Fig. 5. Computational studies. (**A**) Energy diagram of the reaction profile leading to the major enantiomer **2a**. IDPi **7d** is shown as H-X* for simplicity. (**B**) Enantiodetermining steps in formation of **2a**. The important paths are highlighted in red. (**C**) Comparison of IGMH map between **TS4** and **TS3'**. A green isosurface represents the dispersion interactions. Atoms in brighter color contribute to intermolecular interactions.

the less-substituted carbon through **TS1B'**, although the barrier is still high. Whereas interconversions between **II** and **II'** through 1,2-H-shift are less likely, the racemization pathway between **III** and **III'** through **TS7** was found to be competitive with the major

pathway because it is lower than either **TS1A'** or **TS1B'**. Upon the formation of intermediate **III**, a rotation of the methyl group through **TS3** induces the exposure of the secondary carbocation **IV**, followed by the ring flip followed by a spontaneous formation of an elusive co-

valent bond adduct **V** with the IDPi catalyst. Similarly, **III'**, presumably mainly obtained through **TS7**, also undergoes the rotation of the methyl group, and the subsequently adduct **V'** is formed directly through **TS3'**, leading to the minor enantiomer of the product. **V** would

then convert into product **2a** and regenerate the IDPi catalyst **7d** (fig. S14). Overall, the energy difference between the **TS4** and **TS3′**, corresponding to the primal enantiodetermining step, is 1.3 kcal/mol, which is in good agreement with experimental outcome (1.7 kcal/mol at 298.15 K). Other isomers could be likely obtained through these intermediates by rearrangement and deprotonation occurring in the microenvironment as well. Consequently, there would be further enrichment or reduction in enantioselectivity depending on the isomers.

After thoroughly examining the reaction mechanism, we directed our attention to the key noncovalent interactions that drive the discrimination between enantiomers, which enabled us to unravel the origin of enantioselectivity. To visualize the noncovalent interactions, we used the independent gradient model based on the Hirshfeld partition (IGMH) method (51) at the ωB97X-D/def2-TZVPP level of theory, using the Multiwfn program (Fig. 5C) (52). Each transition state structure was divided into two fragments, a catalyst anion and a substrate cation, with solid surfaces representing an interaction between two fragments. It is evident that the aliphatic substrate fits well into the microenvironment of **7d** in **TS4**, with the substituents pointing outside of the pocket, whereas the pocket is slightly expanded by the substrate cation in **TS3′** (fig. S19). The interaction regions were generally composed of green isosurface, suggesting strong van der Waals interactions between the substrate and the BINOL substituents, with electrostatic interactions induced by the heteroatoms in the catalytic active site. Additionally, the green color in the counteranion highlights the atoms having a major contribution to the intermolecular interaction, indicating that the heteroatoms in the catalytic active site as well as the aliphatic substituents expectedly stabilize the substrate through noncovalent interactions. The protonation steps **TS1A**, **TS1B**, **TS1A′**, and **TS1B′** were also investigated, and it appears that both distortion and interaction play important roles in controlling the selectivity in this step as well (figs. S20 and S21).

We have developed a highly enantioselective fragmentation of cyclopropanes using chiral Brønsted acid catalysis with well-defined active sites. A series of saturated cyclopropanes were protonated to afford the corresponding alkenes with excellent enantioselectivities. We believe the approach delineated in this study pushes the boundary of stereoselectivities achieved by artificial molecules and could further expand the utilization of Brønsted acid catalysis in hydrocarbon chemistry.

REFERENCES AND NOTES

1. M. W. Peters, P. Meinhold, A. Glieder, F. H. Arnold, *J. Am. Chem. Soc.* **125**, 13442–13450 (2003).
2. M. Milan, M. Bietti, M. Costas, *ACS Cent. Sci.* **3**, 196–204 (2017).
3. K. Liao, S. Negretti, D. G. Musaev, J. Bacsa, H. M. Davies, *Nature* **533**, 230–234 (2016).
4. J. Fu, Z. Ren, J. Bacsa, D. G. Musaev, H. M. L. Davies, *Nature* **564**, 395–399 (2018).
5. Z. Y. Dai, Z. S. Nong, S. Song, P.-S. Wang, *Org. Lett.* **23**, 3157–3161 (2021).
6. Z.-Y. Dai, Z.-S. Nong, P.-S. Wang, *ACS Catal.* **10**, 4786–4790 (2020).
7. X. Tang, X. Jia, Z. Huang, *Chem. Sci.* **9**, 288–299 (2017).
8. G. A. Olah, G. K. S. Prakash, K. Wade, Á. Molnár, R. E. Williams, *Hypercarbon Chemistry* (Wiley, ed. 2, 2016).
9. M. Stöcker, *Microporous Mesoporous Mater.* **82**, 257–292 (2005).
10. K. B. Wiberg, *Angew. Chem. Int. Ed.* **25**, 312–322 (1986).
11. T. Dudev, C. Lim, *J. Am. Chem. Soc.* **120**, 4450–4458 (1998).
12. C. A. Coulson, W. E. Moffitt, *J. Chem. Phys.* **15**, 151 (1947).
13. A. D. Walsh, *Trans. Faraday Soc.* **45**, 179–190 (1949).
14. C. J. Collins, *Chem. Rev.* **69**, 543–550 (1968).
15. D. J. Sandbeck, D. J. Markewich, A. L. East, *J. Org. Chem.* **81**, 1410–1415 (2016).
16. J. A. Pople, L. Radom, V. Buss, P. R. Schleyer, *J. Am. Chem. Soc.* **93**, 1813–1815 (1971).
17. G. A. Olah, *J. Am. Chem. Soc.* **94**, 808–820 (1972).
18. G. A. Olah, G. K. Surya Prakash, A. Molnár, J. Sommer, *Superacid Chemistry* (Wiley, ed. 2, 2009).
19. P. S. Grant *et al.*, *J. Am. Chem. Soc.* **145**, 5855–5863 (2023).
20. S. C. Bart, P. J. Chirik, *J. Am. Chem. Soc.* **125**, 886–887 (2003).
21. T. Matsuda, T. Tsuboi, M. Murakami, *J. Am. Chem. Soc.* **129**, 12596–12597 (2007).
22. M. Murakami, N. Ishida, *J. Am. Chem. Soc.* **138**, 13759–13769 (2016).
23. B. Wang, M. A. Perea, R. Sarpong, *Angew. Chem. Int. Ed.* **59**, 18898–18919 (2020).
24. S. L. Lin, Y. H. Chen, H. H. Liu, S. H. Xiang, B. Tan, *J. Am. Chem. Soc.* **145**, 21152–21158 (2023).
25. L. Souillart, N. Cramer, *Chem. Rev.* **115**, 9410–9464 (2015).
26. T. F. Schneider, J. Kaschel, D. B. Werz, *Angew. Chem. Int. Ed.* **53**, 5504–5523 (2014).
27. Y. Cohen, A. Cohen, I. Marek, *Chem. Rev.* **121**, 140–161 (2021).
28. V. Pirenne, B. Muriel, J. Waser, *Chem. Rev.* **121**, 227–263 (2021).
29. J. R. Proudfoot, C. Djerassi, *J. Chem. Soc., Perkin Trans. 1* 1283–1290 (1987).
30. J. L. Giner, C. Djerassi, *Steroids* **57**, 258–261 (1992).
31. T. Li, K. D. Janda, R. A. Lerner, *Nature* **379**, 326–327 (1996).
32. P. S. Kaib, L. Schreyer, S. Lee, R. Properzi, B. List, *Angew. Chem. Int. Ed.* **55**, 13200–13203 (2016).
33. N. Tsuji *et al.*, *Science* **359**, 1501–1505 (2018).
34. P. Zhang, N. Tsuji, J. Ouyang, B. List, *J. Am. Chem. Soc.* **143**, 675–680 (2021).
35. R. Maji *et al.*, *J. Am. Chem. Soc.* **145**, 8788–8793 (2023).
36. L. Schreyer, R. Properzi, B. List, *Angew. Chem. Int. Ed.* **58**, 12761–12777 (2019).
37. V. N. Wakchaure *et al.*, *Nature* **625**, 287–292 (2024).
38. S. K. Nistanaki *et al.*, *Science* **378**, 1085–1091 (2022).
39. R. Properzi *et al.*, *Nat. Chem.* **12**, 1174–1179 (2020).
40. G. Dickmeiss *et al.*, *Angew. Chem. Int. Ed.* **48**, 6650–6653 (2009).
41. T. Akiyama, J. Itoh, K. Yokota, K. Fuchibe, *Angew. Chem. Int. Ed.* **43**, 1566–1568 (2004).
42. D. Uraguchi, M. Terada, *J. Am. Chem. Soc.* **126**, 5356–5357 (2004).
43. D. Parmar, E. Sugiono, S. Raja, M. Rueping, *Chem. Rev.* **114**, 9047–9153 (2014).
44. I. Ćorić, B. List, *Nature* **483**, 315–319 (2012).
45. L. Liu, P. S. Kaib, A. Tap, B. List, *J. Am. Chem. Soc.* **138**, 10822–10825 (2016).
46. T. I. Houjeiry, S. L. Poe, D. T. McQuade, *Org. Lett.* **14**, 4394–4397 (2012).
47. A. Sudau, W. Münch, J.-W. Bats, U. Nubbemeyer, *Eur. J. Org. Chem.* **2002**, 3315–3325 (2002).
48. M. Bursch, J. M. Mewes, A. Hansen, S. Grimme, *Angew. Chem. Int. Ed.* **61**, e202205735 (2022).
49. S. Grimme, A. Hansen, S. Ehlert, J. M. Mewes, *J. Chem. Phys.* **154**, 064103 (2021).
50. N. Mardirossian, M. Head-Gordon, *J. Chem. Phys.* **144**, 214110 (2016).
51. T. Lu, Q. Chen, *J. Comput. Chem.* **43**, 539–555 (2022).
52. T. Lu, F. Chen, *J. Comput. Chem.* **33**, 580–592 (2012).

Camera-Based 2D Rack Reader

The DP5 Mirage Rack Reader from Azenta Life Sciences delivers a cost-effective solution for reading SBS format racks in a compact low-height design. With an easy, out-of-the-box set up delivered pre-calibrated and ready to read, the DP5 Mirage is the perfect workhorse rack reader for busy life science laboratories. The small footprint of DP5 Mirage is highly economical as it relates to your valuable bench space, and it can be readily integrated into almost any laboratory workflow protocol. Using fast and reliable cutting-edge communications and control protocols, DP5 Decoding Software simplifies the operation of camera-based readers for users, while simultaneously offering greater networking ability and enhanced security features for more demanding applications in two add-on packages.

AZENTA Life Sciences
For info: +44 -161-777-2098
www.azenta.com

Bispecific Antibody Prototyping and Screening

Announces the expansion of its Pioneer™ Antibody Discovery Platform services with the addition of fast generation and screening of bispecific antibodies, based on the company's proprietary SpyLock Technology. The SpyLock service builds upon the capabilities of the Pioneer Platform, enabling customers to rapidly generate and screen bispecific antibodies to identify effective antibody combinations to take forward into development. The Pioneer Platform comprises a fully human synthetic phage display library of 225 billion unique Fab antibodies, and its proprietary SpyDisplay technology enables the rapid delivery of high-quality antibody candidates. The platform is paired with a high quality and comprehensive suite of discovery services that can be customized to meet the needs of any project. The Pioneer Platform is being harnessed by organizations globally, across applications including immuno-oncology and infectious diseases.

Bio-Rad Laboratories, Inc.
For info: +1 510-356-7909
www.bio-rad-antibodies.com/Spylock

Gravity-based Pump

Meticulously engineered to ensure delicate handling of sensitive media, VentriLabs´ gravity-based pump technology reduces destructive mechanical and shearing effects by up to 95% while pumping blood, as compared to commonly used peristaltic pumps. This innovative technology eliminates pulsation effects inherent in peristaltic and membrane pumps. In filtration applications this equates to improved efficiency and yields that translate into a more cost-efficient process for the user.

Biotech Fluidics AB
For info: +46-300-56-9180
Web   www.biotechfluidics.com

Tube-in-plate Reactor for Flow Chemistry

For synthetic chemistry experiments requiring efficient heat transfer - Uniqsis has introduced a novel Tube-in-plate (TIP) reactor that provides a cost-effective alternative to glass static mixer (GSM) reactor blocks. For highly exothermic or fast, mixing-dependent reactions. GSM reactor blocks produce an efficient, turbulent mixed reagent stream that is not diffusion dependent. However, while the narrow internal channels / large surface area of GSM reactor blocks provides very efficient heat transfer and mixing they can be prone to blockage. To address these issues, in applications where static mixing is not a concern, Uniqsis has developed a new TIP reactor using narrow bore, perfluoro polymer tubing embedded in channels cut into an aluminium block. To help stabilize the internal temperature of the TIP reactor, but retain visibility, the tubing is covered with a glass plate. Beneficially the glass coverplate and tubing can be easily removed, and the tubing replaced—should it become blocked or contaminated. With an internal volume of 9.5ml, a single layer large format TIP reactor can be used for both photochemistry and conventional thermally mediated reactions. For reactions where static mixing is not a concern, these TIP reactors offer a cost-effective alternative to "large-format" GSMs and are fully compatible with Uniqsis HotChip, ColdChip, PhotoChip and Polar Bear Plus GSM reactor modules.

Uniqsis Ltd
For info: +44-1223-942004
web: www.uniqsis.com

LNP-based T Cell Transfection Kit

Cayman Chemical's new LipidLaunch™ C14-4 T Cell Exploration Kit (Item No. 41820) is intended to serve as a starting point for laboratories to explore the feasibility of using C14-4-based lipid nanoparticles (LNPs) for applications, including T cell transfection, without the need for specialized equipment. This kit features conveniently packaged reagents and protocols for preparing LNPs and is adaptable to encapsulation of mRNA or siRNA cargo and testing different formulation parameters. LNPs created using this kit are composed of the lipids C14-4, 1,2-DOPE, cholesterol, and DMG-PEG(2000), which are included.

Cayman Chemical
For info: +1-800-364-9897
www.caymanchem.com/41820

96-Channel Liquid Handling System

Eppendorf introduces its 96-cannel liquid handling system, the epMotion 96 Flex. Pioneering a bridge between handheld electronic pipettes and fully automated pipetting systems, the epMotion 96 Flex assures significantly improved efficiency and throughput compared to manual pipetting. It also provides a cost-effective alternative to automated liquid handling systems, with a smaller footprint on the lab bench. The exchangeable dispensing heads of epMotion 96 Flex allow effortless adaptation to various volume ranges. At launch, two dispensing heads are available for the system, covering volume ranges from 0.5 to 300 µL and 5 to 1000 µL. The epMotion 96 Flex offers a touchscreen interface and smartly placed interaction points at the base of the device ensuring user-friendly and precise operations. Combined with its compact size it can be used comfortably in biosafety cabinets. Calibrate according to the new ISO 23783-2 standard the epMotion 96 Flex offers pipetting accuracy and precision. Beyond its application in manual workflows, the system serves as a valuable addition to large automated systems. In the role of a "Feeder" system, it prepares plates for further automated processing.The versatile epMotion 96 Flex complements both manual and automated workflows, adapting seamlessly to diverse applications, from molecular biology to cell culture.

Eppendorf SE
For info: +49 40 69 21 06 268
www.eppendorf.com

UNIVERSITÉ DE GENÈVE

The University of Geneva is a versatile institution with an international reputation, ranked among the best universities in the world.

Its Faculty of Medicine has 3,500 students and over 200 research groups. Together with the Geneva University Hospitals, it is a major centre for teaching, research and innovation in medicine and life sciences.

The Faculty's Section of Fundamental Medicine hosts the National Centre of Competence in Research « Evolving Language », which brings together biological, linguistic and cultural evolutionary approaches to unravel the emergence of new modes of communication, such as human language. To this aim, a position is opened within the Department of Genetics and Development which brings together Genetics/Gemomics and Developmental Biology to address both human disease and phenotypes :

**ASSOCIATE PROFESSOR or
ASSISTANT PROFESSOR with TENURE TRACK
in the field of genetics/epigenetics of language development**

JOB DESCRIPTION
• Establish and maintain a competitive research programme supported by external funding
• Ensure pregraduate and postgraduate teaching within the Faculty of Medicine, particularly in the domain of fundamental research
• Lead a research team
• Take part in scientific activities of the National Centre of Competence in Research « Evolving Language » and the host Department

REQUIREMENTS
•   PhD or MD-PhD
•   Proven expertise in genetic/epigenetic influences on human language development, through studies of human genetics and associated communication disorders, potentially incorporating evolutionary evidence from ancient DNA
•   Ability to develop a competitive line of research focusing on the genetics of cognitive and/or social components of language through studies of human genetics
•   and associated communication disorders
•   Beneficiary of competitive research funds and author of publications in leading peer-reviewed journals
•   Capacity for interdisciplinary collaboration
•   Candidates without French language skills are encouraged to apply. The successful candidate will be supported in developing French proficiency over the first three years.

STARTING DATE: September 1st 2025, or according to agreement
Mandatory online registration before 25 november 2024 at: http://www.unige.ch/academ
Additional information may be obtained from: olivia.michel@unige.ch

Women are encouraged to apply

Assistant Professor of Medicine- Basic Science Pathway

The UNIVERSITY of OKLAHOMA Health Sciences Center

GENERAL DESCRIPTION
The Division of Nephrology and Hypertension within the Department of Internal Medicine, College of Medicine, at the University of Oklahoma Health Sciences Center (OUHSC) seeks applicants for a full-time, consecutive term, Assistant Professor of Medicine- Basic Science Pathway with an anticipated start date of July 1, 2025. Expected starting salary for the position is $90,000 per year. The Department of Medicine encourages applications from all interested and qualified applicants.

The successful candidate will be expected to develop a nationally recognized, independently funded, research program in the broad field of Nephrology, Immunology, and Computational Biology. Preference will be given to applicants looking to develop an independent research program focused on immune-mediated mechanisms of acute kidney injury (AKI) and polycystic kidney disease (PKD). Applicants with prior experience using multi-omics approaches to studying immune cell involvement in AKI and PKD are strongly encouraged to apply. Additional preference will be given to applicants looking to build on the current research and clinical strengths of the Department which include Trem2 biology, immune-mediated diseases, and cancer biology.

The Department of Medicine at the University of Oklahoma Health Sciences Center is part of a 325-acre comprehensive academic health center. OUHSC is comprised of seven colleges on a single urban campus located in the heart of metropolitan Oklahoma City, an energetic and growing community with mild winters, year-round sunshine, low cost of living, and plenty of activities.

REQUIRED AND PREFERRED QUALIFICATIONS
Required qualifications:
•   Ph.D. or MD degree in the biological sciences.

PREFERRED QUALIFICATIONS
•   Ph.D. degree in Cell Biology or Immunology related field.
•   Post-doctoral fellowship in a Nephrology or Immunology related area.
•   Expertise in immune mediated mechanisms of acute kidney injury or polycystic kidney disease.
•   Experience in computational biology and immunology.

APPLICATION INSTRUCTIONS
Applicants are invited to submit a cover letter (2-page maximum), CV, and three letters of recommendation to Kurt Zimmerman (kurt-zimmerman@ouhsc.edu). The cover letter may address examples of the candidate's approach to teaching, research, service, and graduate student mentorship. Additional materials may be requested at a later date. The search will remain open until filled.

MISSION OF THE UNIVERSITY OF OKLAHOMA
The Mission of the University of Oklahoma is to provide the best possible educational experience for our students through excellence in teaching, research and creative activity, and service to the state and society.

EQUAL EMPLOYMENT OPPORTUNITY STATEMENT
The University of Oklahoma, in compliance with all applicable federal and state laws and regulations does not discriminate on the basis of race, color, national origin, sex, sexual orientation, genetic information, gender identity, gender expression, age, religion, disability, political beliefs, or status as a veteran in any of its policies, practices, or procedures. This includes, but is not limited to: admissions, employment, financial aid, housing, services in educational programs or activities, or health care services that the University operates or provides.

WHY YOU BELONG AT THE UNIVERSITY OF OKLAHOMA
The University of Oklahoma, fosters an inclusive culture of respect and civility, belonging, and access,which are essential to our collective pursuit of excellence and our determination to change lives. The unique talents, perspectives, and experiences of our community enrich the learning, and working environment at OU, inspiring us to harness our innovation, creativity, and collaboration for the advancement of people everywhere.

TENURE-TRACK PROFESSOR IN EVOLUTIONARY BIOLOGY, HARVARD UNIVERSITY

The Department of Organismic and Evolutionary Biology (OEB) seeks to appoint a tenure-track professor in evolutionary biology. The broad search field includes, but is not limited to, evolutionary genetics, quantitative genetics, population genetics, phylogenetics and systematics, adaptation, evo-devo, speciation, eco-evo dynamics, evolutionary biomechanics, evolutionary phenomics, macroevolution, and all experimental, theoretical, or computational approaches. If applicable, the experimental system may be viral, prokaryotic, or eukaryotic. The successful candidate is expected to establish a dynamic, externally funded research program, teach at the undergraduate and graduate levels, mentor a diverse community of students, and contribute to departmental and university service. Should the candidate's expertise fall into an area of available curatorships in the Museum of Comparative Zoology or Harvard Herbaria, a joint appointment as curator could also be considered. The appointment is expected to begin on July 1, 2025.

OEB is a supportive and collaborative research community with interests in a broad range of topics that include evolution, ecology, biomechanics, behavior, development, paleobiology and public health. Junior faculty are assigned a mentoring committee that advises them on all aspects of setting up a successful research group.

An appointment in OEB provides access to state-of-the-art animal and plant facilities and core facilities for imaging, proteomics, genomics, and bioinformatics as well as access to multiple supporting institutions including the Museum of Comparative Zoology, Harvard University Herbaria, Harvard Forest, and the Arnold Arboretum. Please visit the OEB webpage (https://oeb.harvard.edu/) for more information.

OEB celebrates the multiple dimensions of diversity that each member of our community offers, including diversity of background, perspective, and lived experience. We strongly welcome applications from persons from underrepresented groups.

Doctorate or terminal degree in Evolutionary Biology or related discipline required by the time the appointment begins.
Demonstrated strong commitment to teaching and advising is desired.

Please submit the following materials through the ARIeS portal at:
https://academicpositions.harvard.edu/postings/14051
Review of applications will begin on November 1, 2024; applications will be reviewed until the position is filled. We anticipate conducting in-person interviews in January 2025.

1. Cover letter
2. Curriculum Vitae
3. Teaching/advising statement (describing the candidate's teaching philosophy and practices as well as their approach to creating a learning environment in which students are encouraged to ask questions and share their ideas)
4. Research statement
5. Service statement describing efforts to strengthen academic communities, e.g., the candidate's department, institution, and/or professional societies.
6. Names and contact information of 3-5 referees, who will be asked by a system-generated email to upload a letter of recommendation once the candidate's application has been submitted. Three letters of recommendation are required, and the application is considered complete only when at least three letters have been received. At least one letter must come from someone who has not served as the candidate's undergraduate, graduate, or postdoctoral advisor.
7. Up to four representative publications.

The health of our workforce is a priority for Harvard University. With that in mind, we strongly encourage all employees to be up-to-date on CDC-recommended vaccines.

For administrative questions, please contact Christian Flynn (cflynn@fas.harvard.edu). For search-related questions, please contact Michael Desai (mdesai@oeb.harvard.edu).

UT Southwestern
Medical Center

FACULTY POSITION IN MICROBIAL PATHOGENESIS

The Department of Microbiology at UT Southwestern Medical Center (UTSW) seeks a new faculty member at the Assistant Professor (tenure-accruing track) level. Ideal candidates will develop a front-rank, competitive, and independent research program in the field of microbial pathogenesis. Some preference may be given to candidates working on bacterial pathogens. At present, research in the department spans a wide range of topics including viral and bacterial gene regulation, metabolism and motility of bacterial pathogens, animal reservoirs for bacteria and viruses, microbiota, innate immune responses, evolution of microbes and host defense, molecular interplay between viruses and bacteria and their hosts, and resistance to antimicrobials. An important academic responsibility will be the instruction and mentoring of graduate students. An attractive start-up package, including a competitive salary and generous laboratory space in a modern building, is available to conduct research in a highly dynamic and collaborative environment (https://www.utsouthwestern.edu/departments/microbiology/). Candidates will be considered for our $2.2M Endowed Scholars (start-up) Program (http://www.utsouthwestern.edu/education/programs/nondegree-programs/other-programs/endowed-scholars/index.html).

Candidates should have a Ph.D. and/or M.D. degree with 3+ years of postdoctoral experience and an exceptional publication record. Candidates should apply online at https://jobs.utsouthwestern.edu (job ID 808000) and upload a cover letter, C.V., contact information for three letters of recommendation, and a brief statement of current and future research interests. UT Southwestern Medical Center is committed to an educational and working environment that provides equal opportunity to all members of the University community. As an equal opportunity employer, UT Southwestern prohibits unlawful discrimination, including discrimination on the basis of race, color, religion, national origin, sex, sexual orientation, gender identity, gender expression, age, disability, genetic information, citizenship status, or veteran status.

This position is security-sensitive and subject to Texas Education Code 51.215, which authorizes UT Southwestern to obtain criminal history record information.

Appointment rank will be commensurate with academic accomplishment and experience. Consideration may be given to applicants seeking less than a full-time schedule.

To learn more about the benefits UT Southwestern offers, visit https://www.utsouthwestern.edu/employees/hr-resources/

Who's the top employer for 2023?

Science Careers' annual survey reveals the top companies in biotech & pharma voted on by *Science* readers.

Read the article at sciencecareers.org/topemployers

Science 2023
TOP EMPLOYER

ARE YOU THE KIND WHO WORKS TO HELP ALL HUMANKIND?

🔍 **Find your next job at ScienceCareers.org**

Whether you're looking to get ahead, get into, or just plain get advice about careers in science, there's no better or more trusted authority. Get the scoop, stay in the loop with *Science* Careers.

Science Careers
FROM THE JOURNAL SCIENCE ◣AAAS

The Institute for Medical Engineering and Science (IMES) http://imes.mit.edu/ at the Massachusetts Institute of Technology (MIT), in Cambridge, Massachusetts, invites applicants to apply for junior tenure-track faculty positions in the area of health sciences to begin July 1, 2025 or on a mutually agreed date thereafter. Established in 2012, IMES is an inclusive community of students, administrative staff, research staff, and faculty who research, work, educate, and learn, at the convergence of engineering, science, and medicine to transform human health for all.

The appointment will be made in partnership with a relevant science or engineering department. The position will have equal academic responsibilities in IMES and the partner department. Applicants should hold an MD or a PhD in a science or engineering field related to health sciences by the beginning of employment. While all research areas are of interest, IMES encourages applications from candidates whose research and teaching interests would be enhanced and accelerated through deep engagement with both basic and clinical partners, and whose research program intimately involve the development and/or application of advanced computational methods. Examples include, but are not limited to:

- Clinical data generation, consumption and sharing to improve health care delivery;
- The development, testing, and analysis of devices, sensors, interfaces, & vectors (mechanical, optical, electrical, chemical), and their products, to advance health care and its delivery;
- Methods and modeling to understand the impact of climate change on human health;
- Deriving insights into human diseases through epidemiological research and mechanistic modeling;
- Improving our ability to understand, predict, and manipulate human immune responses (e.g., in the context of infectious diseases, neuro-immune interactions, or women's health)

The successful candidate will be expected to fulfill all teaching, research, and service obligations of a junior faculty member at MIT. Faculty duties include teaching courses in health sciences and related subjects at the undergraduate and graduate levels, advising students, conducting original scholarly research, developing course materials at the graduate and undergraduate levels, and service to MIT and the profession.

Interested candidates should submit application materials electronically https://faculty-searches.mit.edu/imes-search/. Each application must include: a curriculum vitae; the names and addresses of three or more references; a strategic statement of research interests; and a statement of teaching interests. It is the responsibility of the candidate to arrange for reference letters to be uploaded at http://imes-search.mit.edu/letters.

Please address questions to imes-search@mit.edu. Responses received by December 1, 2024 will be given priority.

With MIT's strong commitment to diversity in engineering and science education and research, we especially encourage those who will contribute to our diversity and outreach efforts to apply.

MIT is an equal employment opportunity employer. We value diversity and strongly encourage applications from individuals from all identities and background. All qualified applicants will receive equitable consideration for employment based on their experience and qualifications and will not be discriminated against on the basis of race, color, sex, sexual orientation, gender identity, pregnancy, religion, disability, age, genetic information, veteran status, ancestry, or national or ethnic origin. See MIT's full policy on Nondiscrimination (https://policies.mit.edu/policies-procedures/90-relations-and-responsibilities-within-mit-community/92-nondiscrimination). Know your rights.

Employment is contingent upon the completion of a satisfactory background check, including possible verification of any finding of misconduct (or pending investigation) from prior employers.

FACULTY POSITION –
Pediatric Thromboembolic Disease
CENTER FOR CLINICAL AND TRANSLATIONAL RESEARCH

The Abigail Wexner Research Institute at Nationwide Children's Hospital seeks a faculty candidate to establish an independent research program for the prevention and treatment of thromboembolic disease in children. A successful candidate for any rank (Assistant, Associate, or Full Professor) will have:

• a research laboratory focused on thrombosis, hemostasis, and/or vascular biology

• an MD, MD/PhD, or PhD

• a strong track record of publishing in high quality journals

Principal investigators are also faculty members at The Ohio State University College of Medicine. To learn more about the opportunity, visit Center for Clinical and Translational Research (NationwideChildrens.org).

Send correspondence and CV to
Patrick.Kennedy@NationwideChildrens.org
in Scientific Recruitment.

![Nationwide Children's - When your child needs a hospital, everything matters.] | ![The Ohio State University College of Medicine]

W1380850

By **Erika Cyphert**

The Poland experiment

"**W**hat have I gotten myself into?" I wondered as I sat through a lengthy research symposium conducted entirely in Polish, unable to even read the x- and y-axis labels on the plots. I had packed up my life in the United States and moved to Poland for a 10-month fellowship during my Ph.D., having only rudimentary Polish language skills, and I was the only native English speaker in my research group. My experiment in cross-cultural collaboration was looking more challenging than I had expected.

The idea of spending time abroad during graduate school hadn't occurred to me until my adviser suggested it during my first year. I spent that summer in Japan, in the same lab where he had worked some 20 years earlier and formed a lasting collaboration. It was a remarkable experience that left me wanting more. So, with my adviser's support, I applied for a Fulbright fellowship in Poland, which was seeking STEM researchers and was one of the few countries that didn't require language proficiency. To receive the fellowship, I first had to find a lab to host me, and neither I nor my adviser had any connections to Poland.

Relying on Google to translate lab websites, I began to send cold emails. When my eventual host responded, eager to collaborate, I was elated. I was struck that the simple gesture of emailing a scientist across the world whose work I admired could open new avenues. I suddenly felt the possibilities for my research career were endless and global.

But first I faced some practical difficulties. As in Japan, learning how to run equipment in my new lab turned out to be a humbling experience. The software that ran many of the instruments was in Polish. Deciphering simple buttons such as "save" and "run" and setting up an automated protocol seemed like a momentous task with my primitive grasp of the language. I remember my paranoia that I would click the wrong buttons and break the equipment the first time I ran the instrument alone, and the words *plik* and *zapisać* ("file" and "save") have now been ingrained in my psyche.

I was the first visiting scholar my research group had hosted from the United States, and my Polish colleagues were curious what brought me to their country. This led to many fruitful discussions about the similarities and differences in how science is conducted across cultures. For example, much of the scientific funding in Poland is awarded to teams or centers rather than individual investigators. The system promotes formation of cross-disciplinary collaborations across the country and the European Union, which in turn shapes how scientists design and carry out studies.

"My cultural immersion made me a better researcher."

I learned as much about the history and culture in Central and Eastern Europe as I did scientifically. Colleagues related poignant firsthand accounts about challenges the region faced during the Cold War. Many laboratories had limited resources, and they rarely published in English, which hindered timely communication of scientific discoveries. Those conditions helped shape research on antimicrobial treatment. During the Cold War, the "West" prioritized the study of antibiotics, whereas the "East" prioritized therapy with bacteriophages—viruses that attack bacteria. To this day, many Central and Eastern European countries, including Poland, remain leaders in bacteriophage research.

My experience in Poland inspired me to think more broadly about the questions I could tackle as a scientist. I became fascinated by the microbiome and how it might affect the way people from different cultures respond to infectious disease, as well as its role in antimicrobial resistance. As a postdoc I dove into bioinformatics as a way to pursue these questions, temporarily abandoning my Ph.D. training in biomaterials and microbiology.

As I start my independent lab, I'm combining elements of my Ph.D. research and this new focus. I'm also encouraging my students to pursue research in other countries and cultures. I've tried to demystify the application process by sharing my own application materials.

My cultural immersion made me a better researcher. I now see cross-cultural collaborations as a way to overcome communication barriers that can hinder progress in science. By stepping outside my comfort zone, I gained the confidence to pivot my research in new directions that I might have otherwise boxed myself out of. ■

Erika Cyphert is an assistant professor at the University of California San Diego. Send your career story to SciCareerEditor@aaas.org.