AAAS

Science

28 AUGUST 2025

RETURN OF THE HERD

Ecosystem effects of migrating bison **p. 904**

The Chen Institute and *Science* launched the "Chen Institute and *Science* Prize for AI Accelerated Research" in August 2024. This initiative aims to drive advancements in artificial intelligence that can expand scientific research. Young scientists from around the world are invited to submit AI-driven projects that demonstrate significant potential to improve research and lives.

The competition will award a Grand Prize of $30,000, with the winner's essay published in *Science* and an accompanying five-year AAAS membership. Additionally, up to two runners-up will receive $10,000 each, with their essays published on *Science* Online and the same membership benefits, promoting sustained engagement with scientific progress.

Apply at

cheninstitute.org/prize

CONTENTS

28 AUGUST 2025 | VOLUME 389 | ISSUE 6763

870

A museum outside Mexico City houses the largest trove of mammoths ever found in the tropics.

PHOTO: EDGARD GARRIDO/REUTERS

EDITORIAL

NEWS

FEATURES

COMMENTARY

PERSPECTIVES

BOOKS ET AL.

ANALYSIS

REVIEWS

RESEARCH

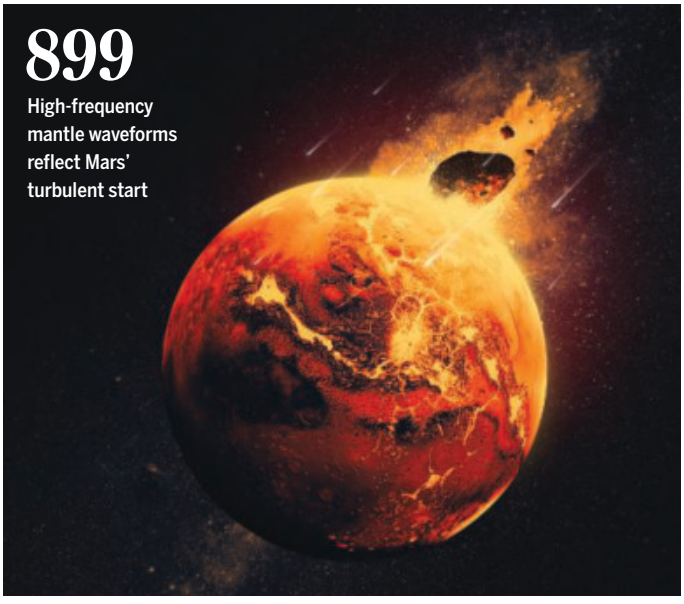


ILLUSTRATION: @VADIMSADOVSKI

ON THE COVER

Bison move through Yellowstone's Lamar Valley at sunrise. Their movements and grazing accelerate the nitrogen cycle, increasing the annual nutrition that plants provide to herbivores. After decades of recovery, bison now add heterogeneity that sustains soil nutrient storage and plant productivity while allowing plant communities to become more diverse, highlighting the importance of restoring native grazers in large numbers and with freedom to move. See page 904. Photo: NPS/Jacob W. Frank

WORKING LIFE

Science **serves as a forum for discussion of important issues related to the advancement of science by publishing material on which a consensus has been reached as well as including the presentation of minority or conflicting points of view. Accordingly, all articles published in *Science*—including editorials, news, commentary, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.** *Science* (ISSN 0036-8075) is published weekly on Thursday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. The title *Science* is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $2865; Foreign postage extra: Air assist delivery: $135. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A. Change of address:** Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling available from backissues.science.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

Science AAAS

EDITORIAL
science_editors@aaas.org

NEWS
science_news@aaas.org

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

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

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

MEDIA CONTACTS
scipak@aaas.org

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

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

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

MEMBERSHIP AND INDIVIDUAL SUBSCRIPTIONS
science.org/subscriptions

MEMBER BENEFITS
aaas.org/membership/
benefits

INSTITUTIONAL SALES AND SITE LICENSES
science.org/librarian

Taking nuclear energy to the Moon

Kathryn Huff

Earlier this month, Sean Duffy, the acting head of the US National Aeronautics and Space Administration (NASA), announced an acceleration of the agency's Fission Surface Power program, with the ambitious goal of placing a nuclear reactor on the Moon by 2030 that can produce 100 kilowatts of electric power. A surface reactor will be essential to enable sustained human exploration on both the Moon and Mars. However, an overly aggressive schedule could compromise both technical readiness and NASA's other scientific priorities.

Space reactors are not new. In 1965, the United States put the first nuclear-powered satellite into orbit, which produced about 500 watts for 43 days before a non-nuclear component failed prematurely. From 1967 to 1988, the Soviet Union launched 33 nuclear fission–powered Earth-observing satellites, including two that produced 5 kilowatts. In 1978, the unfortunate failure of a support system resulted in the uncontrolled reentry into Earth's atmosphere of a Soviet satellite, scattering radioactive debris across 20,000 square miles of Canadian territory. Clearly, reactors can operate in space, but integration with support systems may ultimately determine mission success.

...NASA could make this program a cornerstone of future exploration.

Still, nuclear fission is the best available power source to operate through the 14-day lunar night or on the harsh surface of Mars, though the proposed reactor will be larger than its predecessors. Earlier units were approximately the size of a car, but NASA's new reactor will likely exceed that of a standard freight shipping container.

Mass will be the first constraint. Launch and landing systems must accommodate the reactor's size and weight without sacrificing stability or control. NASA's goal of keeping system mass under six metric tons would have been challenging even for transporting the smaller 40-kilowatt reactor envisioned in 2022. With an increased target power output of 100 kilowatts, the proposed 15-ton lander payload limit presents a formidable reactor design constraint.

Thermal management will be equally challenging. The reactor must control coolant behavior in low gravity, reject waste heat in a vacuum, and maintain stable operation despite the Moon's extreme temperatures, which can swing by 200°C between day and night. Accordingly, the directive requires a specific type of system to convert heat into electricity without exhausting fluids into the thin lunar atmosphere. It must maximize thermal efficiency while powering habitats, life-support systems, rovers, communications equipment, and devices that convert water ice to oxygen and hydrogen.

Safety begins before launch. To limit the hazard if a launch fails, the reactor will not operate until it reaches the Moon and will only contain fresh (very low radioactivity) uranium fuel during its ascent. NASA, the Department of Energy, and the Federal Aviation Administration have experience launching nuclear materials, and a reactor of this scale will require extensive safety review and radiological contingency planning. Once deployed, it will need shielding to protect nearby crew and autonomous power control systems, including the ability to shut down automatically as terrestrial reactors do.

These challenges are solvable. The greater question is whether the 2030 target aligns with the mission it is meant to serve. No crewed lunar base is currently scheduled to operate in 2030. Building a reactor years before its intended users arrive risks wasting resources. Unless NASA's budget grows substantially, prioritizing this program could slow or cancel other critical missions that enable observations of the Universe, climate, and Earth. The directive frames the program's urgency in competitive terms, noting Chinese and Russian plans to deploy a lunar reactor by the mid-2030s. No international law prohibits peaceful nuclear power on the Moon. And according to the 1967 Outer Space Treaty, building infrastructure on the Moon establishes a country's presence but not a territorial claim. However, establishing a safety perimeter around a reactor could limit access to that region by other countries. Duffy hence warned of potential "keep-out zones" on the Moon and declared, "We want to get there first." Although innovation is sometimes spurred by competition, modern science thrives on collaboration. NASA's Artemis program, which aims to establish a long-term Moon base and prepare for future human missions to Mars, has demonstrated that international partnerships can accelerate progress toward shared goals. Although cooperation with every nation may be infeasible, space exploration should remain a collective human endeavor, not a national possession.

With technical discipline, realistic scheduling, and a commitment to science, NASA could make this program a cornerstone of future exploration. If urgency overwhelms readiness, the United States risks undermining its own ambitions in space and climate resilience at home. □

Kathryn Huff is an associate professor in the Department of Nuclear, Plasma, and Radiological Engineering, University of Illinois at Urbana-Champaign, Champaign, IL, USA and is a Public Voices fellow of The OpEd Project. kdhuff@illinois.edu

10.1126/science.aeb6479

Changing the course: Stopping Alzheimer's before it starts

Scientists are redefining Alzheimer's disease, with new tools that can measure biological signals before late-stage cognitive symptoms appear. This new understanding is leading to a generation of therapies designed to intercept the disease before it takes hold.

For more than a century, Alzheimer's disease has been defined by its tragic endpoint: the slow, heartbreaking erosion of memory, cognition, and self. A diagnosis was often a confirmation of what families already knew and feared, coming long after the disease had taken root in the brain. But scientists are redefining Alzheimer's disease from a late-stage cognitive syndrome to a biological condition that can be identified, and potentially intercepted, before overt cognitive symptoms appear.

"We will look back in 20 years and say, 'I can't believe we waited until symptoms appeared before we started treating Alzheimer's disease,'" says Bill Martin, the global therapeutic area head for neuroscience at Johnson & Johnson (J&J). "We know now that the disease pathology is detectable before symptoms even appear." This shift represents one of the most significant leaps forward in neuroscience and could completely change the fates of the patients affected by Alzheimer's as well as the people who care for them.

From post-mortem to predictive science

When Alois Alzheimer, a German psychiatrist and neuropathologist, first described the disease in 1906, the markers he identified, amyloid and tau proteins, could only be seen in the brain post-mortem. For decades, these clumps of misfolded proteins remained the only definitive confirmation. However, over the past 20 years, advances in science have shifted the process from diagnosing based on symptoms to detecting biological changes in living patients. "We are redefining diagnosis of the disease," Martin explains.

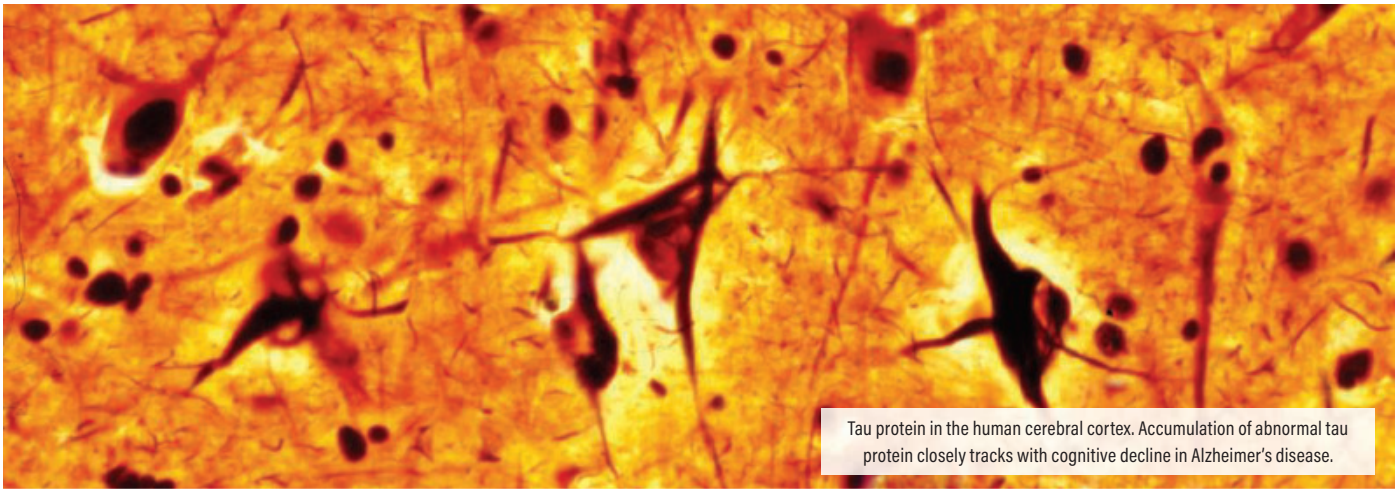
Though amyloid protein buildup is believed to trigger the disease process decades before symptoms manifest, it is the subsequent spread of abnormal tau protein that most closely tracks with the neurodegeneration and cognitive decline that patients experience. This strong connection suggests that tau may be a critical therapeutic target. As Martin explains, focusing on tau could be a reliable way to target and track disease. Because the presence of pathological tau proteins is so directly linked to a patient's brain health, any therapy that successfully slows, or even reverses, the spread and accumulation of tau could produce a distinct and measurable benefit to cognitive function in a clinical trial.

The power of precision

This focus on tau is a key part of J&J's "precision neuroscience" strategy to deliver the right therapy to the right patient at the right time, Martin says. In practice, this means using a suite of sophisticated tools to build a detailed picture of a person's individual disease biology.

One of these tools is positron emission tomography (PET), which can visualize tau in the brain, allowing researchers to see not only if the pathology is present, but also how much there is and where it has spread. In fact, the location of tau in the brain

IMAGE: PROVIDED BY JOHNSON & JOHNSON



Tau protein in the human cerebral cortex. Accumulation of abnormal tau protein closely tracks with cognitive decline in Alzheimer's disease.

is closely related to specific symptoms.

PET imaging provides the most precise map, but its accessibility and cost make it impractical for wide-scale screening. Researchers are therefore putting considerable effort into finding blood-based biomarkers. J&J, for example, has developed and utilized a test for its clinical trials that measures a specific fragment of tau called phosphorylated tau 217 (ptau217), a blood-based biomarker closely associated with disease pathology.

> **"We will look back in 20 years and say, 'I can't believe we waited until symptoms appeared before we started treating Alzheimer's disease.'"**
>
> **— Bill Martin, global therapeutic area head for neuroscience at Johnson & Johnson**

At J&J, researchers are using this blood test as a pre-screening tool to identify candidates for their anti-tau clinical trials, who then go on to have more detailed PET imaging to confirm the findings of the blood test. This precision approach helps ensure that trials are populated with the right patients at the right stage of disease. It also allows for a more streamlined approach to screening, as the more costly PET is reserved for patients whose tau status has already been confirmed via blood test.

A new generation of tau therapies

This move toward early and precise diagnosis is also reshaping how researchers are approaching therapeutics. The ability to define Alzheimer's disease by its specific protein signatures and patient subtypes is a first step toward personalized medicine. New interventions have already been shown to modify the course of the disease at a biological level in the brain, and now, researchers are building on that progress to develop even more precise treatments that can slow or stop disease progression.

One approach is a tau-targeting monoclonal antibody, which received United States Food & Drug Administration (FDA) Fast Track designation earlier this year. It is designed to slow the spread of tau pathology in the brain. Another J&J anti-tau immunotherapy, JNJ-2056 (formerly known as ACI-35.030, in collaboration with AC Immune), was granted Fast Track designation by the FDA in July of last year. It is the first active immunotherapy targeting tau to be investigated for people with preclinical Alzheimer's disease, and it works by triggering the patient's own immune system to attack disease-causing tau molecules, says Martin.

This research is a starting point that sets the stage for a more comprehensive future, says Martin. "Very few diseases, let alone of this complexity, are managed by a single intervention." Instead, he believes the field is headed toward combination therapies, in which a patient's personal biological profile, identified through advanced diagnostics, will guide the selection of a specific combination of treatments. "Two years ago, I told my daughter, '[Your generation] is the last that's going to know Alzheimer's disease as we know it today,'" Martin shares. "As I look at it today, I stand by that prediction. I remain really inspired by the progress that the field is seeing."

SCIENCE POLICY

Supreme Court quashes restoration of NIH grants

Emergency ruling may prompt Trump administration to again cancel research it dislikes

MEREDITH WADMAN and **JOHN TRAVIS**

Annie Luetkemeyer, an infectious disease researcher at the University of California (UC) San Francisco, was leaving a research meeting last week when the U.S. Supreme Court ruined her day. She had been about to relaunch a study on the effects of using an antibiotic to prevent sexually transmitted infections in certain sexual and gender minorities. The National Institutes of Health (NIH), which funded the project, had killed it in March, not long after President Donald Trump's executive order barring grants that "promote gender ideology," but a federal judge in June had ordered the funding and many other canceled NIH grants restored.

Now, in a 5-4 decision, a Supreme Court majority had granted an emergency stay of that ruling and signaled it was unlikely to reverse any federal grant cancellations upon appeal. The news was "very disheartening," Luetkemeyer says, as NIH is free to again terminate her grant and more than 800 others on politically charged topics such as COVID-19 and diversity, equity, and inclusion.

The high court's verdict threatens an estimated $2 billion in multiyear NIH funding and left medical research proponents aghast. "This was a decision critical to the future of the nation, and the Supreme Court made the wrong choice," the Association of American Medical Colleges said in a statement. "History will look upon these mass [NIH] research grant terminations with shame."

This legal battle arose soon after Trump took office, when NIH began to terminate as many as 2300 grants. In early April, two lawsuits challenging some of the terminations were filed and on 16 June, William Young of U.S. District Court for the District of Massachusetts found the cancellations were arbitrary and capricious, making them illegal under a law called the Administrative Procedure Act (APA). He also found the terminations discriminated against racial, sexual, and gender minorities.

Young ordered the reinstatement of 848 grants held by plaintiffs in the case, studies scheduled to get nearly $800 million this fiscal year. An appeals court rejected the administration's request to stay the order, and NIH had restored most of the grants by this month. But Trump officials had appealed to the Supreme Court, leading to last week's decision.

> "Labs will shutter, people will lose jobs."
>
> **Gregg Gonsalves**
> Yale University

The high court was divided on whether Young had the legal authority to order the NIH grants be paid and whether the government would suffer "irreparable harm" if forced to deliver the funding before legal appeals are resolved. Although five of the conservative justices scolded the district court for asserting a jurisdiction they said it did not have, and should have known it did not have because of a previous high court ruling, Chief Justice John Roberts joined the court's three liberals in a dissent.

Legal observers say the stay allowing the grant terminations appears to be a clear win for the White House's efforts to end scientific research it dislikes. "The reasoning [the majority] gave for putting the district court's injunction on hold seems to make clear that their view is this case shouldn't proceed at all," says Samuel Bagenstos, a University of Michigan law professor who from 2022 to 2024 was the senior lawyer in the Department of Health and Human Services (HHS), NIH's parent agency.

The majority in the ruling stated that the plaintiffs, which include state attorneys general, scientists, and groups representing them, should have brought their challenge of NIH's terminations not to the district court, but to the Court of Federal Claims, a specialized venue for contract disputes. But Bagenstos says the claims court has no jurisdiction to restore federal grants. "The take-home message is the Supreme Court is going to deny you the remedy of getting your grant reinstated, and instead, the Supreme Court is essentially going to send you to a different court where, after potentially many years, you might possibly get some damages."

The Supreme Court ruling may also indicate rough times ahead for UC faculty who filed a class action lawsuit against research grant terminations from multiple federal agencies and also in June won an injunction from a different district court. The U.S. government similarly argues in that case that the judge didn't have authority over the matter and last week's ruling suggests the Supreme Court may concur.

"They effectively said all grant termination litigation needs to go to the Court of Federal Claims. … There's a very good chance that that whole case will disappear," says Scott Delaney, a former attorney and an environmental epidemiologist at the Harvard T.H. Chan School of Public Health who maintains a database of science grants canceled by the Trump administration.

He predicts other legal challenges to federal grant cancellations will also be quashed. "It gives the government

a very, very big arrow in its quiver for all of those cases."

NIH had not yet terminated previously restored grants as *Science* went to press, and HHS declined comment, citing the ongoing litigation. But many researchers are pessimistic. "Labs will shutter, people will lose jobs," Gregg Gonsalves, an epidemiologist at Yale University, wrote in a Bluesky post. For her part, Luetkemeyer is now hesitating to rehire for her project as she awaits NIH's response to the ruling.

Justice Amy Coney Barrett did join Roberts and the three liberal justices in concluding Young had the jurisdiction to find that NIH probably violated the APA with its policy guidance that prompted the grant cutting. But even if that finding is sustained on appeal, it would still leave grantees whose projects are canceled in the future with no opportunity to go to federal district court to get them reinstated.

In a separate dissent, Justice Ketanji Brown Jackson called the notion that the plaintiffs would have to pursue legal remedies in two courts "bizarre." She also scoffed at the rationale several justices gave for the stay: that the plaintiffs had made no offering that they would return their NIH grant money if they ultimately lost their case on appeal. Jackson wrote: "Government does not come close to offering any reason for us to intervene. Its asserted harm—its only asserted harm—is that it might have to keep paying out grants it has already committed to paying for the few months it will take to appeal the District Court's decision."

She contrasted that with the harm scientists would likely suffer if they have to wait for appeals to play out, writing: "Make no mistake: Per the evidence in front of the District Court, the forward march of scientific discovery will not only be halted—it will be reversed. … Yearslong studies will lose validity. Animal subjects will be euthanized. Lifesaving medication trials will be abandoned. Countless researchers will lose their jobs. And community health clinics … will close."

For Luetkemeyer, the past 6 months have already exacted a steep price. She expected to have hundreds of participants in her study at this point, rather than the single person who enrolled just before NIH cut the funding. "We're all in this field because we're trying to do good science and serve people and answer important questions," she says. "And it feels like [we've spent] an incredible amount of time and heartache and energy not doing those things because of this." □

IN FOCUS

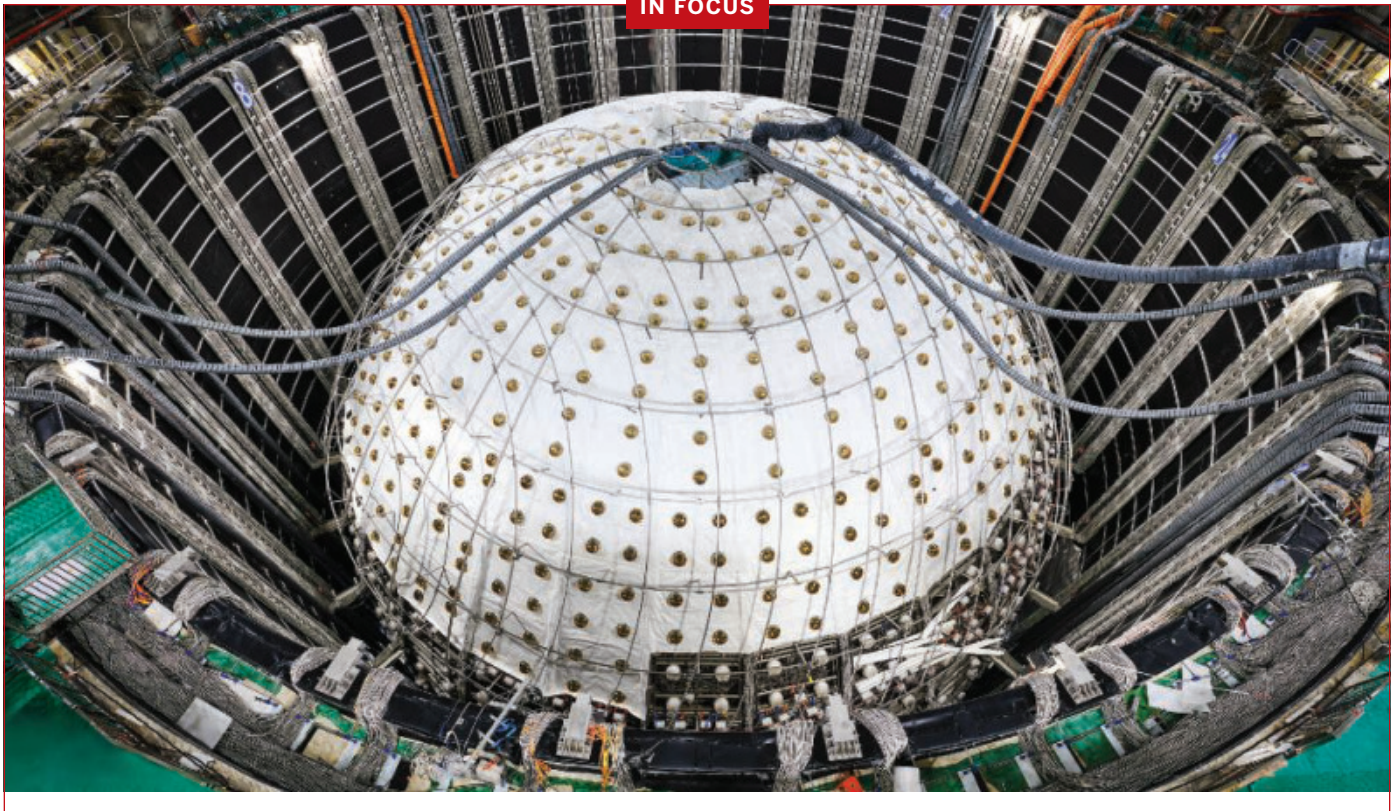


PHOTO: ENRICO SACCHETTI

China's new neutrino observatory comes online

China's Jiangmen Underground Neutrino Observatory (JUNO) started taking data this week that may resolve a puzzle about the masses of the ghostly particles. Built by the Chinese Academy of Sciences's Institute of High Energy Physics, JUNO has a 35.4-meter-diameter vessel (above) holding 20,000 tons of a liquid scintillator surrounded by 45,000 photomultiplier tubes to pick up the telltale flashes of light that result when antineutrinos from distant nuclear reactors slam into atomic nuclei in the liquid. Data from JUNO should indicate which of three neutrino mass states is heaviest, a finding affecting physics, astrophysics, and cosmology. —Dennis Normile

IN OTHER NEWS

PIG LUNG WORKS, BRIEFLY, IN HUMAN

In a small step forward for the long-struggling field of xenotransplantation, surgeons in China have put a pig lung into a brain-dead man and shown it can function, although the organ only lasted 9 days before succumbing to an immune assault. The organ came from a miniature breed of pig with three of its genes for cell surface sugars deleted and three human genes added, changes made to help the lung work in a person without being rejected. The lung delivered oxygen to the man's blood and removed carbon dioxide, according to a report this week in *Nature Medicine*. The organ also avoided an initial antibody attack, but soon swelled with edema as immune factors reacted to the foreign tissue. The experiment "clearly illustrates that lung xenograft survival in humans will require more genetic engineering and optimized immunosuppression," says animal cloning specialist Dengke Pan, a co-author of the study and founder of Chengdu ClonOrgan Biotechnology, which makes the pigs. Meanwhile, a gene-edited pig kidney has continued working in a U.S. man for 8 months, nearing the 9-month record for a xenotranplant in a human. —Jon Cohen

PHOTO: RICH VON BIBERSTEIN/ICON SPORTSWIRE VIA AP IMAGES

GENETICS

Gaps remain in college sports' 15-year-old sickle cell policy

Report recommends education, genetic counseling with NCAA screening

RODRIGO PÉREZ ORTEGA

In September 2006, Dale Lloyd II, a first-year football player at Rice University, collapsed during a routine team workout and never got back up. The 19-year-old's death was later linked to sickle cell trait (SCT)—a genetic condition he hadn't known he carried, and that can make intense exercise dangerous.

The tragedy was a turning point in college sports. Lloyd's parents sued the university and the National Collegiate Athletic Association, and, as part of a settlement, NCAA mandated in 2010 that all Division I athletes either show proof of prior screening for SCT, undergo new testing, or sign a liability waiver. The policy soon expanded to all NCAA sports. At the time, it was one of the most ambitious genetic screening efforts by a private organization in the United States.

The policy has likely saved lives, but a comprehensive new study funded by the National Human Genome Research Institute also finds shortcomings in its implementation. Conducted by researchers at Duke University and the University of North Carolina at Chapel Hill and described in a report published this week on a Duke website, the study finds the testing itself can be inadequate, and that many athletes who test positive don't get comprehensive advice about how to reduce their risk, or genetic counseling about the implications of the trait for future generations.

"I'm really impressed" with the report, says Lakshmanan Krishnamurti, a pediatric hematologist at the Yale University School of Medicine. The findings are "not a surprise," he adds. It's particularly troubling that protocols long used in the military to mitigate SCT-related risks, such as ensuring good hydration and not exercising outside on very hot days, are still not universal in college sports, he says.

A spokesperson for NCAA's national office declined to comment on the study.

SCT is caused by a mutation in the gene for hemoglobin, the oxygen-carrying protein in red blood cells. The condition disproportionately affects Black people, about one in 13 of whom carry the mutation. People with two copies have sickle cell disease (SCD), in which misshapen red blood cells can lead to chronic pain, anemia, and organ damage. But those with SCT carry a single copy and often have no symptoms or health risks—unless they engage in extreme physical exercise. Most of the reported deaths related to SCT are from exertional rhabdomyolysis—a condition in which muscle breaks down, releasing toxic proteins into the bloodstream that can damage organs. A 2012 study in the *British Journal of Sport Medicine* reported that athletes with SCT had a risk of exertional death 37 times higher than those without the trait.

Duke University defensive tackle Aaron Hall tested positive for sickle cell trait in a screening program mandated by the National Collegiate Athletic Association.

Since the NCAA policy was implemented, the number of football players who died of complications related to SCT has plummeted, according to a 2020 paper in *Sports Health: A Multidisciplinary Approach*. Nine players died in the 10 years before July 2010, and only one in the 9 years after. But deaths still happen: Last year, 18-year-old Calvin "C.J." Dickey Jr. collapsed during football practice at Bucknell University and died 2 days later from exertional rhabdomyolysis. He had tested positive for SCT weeks earlier.

In the new study, Duke geneticist Charmaine Royal and her colleagues interviewed nearly 1900 athletes, trainers, coaches, and team physicians across 172 schools about their experiences with the NCAA policy. About 2% of the participating athletes carried SCT. And even though newborns in the U.S. have been routinely screened since the mid-1970s for hemoglobin abnormalities, more than 40% of those athletes found out they had SCT during the NCAA screening.

> **We shouldn't just treat these people as football players ... [but] as whole, holistic human beings.**
>
> **Lainie Ross**
> University of Rochester

More than 85% of participants supported SCT screening. But the report raises concerns about the testing method. NCAA recommends a "solubility test," which detects the presence of abnormal hemoglobin caused by the sickle cell mutation in blood with very high sensitivity but can miss abnormalities from other harmful gene variants that can pose risks to athletes. The report says players who test positive should have a confirmatory "electrophoresis" test, which uses electrical current to separate different types of hemoglobin and can distinguish SCT, sickle cell disease, and other hemoglobin disorders such as beta thalassemia. "I think it's unconscionable to test these kids with solubility testing and not do hemoglobin electrophoresis," says pediatrician and philosopher Lainie Ross of the University of Rochester, who attended summits convened by the study's authors to discuss the NCAA policy.

The researchers also found that many schools don't have clearly documented protocols for athletes with SCT. An NCAA fact sheet recommends precautions, including limiting all-out exertion to 3 minutes without resting, but the organization doesn't say how coaches and athletes should handle SCT-related complications.

Aaron Hall, a Duke student athlete who participated in one of the summits, found out he carried SCT during the NCAA screening as a freshman. He says coaches and staff gave him and his teammates plenty of information about the trait. "If I ever do encounter symptoms, I kind of know what to do myself," he says. But he may be an exception; the study found that only 21% of coaches felt knowledgeable about SCT facts, and only 11% of all athletes did.

The surveys also revealed that although NCAA recommends that people with positive results seek genetic counseling, only 20% of the colleges offered it, and no athlete in the study had received it after a positive SCT test. Schools have "an obligation to do pre- and post-test counseling," Ross says. "We shouldn't just treat these people as football players. ... You should treat them as whole, holistic human beings."

On the positive side, the report finds little evidence to support concerns that screening players might lead to discrimination based on SCT status. At Duke, athletes are required to wear red wristbands to alert staff and teammates they have SCT. "I am totally against those bands," Ross says. "This is private health information." But Hall says he's "extremely comfortable" wearing a risk band because it allows teammates and staff take care of each other.

Royal hopes NCAA takes the report seriously and doesn't "throw the baby out with the bathwater" in addressing the recommendations. "We propose a whole bunch of strategies" for improving the policy, she says, "so they have lots of options." □

ARTIFICIAL INTELLIGENCE

AI hypotheses lag human ones when put to the test

Machines still face hurdles in identifying fresh research paths, study suggests

JEFFREY BRAINARD

In May, scientists at FutureHouse, a San Francisco–based nonprofit startup, announced they had identified a potential drug to treat vision loss. Yet they couldn't fully claim the discovery themselves. Many steps in the scientific process—from literature search to hypothesis generation to data analysis—had been conducted by an artificial intelligence (AI) the team had built.

All over the world, from computer science to chemistry, AI is speeding up the scientific enterprise—in part by automating something that once seemed a uniquely human creation, the production of hypotheses. In a heartbeat, machines can now scour the ballooning research literature for gaps, signaling fruitful research avenues that scientists might otherwise miss.

But how good are the ideas? A new study, one of the largest of its kind, finds the AI-generated hypotheses still fall short of human ones, when researchers put them through real-world tests and get human evaluators to compare the results. But not by much. And maybe not for long.

A paper describing the experiment, posted to the arXiv preprint server in June, suggests AI systems can sometimes embellish hypotheses, exaggerating their potential importance. The study also suggests AI is not as good as humans at judging the feasibility of testing the ideas it conjures up, says Chenglei Si, a Ph.D. student in computer science at Stanford University and lead author of the study.

The research is drawing praise but also caution from others in the field, in part because judging originality is so difficult. "Novelty is the bugbear of scientific evaluation and one of the most difficult tasks in peer review," says Jevin West, a data scientist at the University of Washington.

The study examined hypotheses about AI itself, in particular natural language processing (NLP), which underpins AI tools called large language models (LLMs). The researchers tasked Claude 3.5 Sonnet, an LLM developed by the startup Anthropic, with generating thousands of ideas based on an analysis of NLP studies in the Semantic Scholar database and ranking the most original ones. The researchers then paid human NLP specialists to come up with compet-

ing ideas. The team also recruited another group of computer scientists to judge the novelty, feasibility, and other qualities of the two sets of ideas, to which the reviewers were blinded. They gave the AI ideas higher marks on average, a surprising finding the team reported in a 2024 preprint that drew media attention.

But the tables turned in the study's second phase. After advertising via social media and other routes—including on a T-shirt Si wore to conferences—the team recruited a new team of paid NLP specialists to run experiments for 24 of the AI-generated ideas and 19 human ones. The tests typically examined how a proposed algorithm could improve an aspect of an LLM, such as its language translations, and the experimenters were empowered to tweak the study design by, for example, choosing a data set better suited to evaluate the hypothesis. The team once again got independent evaluators to judge the hypotheses. Average overall scores for the AI ideas dipped on a 10-point scale from 5.382 to 3.406, whereas human ideas fell from 4.596 to only 3.968.

Si says the results show the importance of putting hypotheses to the test. "If you only look at the ideas, some reviewers can get fooled by how exciting certain words sound, but when you actually look at that code execution or interpretation of it, you'll realize it's just a fancy or novel phrasing of a known technique." (That concern was echoed in a February study of 50 AI hypotheses: Human evaluators judged one-third to have been plagiarized, with another third partially borrowed from previous work. Only two were mostly novel and none were completely novel.)

The study is "really exciting" but has limitations, says Dan Weld, chief scientist at the nonprofit Allen Institute for Artificial Intelligence. For one, he says, it relied on a single LLM to generate hypotheses based on a wide body of relevant research rather than using multiple AI tools to scour highly cited studies written by prominent specialists. Also, humans are not necessarily the best judges of novelty either: Previous studies have found that actual researchers disagree substantially in how they score the same computer science papers. An experiment's novelty is best evaluated in hindsight, after years of accruing citations, West adds.

Si says it would be too time-consuming to have humans groundtruth AI-generated hypotheses as a matter of course. But LLMs could get better at recognizing novel hypotheses if they are trained on the details of past, successful experiments, he suggests.

Despite the questions, the AI and human scores were remarkably close—something that would have shocked researchers even a few years ago. Weld wouldn't be surprised if, eventually, AI comes up with most hypotheses and scientists are left to carry out the parts of the experiments that can't be automated with robots. But if that's the case, it removes "the most fun part of science" and leaves scientists to conduct lab work that "is sometimes brain-numbing," West says. "Science is a social process that involves humans. You take that out, then what is it?" □

BY THE NUMBERS

4

Distinct giraffe species, according to an analysis of scientific data released last week by the International Union for Conservation of Nature. The world's tallest mammal had long been labeled a single species with as many as nine subspecies. The new classification may help conservation managers design efforts specific to each species.

ORIGIN OF LIFE

New clues found about the assembly of life's first proteins

Lab study shows how RNA could have helped amino acids join up—without preexisting protein machinery **ROBERT F. SERVICE**

Life today depends on proteins, cellular workhorses that do everything from flex muscles to ferry oxygen. And proteins, in turn, depend on RNA, which carries the recipes for making them and also helps with their assembly. In modern cells, large protein-based enzymes help connect RNA snippets to amino acids, the building blocks of proteins. Then, the RNA- and protein-based cellular machine called the ribosome stitches the amino acids together into a protein chain, reading the correct sequence from a longer strand of messenger RNA. But billions of years ago, before the evolution of the enzymes and the ribosome, how could life's first proteins have been constructed?

Researchers say they've now come up with a plausible route by which RNA and amino acids could have paired up to assemble small proteins called peptides, without the help of complex enzymes or the ribosome. The work, published this week in *Nature*, offers a glimpse of how RNA might have helped form the first simple proteins, an event that could have set the stage for evolution.

"This is the first step to allowing life's information molecules to encode peptides," says Matthew Powner, a chemist at University College London (UCL) who led the study. Thomas Carell, a chemist at the Ludwig Maximilian University of Munich, says the study "gives wonderful insights" into "one of the big riddles in prebiotic chemistry."

At the heart of the protein-nucleic acid partnership is the process by which RNA gets amino acids to link up. Amino acids don't naturally connect into proteins. They must first be chemically "activated," whereupon they link up like boxcars in a train. Today, biology speeds all this up with large enzymes that prompt amino acids to react with a cellular energy molecule called ATP. This creates a form of an amino acid that can chemically join with RNA to create an activated combo known as aminoacyl-RNA. Only then can these compounds be fed to the ribosome, which snips off the RNA bits as it links the amino acids together into peptides.

Researchers believe alkaline lakes such as Turkey's Yarişli Lake might have been ripe for the assembly of Earth's first proteins.

PHOTO: EKREM SAHIN/ANADOLU VIA GETTY IMAGES

Since the 1970s, numerous groups have tried an array of chemical strategies to produce activated aminoacyl-RNAs under conditions simulating a primitive environment like early Earth's. But the reactions have typically worked poorly, if at all, and they often produce compounds that are unstable in water, suggesting they wouldn't have had much staying power on early Earth, Powner says.

In hopes of doing better, Powner and his colleagues turned to an energy-rich compound called pantetheine that's involved in numerous metabolic reactions in cells and could have formed in lakewater environments on early Earth, as the researchers reported last year. In the new study, the researchers added pantetheine to water containing amino acids. The amino acids reacted with it, creating a combo known as an aminoacyl-thiol that was ready to react with RNA. "It forces the amino acids to react with RNA and not with each other," says Moran Frenkel-Pinter, an origin of life chemist at the Hebrew University of Jerusalem.

RNA can take two forms, either as a single- or double-stranded molecular helix. When Powner and his colleagues tried to react the aminoacyl-thiols with single-stranded RNA, the amino acids bound to it chaotically, all along the RNA. But when Powner's team used double-stranded RNA, which more closely resembles the kind that pairs with amino acids in cells, the reactions yielded activated aminoacyl-RNAs with structures matching those normally produced by enzymes. "It's quite an impressive achievement," says Nick Lane, an origin of life chemist at UCL who is not affiliated with Powner's group. "It looks like life's chemistry even if it's a bit different."

Powner and his colleagues took their experiment one step further. When they enriched their brew with hydrogen sulfide and compounds called thioacids, both of which were also likely abundant on early Earth, the amino acid pieces of the aminoacyl-RNAs were triggered to link up into peptides, all without the need for enzymes or ribosomes.

Lane cautions that the new work still only produces peptides with a random arrangement of amino acids, unlike the genetically controlled ordering carried out by ribosomes. "They still have not cracked that problem," Lane says. Powner agrees, but says the current work already reveals that some RNAs have a slight preference for interacting with particular amino acids over others. "This is just the first step," he says.

Even with that piece outstanding, Lane says he believes the new work is helping bridge the gap between divergent origin of life camps. For decades, one group of researchers has emphasized the central role of RNA, whereas another group—the so-called "metabolism first" camp—has argued that self-sustaining chemical networks needed to appear before self-replicating genetic molecules like RNA could arise to make use of them. With the new evidence that energy-rich compounds played a key role in linking amino acids to RNA, Lane says, "this whole field is moving together a little bit." □

IN OTHER NEWS

MEDICAL MUSEUM'S NEW SENSITIVITY Philadelphia's Mütter Museum, infamous for its medical curiosities such as a 2.5-meter colon and abnormal skeletons, outlined new policies for its handling of human remains earlier this month. Critics had said its displays disrespected the dead and disregarded the medical field's racist history. The Mütter's 2-year Postmortem Project had since scrutinized the museum's collections and approach to displaying them. The museum says it will now add context and biographical details to its specimens and that from now on, human remains will only be accepted if pledged by a living person or donated by their descendants. It also commits to returning specimens to communities that did not consent to their acquisition, and to having Indigenous American tribes consult on the Mütter's collection. —Hannah Richter

CHINA VISA FOR STEM SCHOLARS Hoping to bolster its international competitiveness and foster academic cooperation, China has created a new K visa category intended to make it easier for young science and technology professionals to visit the country. Currently, non-Chinese researchers seeking to attend conferences, meet collaborators, or investigate exchange opportunities must find official sponsors. Starting in October, young graduates of renowned universities with bachelor's or higher degrees in science, technology, engineering, and math will be able to apply for K visas on their own, according to the Ministry of Justice website. The K visa will also offer greater leeway on the number of times a person can visit and their length of stay. —Dennis Normile

FAMINE VERDICT CHALLENGED Gaza City and its surrounding governorate are now experiencing famine and other parts of Gaza may follow or be in worse shape, according to a global expert body on food insecurity. But even before the Integrated Food Security Phase Classification group released its official assessments last week in two reports, the Israeli military and other critics attacked the science underlying them, arguing that IPC changed its criteria to make declaring a Gaza famine easier. Independent food security experts and others close to IPC's deliberations say those assertions are false. The contested changes to the body's famine criteria were made 6 years ago, long before the start of the latest Israel-Hamas war, they point out. —Leslie Roberts

IMAGE: NORTH WIND PICTURE ARCHIVES/ALAMY

EARTH SCIENCE

Europe's biggest quake may foretell Atlantic 'ring of fire'

Earth's mantle is peeling from the crust in the eastern Atlantic, a possible sign of the ocean's eventual closure **EVAN HOWELL**

On 1 November 1755, the earth shuddered violently beneath Lisbon, Portugal. The seas swelled and a firestorm followed, killing tens of thousands who had gathered on All Saints' Day. Surveys in the aftermath revealed 85% of Lisbon's buildings lay in ruins, allowing scientists centuries later to estimate the earthquake's magnitude at 8.7, cementing it as Europe's most destructive.

Mysteriously, the quake struck far from any known subduction zone—regions such as those off Japan and South America where one tectonic plate dives beneath another, forming a deep ocean trench, fueling volcanoes, and driving some of Earth's largest quakes. A study published this week in *Nature Geoscience* may finally explain the Lisbon disaster, while also hinting at the Atlantic Ocean's geological future.

The study shows that, like the sole peeling off an old shoe, the mantle—the hot, dense layer that makes up much of Earth's interior—is sloughing off beneath offshore Portugal. This process, called delamination, was thought to be impossible in old, strong oceanic crust. "To our knowledge," says University of Lisbon geologist and lead author João Duarte, "delamination has never been identified in a pristine oceanic basin."

The delamination process could generate enough stress to trigger major quakes. It may also offer a rare look at the origins of subduction itself—and point to a future millions of years from now when subduction could spread into the Atlantic. The result might resemble the Pacific Ocean's vast chain of subduction zones, which form a "ring of fire" that fuels roughly 450,000 earthquakes each year and about 68% of all global volcanic eruptions since 1960.

"Subduction initiation is a big enigma and a domain of frontier research," says Taras Gerya, a geologist at ETH Zürich who was not involved

A devastating magnitude 8.7 earthquake rocked Lisbon in 1755. New research suggests sinking of the mantle off Portugal's coast was the culprit.

in the study. Most Pacific subduction zones are mature, so finding new ones forming is critical, he explains. "Seeing thousands of chickens doesn't help you imagine an egg."

Tectonically speaking, the Atlantic is fairly calm. Down its center runs a jagged midocean ridge that quietly oozes basalt, forming new crust as the plates on either side spread apart at 2.5 centimeters per year. Offshore Portugal's relatively featureless ocean basin has none of the usual calling cards associated with large, tsunami-spawning subduction zone earthquakes, like deep ocean trenches and volcanic arcs. When another earthquake struck off Portugal's coast in 1969—a magnitude 7.8 or 7.9 event that killed 25 people—scientists were left scratching their heads.

Yet earlier seismic studies had revealed a puzzling geophysical anomaly: a curious blob plunging 250 kilometers into the mantle. Additionally, the epicenters of the 1969 quake and other recent earthquakes clustered unusually deep beneath the Horseshoe Abyssal Plain, an expanse of deep Atlantic sea floor southwest of Portugal made of some of Earth's oldest, densest oceanic crust. Whereas most earthquake epicenters lie within 10 kilometers of the surface, these quakes originated below 20 kilometers, suggesting something unusual was taking place at that depth.

To explain these observations, Duarte and colleagues built a computer simulation of the crust and mantle in this region, testing a range of different conditions to explain the data. In their "best fit" model, a chunk of the mantle is delaminating from the crust and sinking into Earth like a teardrop. This process, researchers say, could generate enormous stresses and major earthquakes.

This kind of unpeeling is highly unusual in oceanic crust. Normally, oceanic crust and its underlying mantle form what Duarte calls a "crème brûlée structure": a brittle, buoyant crust resting on top of a denser, rigid mantle layer with a squishier layer underneath. This arrangement typically keeps the mantle layer from sinking. But along the Horseshoe Abyssal Plain, two tectonic plates converge at a slight angle, applying huge amounts of stress to the crust there. Seismic surveys also reveal a section where the basalt crust is missing and seafloor sediments rest directly on a mantle layer. When the Atlantic first opened, extreme stretching—and too little magma—left the mantle bare.

Duarte and his colleagues believe this shallow mantle was weakened when water seeped into its fractures, triggering chemical reactions that transformed it into a greenish rock type called serpentinite. "It's weak, it's slippery, it's less dense: all that good stuff that would allow the dense plate underneath to decouple from the overlying crust," says Durham University geologist Mark Allen, who was not involved in the study.

Researchers now believe the central block of this patch of mantle has begun to sink, its passage eased by two enormous fracture zones that flank it. As the thin crust peels away, more mantle is drawn into the depths. This self-reinforcing process may reveal how subduction begins in the first place—a mechanism that "hasn't been proposed before," Gerya says. He adds that large earthquakes of the kind that rocked Lisbon in 1755 could be the first "fingerprint" of delamination in action.

Marc-André Gutscher, a geologist at the University of Brest, questions whether it explains the 1755 quake. For one, he's skeptical that the study's proposed arrangement of undersea faults could have produced a transatlantic tsunami capable of reaching the Caribbean Sea, as one did in 1755. If the fault runs northeast-southwest as proposed, he explains, any such wave will spread toward Greenland and Morocco—but not toward the Caribbean. "When a fault slips, the tsunami wave shoots out perpendicular to its length," Gutscher says. "It's like headlights on your car."

But Gutscher agrees the discovery of Atlantic delamination has implications for the future. A small subduction zone in the western Mediterranean, known as the Gibraltar Arc, is already "invading" westward. Delamination could be the first sign of subduction extending into the Atlantic's eastern margin. Eventually, the ocean could stop spreading and begin to close. The Americas would start to converge with Africa and Europe, setting the stage for Earth's next supercontinent.

Although researchers don't expect the Atlantic to change course anytime soon, the seismic hazard remains high. In Duarte's native Portugal, he laments, "there is no culture" of seismic risk; Lisbon didn't have a tsunami warning system until 2017. "For most people, earthquakes are part of history," he says. "This is very dangerous." □

Evan Howell is a science journalist in Colorado.

Archaeologists found hundreds of mammoth skeletons during the 3-year construction of the new Felipe Ángeles International Airport north of Mexico City.

PHOTO: RICARDO FLORES/XINHUA VIA ZUMA

BONES BENEATH THE RUNWAY

How a military megaproject led to Mexico's biggest paleontological discovery—and is reshaping what we know about mammoths

RODRIGO PÉREZ ORTEGA, in Mexico City

The story of Mexico's biggest fossil discovery begins with politics, not science. During his campaign for the country's presidency in 2017, Andrés Manuel López Obrador pledged to scrap a half-built new airport in eastern Mexico City, citing cost overruns and corruption over the course of its construction. Once in office, he replaced it with an alternative plan: to build the Felipe Ángeles International Airport at a military base in Santa Lucía, 50 kilometers north of the capital. In October 2019, the army began to dig.

Massive bones appeared almost immediately.

The first mammoth emerged on 5 November. Then another. And another. Within weeks, the six archaeologists who had been called in had swelled to 56, supervising more than 400 construction workers. Excavators paused whenever bone fragments surfaced. To meet construction deadlines, several excavation areas were opened simultaneously across the site. "We were overwhelmed because every day we were finding them—every day, every day, every day," recalls Rubén Manzanilla López, an archaeologist with the country's National Institute of Anthropology and History (INAH) who supervised the salvage excavation, which was supported by the Mexican military.

By 2022, Manzanilla López and his team had amassed more than 50,000 Pleistocene bones from just 3700 hectares. Among them are at least 500 mammoths, 200 camels, 70 horses, 15 giant ground sloths, as well as the remains of dire wolves, saber-toothed cats, bison, armadillos, birds, freshwater snails—and one human skeleton. Nicknamed Yotzin ("unique" in Nahuatl), the man may have died during a hunt or been trampled by a mammoth. The scale of the discovery rivals—and in some ways surpasses—California's La Brea Tar Pits, the most famous ice age fossil site in North America.

It is also bringing a once-obscure population of tropical mammoths into the light. In the popular imagination, mammoths are cold-adapted furry giants plodding across the frozen tundra. And indeed, most mammoth fossils are found in high-latitude landscapes such as Siberia, Alaska, or Canada. Yet one species, the Columbian mammoth (*Mammuthus columbi*), ranged as far south as Costa Rica. Little was known about these peculiar southern populations and how they relate to their more numerous northerly relatives.

The Santa Lucía mammoths, trapped not in tar like the La Brea fauna, but in the shallows of an ancient lake, are now filling in the picture of the species. From the bones, Mexican scientists have recovered the first-ever DNA from tropical mammoths, gleaning new insights into the beasts' evolutionary history.

"If you had told me 5 years ago that I would be collecting these samples, I would have said, 'You're crazy,'" says the scientist who led the DNA analysis, Federico Sánchez Quinto, a paleogenomicist at the International Laboratory for Human Genome Research of the National Autonomous University of Mexico (UNAM).

In a paper published online in *Science* this week, Sánchez Quinto and his team depict the Santa Lucía mammoths as a previously unknown lineage that split from northern Columbian mammoths hundreds of thousands of years ago; they also shed light on how these animals fared before ultimately dying out. The findings open a new, more complex chapter in mammoths' evolutionary story, says Adrian Lister, a paleobiologist at the Natural History Museum in London who was not involved in the research. "This paper really is an exciting beginning of something."

MAMMOTHS HAVE an evolutionary history as complex as the lands they once crossed. The massive, tusked herbivores evolved from ancient African ancestors about 6 million years ago and migrated across Eurasia starting about 3 million years ago. The woolly mammoth (*M. primigenius*), adapted to cold climes, emerged in Siberia and crossed into North America. Columbian mammoths—larger and less hairy—appeared later in North America and spread south into Central America.

In 2021, Love Dalén, a paleogenomicist at the Centre for Palaeogenetics in Sweden, and his colleagues sequenced the genomes of three mammoths from Siberia and compared them with previously sequenced genomes of other mammoths. They found evidence that Columbian mammoths were hybrids. Dalén's team argued they emerged when descendants of a 1-million-year-old male steppe mammoth (*M. trogontherii*)—a Siberian species found in the Krestovka locality—mated with female woolly mammoths. This hybridization event, Dalén and his colleagues hypothesized, happened at least once in the early Middle Pleistocene, about 800,000 to 400,000 years ago. But the subsequent evolutionary history of the Columbian mammoths remained cloudy.

Like all Mexicans, Sánchez Quinto had heard the news of the mammoth trove at Santa Lucía. Trained in Europe, he had just returned to Mexico to set up a lab. His expertise was in human paleogenomics, but the chance to work on mammoths—especially on ones from such an unusual setting—was irresistible.

The noisy, dusty excavations were nothing like the hushed genetics labs Sánchez Quinto was used to. "It was an apocalyptic scene," he recalls. The combination of multiple open excavation areas, heavy machinery, and the sprawling footprint of the airport meant archaeologists were often rushing from one find to another. In many cases, mammoth skulls were quickly cataloged and stabilized with wooden structures. Tusks especially, softened by millennia of exposure to water and minerals, crumbled easily because they weren't fully fossilized.

> "Every day we were finding them—every day, every day, every day."
>
> **Rubén Manzanilla López**
> National Institute of Anthropology and History

"We tried to save as much as we could," says Joaquín Arroyo Cabrales, a paleontologist and mammoth expert with INAH who was leading a simultaneous research project at the site. At one point, bones filled a whole hangar. Researchers also took samples from soil and from fossils of small animals—axolotls, frogs, rabbits, and even flamingos—to figure out everything from the chemistry of the ancient lake to the ecosystem the mammoths were part of. So far, there are five ongoing research projects.

One is Sánchez Quinto's. With permission from Arroyo Cabrales,

he and his team set up a sterile tent in a military barracks to drill into 73 massive teeth—each the size of a shoebox—in hopes of finding ancient DNA. Although several mammoths had been found around the Basin of Mexico since the 20th century, nobody had extracted enough DNA from them in good enough condition to be analyzed.

Recovering DNA from fossils in the tropics is notoriously difficult: Heat and humidity break down the molecules far faster than in the frozen north. But the inner dentine of the molars from the Santa Lucía mammoths had preserved their genetic history. Still, Sánchez Quinto feared too little nuclear DNA had survived to be studied, and he decided to focus on mitochondrial DNA (mtDNA), inherited only from the mother. Although mtDNA carries less information than nuclear DNA, it is simpler, shorter, and far more abundant.

Sánchez Quinto remembers the moment he first saw the genetic results from the Santa Lucía molars. "My heart skipped a beat," he says. "I literally lost my breath." He and his team succeeded in sequencing 61 complete mitochondrial genomes, more than doubling the global total for Columbian mammoths.

"It's impressive from Federico's team to get mitochondrial genomes out of so many mammoths from such a low latitude and such an old age," Dalén says. "It's a tour de force."

In the new *Science* paper, Sánchez Quinto and his colleagues draw conclusions from the DNA about the evolution of both the Columbian mammoths and the population that once roamed the Basin of Mexico. They find the Mexican mammoths are genetically distant from Columbian mammoths from Canada and the United States—the tropical giants formed an entirely separate branch even further removed from woolly mammoths from North America and Eurasia.

Genetic dating suggests this lineage diverged from northern populations between 400,000 and 300,000 years ago and then persisted in isolation, becoming so distinctive that Lister asks, "Should we really be calling it a Columbian mammoth? [Or] give it a new name—a Mexican mammoth?"

Within the Mexican lineage, the researchers identified three distinct sublineages, suggesting each experienced long periods of isolation in the grassy highlands of the Basin of Mexico. Volcanic ranges may have acted as natural enclosures, sequestering these groups into genetic clusters.

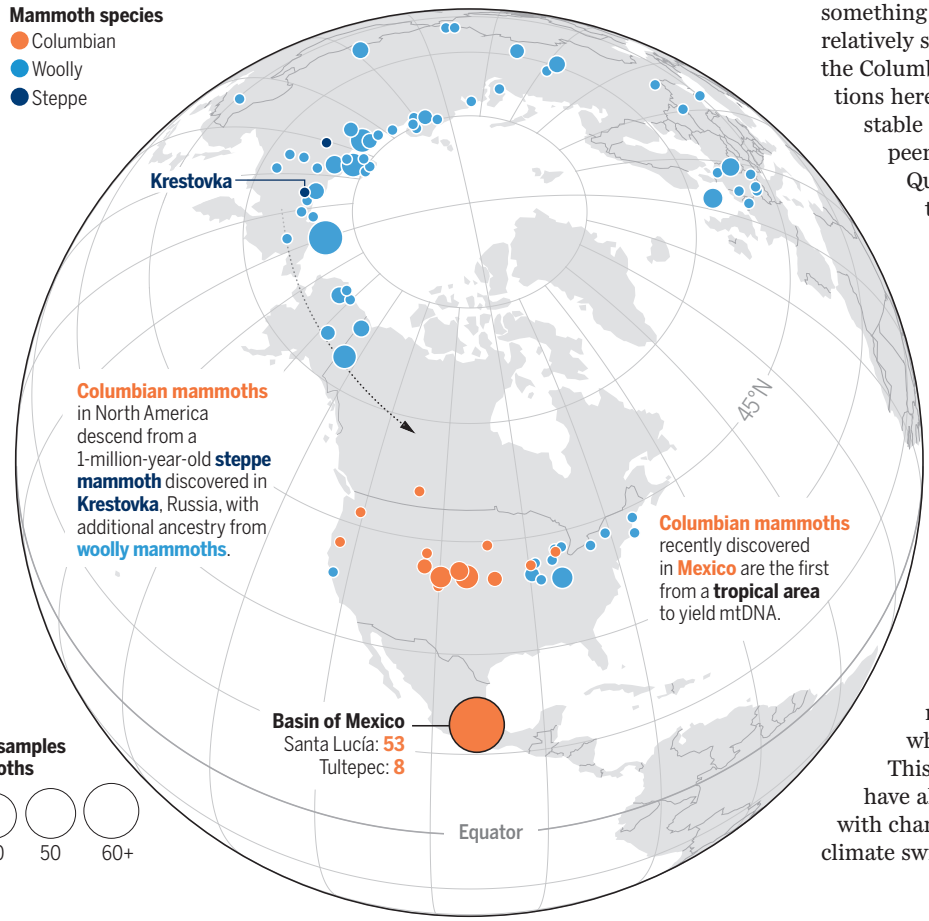
Because the Mexican lineage is substantially different from the ones in Canada and the U.S., the researchers suggest hybridization events between woolly and steppe mammoths happened more than once around the same time, producing multiple Columbian mammoth lineages that spread across the continent.

The Mexican mammoths are "not just a little offshoot" of the larger northern population, Dalén says. "Rather, [they] have hundreds of thousands of years of history in Mexico. It's a proud lineage."

RADIOCARBON DATING of five samples puts the Santa Lucía mammoths at between 16,000 and 11,000 years old. At the younger end of that range, other mammoths in North America were in serious decline because of shifting climates, human hunting pressures, or some combination thereof. Yet the genetic data from Santa Lucía suggested something unusual: Despite having relatively small population sizes, the Columbian mammoth populations here appear to have remained stable even as their northerly peers died off. Sánchez Quinto cautions, however, that this remains to be confirmed with evidence from younger samples. Environment and diet may explain that stability. In 2023, another group of Mexican researchers analyzed carbon isotopes—which contain clues to an animal's diet—in mammoth teeth from Santa Lucía. They found the mammoths ate a mix of shrubs, trees, and grasses, unlike woolly mammoths farther north, which ate mainly grasses. This dietary flexibility would have allowed them to cope with changes in vegetation amid climate swings.

Stomping grounds

Different species of mammoths lived throughout much of Eurasia and North America from 3 million to about 4000 years ago. Mammoths from cold, northern latitudes have yielded most of the genetic material analyzed so far, but Mexican researchers recovered mitochondrial DNA (mtDNA) from a mammoth population near what is now Mexico City. DNA comparisons are shedding light on mammoth family history.



CREDITS: (GRAPHIC) M. HERSHER/*SCIENCE*; (DATA) D. CHANG *ET AL.*, *SCIENTIFIC REPORTS* (2017); P. PEČNEROVÁ *ET AL.*, *EVOLUTION LETTERS* (2017); M. WOOLLER *ET AL.*, *SCIENCE* (2021); T. VAN DER VALK *ET AL.*, *NATURE* (2021); ARRIETTA-DONATO *ET AL.*, *SCIENCE* (2025)

Bones found over 3 years of excavation near Mexico City overwhelmed multiple storage spaces and were eventually moved to a giant warehouse.

PHOTO: MARIA AVILA-ARCOS

Still, their low population numbers may have eventually caught up with them.

Silvia González, a geoarchaeologist at Liverpool John Moores University, had previously suspected Mexican mammoths suffered from low genetic diversity. Nearby, at a site called Tocuila, she had seen deformed molars and skeletal anomalies—signs of inbreeding. "It's clear that this is a terminal population," she says.

Ultimately, the Mexican mammoths vanished about 11,000 years ago, alongside much of North America's megafauna. Whether climate change, human hunting, or both sealed their fate remains uncertain, Sánchez Quinto says.

THE ROAR OF jet engines now fills the air where the mammoths once grazed. Political attention faded and military funding for research dried up once the airport was completed in 2022. But while the army was still flush with construction money, it built a paleontological museum at Santa Lucía, as well as a research center and a storage facility—both donated to INAH to study and store the fossils. In February 2022, the Quinametzin Museum opened, named after the Nahuatl word for "giant." In its main hall, a mounted female mammoth skeleton named Nochipa ("eternal" in Nahuatl) towers over visitors. The museum displays a fraction of the site's richness: skulls of different species, and vertebrae stacked like prehistoric beads. Interactive exhibits explain the excavation process and the late Pleistocene ecosystem of the Basin of Mexico.

Just a few meters away, INAH paleontologist Felisa Aguilar Arellano leads a small team that safekeeps a huge, climate-controlled storage facility full of fossils. Packaged bones or various sizes fill endless racks. Between them, mammoth skulls—some with their tusks still attached—are cradled in yellowish foam and wrapped in plastic. It's impossible to move through the fossil riches. Aguilar Arellano says the facility contains only 40% of the total collection. The rest is stored elsewhere on the base.

Tens of thousands of bones remain unstudied. Scientists are now seeking grants and forging collaborations, including with researchers at the La Brea Tar Pits, to date key specimens and refine extinction timelines, Aguilar Arellano says. Sánchez Quinto, for his part, says his team now aims to isolate the nuclear DNA from several mammoths, hoping to sharpen their picture of the animals' ancestry.

Aguilar Arellano hopes the collection spurs a shift in Mexico, where archaeology has long overshadowed paleontology. "I believe that the paleontological [potential] … continues to be neglected," she says. Sánchez Quinto is already witnessing such shift—his project has inspired graduate students to study mammoth genetics and Pleistocene ecosystems. "The knowledge generated here has a local impact and contributes to forming new lines of research, and training tomorrow's scientists."

María Ávila Árcos, a human population geneticist at UNAM who led the sequencing with Sánchez Quinto, takes an even broader view. "For me, the message that I'm left with is beyond just the history of mammoths. By studying these underrepresented places, it opens up your perspective a lot about the history of a species." □

EVOLUTIONARY BIOLOGY

The rise of rideable horses

Early horse riders selected a rare mutation in a single gene to enhance rideability

Laurent Frantz[1,2]

Although horses have played a transformative role in the recent evolution of humans, altering communication, transportation, farming, and warfare, their domestication occurred unexpectedly late, only around 4500 years ago—more than 5000 years after the domestication of pigs, goats, cattle, and sheep. On page 925 of this issue, Liu *et al.* (*1*) report using a combination of ancient DNA analysis and experimental work with mouse models to provide new evidence of the biological changes that transformed wild horses into rideable animals, unlocking immense possibilities for societal and cultural change.

That horses were domesticated so late does not mean that people did not recognize the potential offered by horselike species (equids) much earlier. Domestic donkeys were already an integral part of societies in Africa (including Egypt) and the Middle East, serving as beasts of burden for more than 2000 years before the widespread domestication of the horse (*2*). In Mesopotamia, the lack of horses led elites to breed kungas—a hybrid of a domestic donkey and a wild Syrian ass (*Equus hemionus hemippus*, or hemippe) (*3*)—around 4500 years ago. Although these animals potentially offered advantages over donkeys for warfare, they were likely less controllable, and breeding required the constant capture of wild hemippes, which restricted their use to elites.

Mounted archers could shoot while riding horses. This came to define warfare throughout Central Asia, as depicted in this tablet from 13th-century Mongolia, China.

Liu *et al.* report that the wild ancestor of modern horses, which lived in the Eurasian Steppe (the region spanning from Eastern Europe to Mongolia), had a key advantage for domestication not present in other wild equids: A few individuals had key mutations that made them more docile and—perhaps more notably—altered their morphology and endurance to make them more rideable. The study shows that mice genetically modified to carry a variant of the *GSDMC* gene (which encodes the protein gasdermin C) found in these steppe horses displayed altered spinal anatomy, superior motor coordination, and enhanced forelimb strength—traits likely to make horses better suited for riding. Gasdermin C is a pore-forming protein that plays roles in multicellular organism development and immune responses.

Although the mutations identified by Liu *et al.* were already present in the wild ancestor of modern horses centuries before rideable horses spread across the world, their potential remained untapped for generations. This was possibly because the earliest horse herders from Central Asia (present-day Kazakhstan) around 5500 years ago used horses primarily for meat and milk (*4*), unlike the societies of Bronze Age Mesopotamia that sought animals for transport and warfare. The rise of the rideable horse required two elements: the presence of rare mutations that conferred suitability for riding, including a key variant in the *GSDMC* gene causing structural back changes and increased endurance, and the emergence of a human community that desired horses for warfare and transport. The right conditions for the rise of the rideable horse materialized ~3500 years ago in the Eurasian Steppe, north of the Caspian Sea in the lower Volga-Don (present-day Russia) (*5*).

Liu *et al.* demonstrate that once people actively pursued riding, the process accelerated markedly. The specific *GSDMC* variant linked to rideability surged from a 1% frequency to almost 100% in the horse population within just a few hundred years as a result of selective breeding (also known as artificial selection). By tracking the increase in frequency of the *GSDMC* variant over time using ancient genomes from early domestic horses, Liu *et al.* estimated its selection coefficient at around 20%. This indicates that horses with this mutation had a notable 20% increase in reproductive success compared with other horses that did not have the mutation. This selection pressure is almost unprecedented in evolution because reproductive success is rarely so heavily influenced by a single trait.

For instance, one of the most strongly selected genetic variations in human evolutionary history, a mutation enabling lactase persistence that allows adults to digest milk, has a selection coefficient of only 2 to 6% (*6*). This is despite the large survival advantage that it offered to human populations once milk consumption became widespread after the domestication of cattle, sheep, and goats.

Horses with the *GSDMC* mutation were spread rapidly to Bronze Age communities across Eurasia from the domestication center north of the Caspian Sea, reaching the farthest corner of Eurasia by ~3000 years ago (*5*). Early iconographic evidence indicates that riding, rather than use in chariots, was the primary reason that people first adopted horses across Eurasia (*5*). The advantage conferred by this mutation was so profound that, aside from donkeys, these new horses replaced nearly all other forms of managed equids in these regions (*7*).

As the confluence of events necessary for domesticating horses begins to be untangled, it is possible to speculate how, with a little less luck, the history of humans might have been radically different. Mutations transmitted to the next generation are inherently rare events. They happen in germline cells (sperm or eggs), stemming from mistakes made by enzymes (DNA polymerase) during DNA copying or from the failure of other enzymes to repair environmentally induced DNA damage. For example, data from genome sequencing of mother-father-offspring trios suggest that humans transmit only ~150 mutations to children each generation, despite the genome containing almost 3 billion nucleotides (*8*).

Furthermore, such rare mutations are often lost to genetic drift before selection can really put them to the test. If an individual carrying the mutation dies without passing it on, the variant is lost forever. This *GSDMC* gene mutation, therefore, had to occur in the germ line and persist within horse populations for hundreds or perhaps thousands of years before it encountered humans who actively desired horses for riding. The selection applied by these herders had to be strong enough to ensure that the mutation did not disappear by chance. This type of intense artificial selection requires sophisticated logistics (such as corrals) and a system to evaluate the breeding value of individual horses, which represents the expected performance of their offspring. Such sophisticated breeding schemes were previously thought to be a much later development, emerging only in the 19th century.

Although the precise circumstances and the cultural identity of the people responsible for this early, intensive breeding remain a mystery, they must have had the necessary ingenuity, technology, and foresight. Future archaeological work in the region north of the Caspian Sea will help identify which culture was likely responsible for this achievement and provide a better understanding of the circumstances of horse domestication. What is certain is that these first riders kick-started a revolution that changed the world, demonstrating how the immense currents of history can turn on the smallest of biological changes. □

REFERENCES AND NOTES

1. X. Liu *et al.*, *Science* **389**, 925 (2025).
2. E. T. Todd *et al.*, *Science* **377**, 1172 (2022).
3. E. A. Bennett *et al.*, *Sci. Adv.* **8**, eabm0218 (2022).
4. A. K. Outram *et al.*, *Science* **323**, 1332 (2009).
5. P. Librado *et al.*, *Nature* **598**, 634 (2021).
6. R. P. Evershed *et al.*, *Nature* **608**, 336 (2022).
7. J. Lira Garrido *et al.*, *Nat. Commun.* **16**, 7098 (2025).
8. D. Porubsky *et al.*, *Nature* **643**, 427 (2025).

10.1126/science.aea6151

[1]Palaeogenomics Group, Institute of Palaeoanatomy, Domestication Research and the History of Veterinary Medicine, Ludwig-Maximilians-Universität München, Munich, Germany. [2]School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK. Email: laurent.frantz@lmu.de

NEUROSCIENCE

Mapping the anatomy of placebo analgesia

The identification of somatotopy in brainstem pain modulatory pathways could help treat chronic pain

Massieh Moayedi[1,2,3] and **Lauren Y. Atlas**[4,5,6]

Placebo analgesia refers to the reduction of pain in response to an inert treatment and is produced through psychosocial factors, including learning, verbal instruction, and social context. Placebo analgesia has been extensively studied to understand the processes that modulate pain in humans (*1*, *2*), and neural circuits involved in placebo-related learning have recently been identified in preclinical models (*3*). A key unanswered question is whether placebo analgesia follows a somatotopic organization—that is, can placebo analgesia target specific spinal segments, or does it engage widespread pain control regardless of location? On page 893 of this issue, Crawford *et al.* (*4*) report that key brainstem regions involved in pain modulation have somatotopic organization. The findings could inform the selection of targets for brain stimulation to treat intractable chronic pain.

The perception of pain can be modulated by psychological factors, including attention and expectation (*5*). This psychological modulation can engage the body's pain modulatory mechanisms, including the release of endogenous opioids from brainstem neurons that can block nociceptive signals coming from the periphery (*6*). The discovery of neural pathways that project from the brainstem to the dorsal horn of the spinal cord—where touch, pain, and temperature signals from the body are transmitted to the brain—provided a structural substrate for this descending modulation. Specifically, projections from the periaqueductal gray (PAG) to the dorsal horn via the rostral ventromedial medulla (RVM) can either amplify or reduce the transmission of nociceptive information to the brain (*7*). The PAG receives top-down inputs from diverse cortical and subcortical brain regions as well as ascending, somatotopically organized inputs from the spinal cord and the trigeminal system, which carries nociceptive and thermal information from the face.

Consistent with descending modulation, electrical stimulation of the PAG leads to analgesia in rodents and humans (*7*), which is dependent on the release of endogenous opioids in the RVM (*7*). The RVM has dense reciprocal connections with the PAG, receives input from other brain regions involved in nociception and pain modulation, and is considered the final common relay in descending pain modulatory circuits. Descending modulation has been linked to placebo analgesia in humans; brain imaging studies indicate that placebo administration can modulate nociceptive activation in the spinal cord (*8*), and placebo-related increases in PAG and RVM activation are abolished by the opioid receptor antagonist naloxone (*9*). The PAG comprises four bilateral columns: the lateral, ventrolateral, dorsolateral, and dorsal. Spinal inputs encoding cutaneous pain project to the caudal lateral PAG, whereas trigemi-

nal inputs reach the rostral lateral PAG (*10*). However, whether somatotopy is preserved in PAG-mediated descending modulation is less clear. In other words, can PAG-driven analgesia target specific body regions, or is there systemic release of neuropeptides (i.e., endogenous opioids) that reduces pain throughout the body? Notably, different pain modulatory paradigms mediated by the PAG-RVM circuit can have analgesic effects in distant body regions, which suggests that there are generalized analgesic effects.

Crawford *et al.* used ultrahigh-field functional magnetic resonance imaging (fMRI) to detect changes in neural activity in subregions of the PAG and RVM during placebo analgesia in humans. They used a common placebo paradigm that induces expectations for pain relief in response to an inert cream. During a test phase, heat was applied to the placebo-treated skin site and to a site with a control cream, and differences in pain between the two sites provided a measure of the placebo effect. The authors applied this paradigm to the foot, the arm, and/or the face and imaged the brain during the placebo test phase. They observed reduced activation in the lateral PAG in individuals who reported less pain with placebo (placebo responders) and found that these PAG changes correlated with the magnitude of placebo analgesia. These effects were anatomically distinct, such that placebo effects on the face reduced activity in more rostral parts of the lateral PAG, whereas placebo effects on the arm and leg were associated with reduced activation in the caudal parts of the lateral PAG.

Descending modulation has been linked to placebo analgesia in humans...

By contrast, previous, less anatomically precise studies reported that placebo is associated with increases in PAG activation and connectivity (*9*). This could suggest that the findings of Crawford *et al.* in the lateral PAG reflect the somatotopically organized ascending inputs to the PAG rather than descending modulation. If placebo administration reduces pain through systemic descending modulation that is not somatotopically targeted, an overall reduction of ascending nociceptive input would be expected. Therefore, somatotopic reductions in PAG activation could result from the attenuation of somatotopic ascending input. However, the authors observed somatotopically organized increases in activity in the RVM associated with placebo analgesia. Given the key role of the RVM in descending modulation, these placebo-induced increases provide stronger support for somatotopically organized placebo analgesia than the findings in the PAG. In most prior studies of placebo analgesia the lateral PAG could not be clearly distinguished from the ventrolateral PAG, which is linked to opioid-induced analgesia and nocifensive behaviors. Although Crawford *et al.* had the resolution to resolve these nuclei, the location of placebo-related changes in activity was inconsistent across analyses. Understanding the link between the two regions will be important, as a recent study reported a role for opioid-releasing neurons in the ventrolateral PAG in a rodent model of placebo analgesia (*3*).

Although site specificity of placebo analgesia was originally thought to be an argument against global mechanisms, such as reductions in anxiety or descending modulation triggered by systemic opioid release (*11*), one study showed that opioid antagonists can have site-specific effects (*12*), which suggests that opioid-mediated placebo analgesia can be targeted rather than systemic. Such anatomical specificity could result from the somatotopy observed by Crawford *et al.* The authors manipulated whether individuals were tested on a single anatomical location or multiple locations, which provides an opportunity to formally test site-specific versus systemic effects. Participants who reported a placebo effect on the face also reported reductions in pain when a treatment (either placebo or control) was applied to the arm, which suggests a systemic modulatory effect. However, there was large variability in placebo effects across individuals, and thus the study was likely underpowered to examine specificity of placebo effects. Future studies should combine high-resolution imaging with tests of site-specific versus systemic placebo analgesia. It will also be important to investigate whether and how PAG-mediated pain modulation can "select" whether to be generalized or site specific, given that previous studies indicate that placebo effects can transfer across not only bodily sites but across modalities, such as between aversive sound and negative emotion (*13*).

The findings of Crawford *et al.* have implications for brain stimulation–based pain management paradigms. Deep brain stimulation (DBS) is a neurosurgical technique where fine electrodes are placed in brain structures to treat neurological diseases. In cases of intractable chronic pain—where all other treatments have failed—there are several potential brain targets for DBS, including the PAG. The analgesic effects of PAG stimulation vary substantially among individuals (*14*), and most studies have targeted the ventrolateral PAG. Given that target precision is considered paramount in the success of DBS, future studies of the utility of this approach for the treatment of chronic pain must identify the somatotopic representation of the targeted body region and ideally differentiate between ascending and descending inputs. However, before this can occur, whether the PAG and RVM somatotopic maps identified by Crawford *et al.* are maintained in chronic pain states needs to be determined. Evidence from animal studies suggests that the PAG-RVM–mediated descending modulation shifts from inhibition to pain facilitation in chronic pain states (*15*). Therefore, individualized mapping of PAG and/or RVM somatotopy may be required for optimal targeting of DBS. □

REFERENCES AND NOTES

1. L. Y. Atlas, T. D. Wager, in *Placebo*, F. Benedetti, P. Enck, E. Frisaldi, M. Schedlowski, Eds., vol. 225 of *Handbook of Experimental Pharmacology* (Springer, 2014), pp. 37–69.
2. M. Zunhammer *et al.*, *Nat. Commun.* **12**, 1391 (2021).
3. H. Neyama *et al.*, *Sci. Adv.* **11**, eadp8494 (2025).
4. L. S. Crawford *et al.*, *Science* **389**, eadu8846 (2025).
5. L. Y. Atlas, *Trends Cogn. Sci.* **25**, 992 (2021).
6. J. D. Levine, N. C. Gordon, H. L. Fields, *Lancet* **2**, 654 (1978).
7. H. Fields, *Nat. Rev. Neurosci.* **5**, 565 (2004).
8. F. Eippert, J. Finsterbusch, U. Bingel, C. Büchel, *Science* **326**, 404 (2009).
9. F. Eippert *et al.*, *Neuron* **63**, 533 (2009).
10. A. Depaulis, R. Bandler, *The Midbrain Periaqueductal Gray Matter: Functional, Anatomical, and Neurochemical Organization*, vol. 213 of *NATO ASI Series* (Springer, 1991).
11. G. Montgomery, I. Kirsch, *Psychol. Sci.* **7**, 174 (1996).
12. F. Benedetti, C. Arduino, M. Amanzio, *J. Neurosci.* **19**, 3639 (1999).
13. Y. Zhao *et al.*, *J. Ment. Health* **24**, 230 (2015).
14. N. Shaheen *et al.*, *Front. Hum. Neurosci.* **17**, 1297894 (2023).
15. R. Kuner, *Nat. Med.* **16**, 1258 (2010).

ACKNOWLEDGMENTS

M.M. holds a Canada Research Chair in Pain Neuroimaging and acknowledges support by the Bertha Rosenstadt Endowment at the Faculty of Dentistry, University of Toronto, and a University of Toronto Centre for the Study of Pain Scientist Award. L.Y.A. acknowledges support by the Intramural Research Program of the National Institutes of Health (NIH) (ZIA-AT000030). The views here are those of the authors and do not necessarily reflect the views of the NIH or the US Department of Health and Human Services.

10.1126/science.aea2370

[1]Centre for Multimodal Sensorimotor and Pain Research, Faculty of Dentistry, University of Toronto, Toronto, ON, Canada. [2]University of Toronto Centre for the Study of Pain, Toronto, ON, Canada. [3]Krembil Research Institute, University Health Network, Toronto, ON, Canada. [4]National Center for Complementary and Integrative Health, National Institutes of Health, Bethesda, MD, USA. [5]National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA. [6]National Institute on Drug Abuse, National Institutes of Health, Baltimore, MD, USA. Email: m.moayedi@utoronto.ca; lauren.atlas@nih.gov

TEXTILE ENGINEERING

Sustainable personal cooling in a warming world

Advanced textiles and intelligent wearable devices can provide cooling under extreme heat

Dahua Shou and **Ziqi Li**

Between 2000 and 2019, heat-related causes claimed more than 480,000 lives annually worldwide (*1*). On average, air conditioning systems account for 12.3% of total greenhouse gas emissions in the United States (*2*), contributing to a feedback loop that amplifies global warming: Cooling technologies increase global temperatures, which in turn leads to greater demand for cooling. Growing global heat waves are pushing conventional clothing to its limits, fueling the demand for personal cooling solutions. Sustainable cooling wearables, such as breathable textiles that are derived from recycled plastic bottles (*3*) and solar-powered active cooling gears (*4*), have been developed to meet the emerging need. These technologies can provide protections against extreme heat while reducing reliance on air conditioning, which often consumes considerable energy.

Cooling is achieved with four primary mechanisms: radiation, conduction, convection, and evaporation. Natural personal cooling processes, such as sweating, rely on evaporative mechanisms. For example, as water evaporates from the skin, it takes heat away from the body and lowers the body temperature. By contrast, engineered materials manipulate radiation, conduction, and convection to control temperature.

Spectrum-selective materials regulate transmission, reflection, and absorption of electromagnetic radiation, such as light and heat, at specific wavelengths. These textiles can block radiative heat from the external environment while dissipating the body heat effectively. For example, a nanoporous polyethylene textile that is transparent to mid-infrared radiation emitted by the human body (7 to 14 μm) effectively cools artificial skin by 2.0°C more as compared with cotton (*2*). A multilayered metafabric, which is an engineered textile for regulating spectral properties, can scatter and reflect solar radiation across a broad spectrum that spans from ultraviolet to near-infrared wavelengths (0.3 to 2.5 μm) by using randomly dispersed particles with optimized size and refractive index (*5*). These textiles radiate heat within the atmospheric transparent window—a range of wavelength (8 to 13 μm) that can escape the atmosphere—to the cold outer space, which acts as a giant heat sink. However, they can absorb urban heat across a broad spectrum (2.5 to 20 μm), which is emitted from the ground and nearby buildings. A hierarchical fabric has been developed to emit mid-infrared radiation from the body while selectively minimizing urban heat absorption (*6*). Although these advances are promising, radiative cooling is limited by environmental factors such as moisture and pollution, which can disrupt radiation pathways and retain heat. Furthermore, radiant heat transfer, which is governed by the Stefan-Boltzmann Law, increases exponentially with temperature. Thus, radiative cooling requires a sufficient temperature difference between the body and the environment, which is often limited in indoor settings. Integrating other mechanisms such as convection and evaporation could enhance the cooling performance in practical conditions.

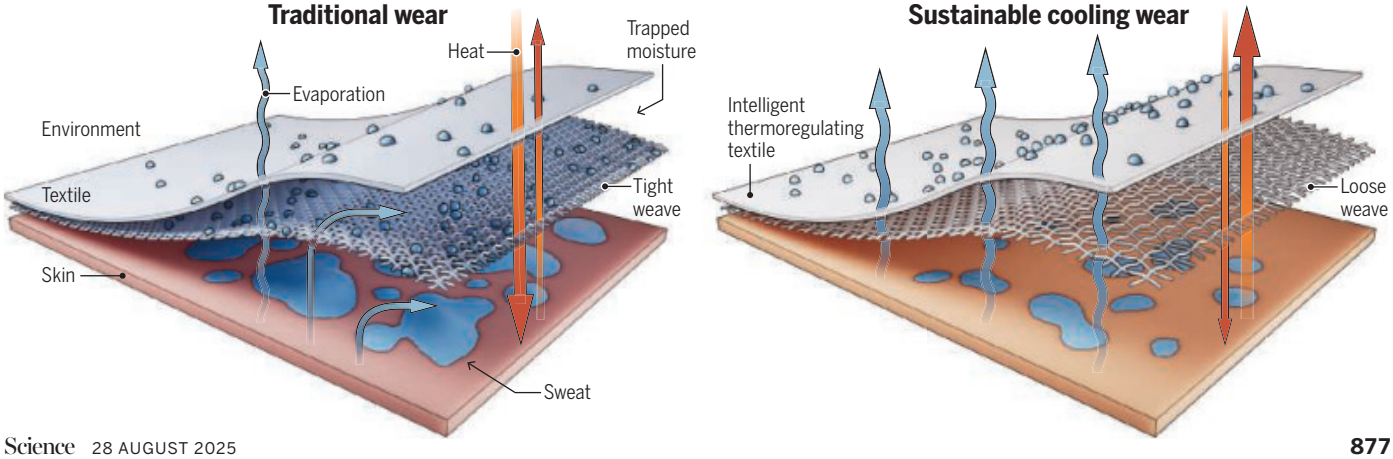
Conduction transfers heat through atomic vibrations, molecular collisions, or electron transport. Consequently, it governs how textiles are perceived as cool or warm upon contact with skin, depending on the heat capacity (the energy required to raise temperature to a specific amount) and thermal conductivity of the textile. Fillers that have high thermal conductivity such as boron nitride particles are often added to textiles to enhance conductive cooling (*7*). By contrast, exceptional thermal insulators (such as still air) could be used to minimize heat transfer from extremely hot environments (*8*). Knittable aerogel-encapsulated fibers, which are inspired by hollow polar bear hair, have demonstrated superior thermal insulation with a high fraction of embedded air (*9*).

Convection, which is driven by air or fluid flow, dissipates body heat through ventilation. Porous, breathable fabrics with strategically placed flaps, made from moisture-responsive fibers that adapt to humidity changes, improve thermoregulation by promoting warm air escape and cool air intake (*10*). Convection can also enhance the evaporative cooling mechanism by promoting efficient evaporation of sweat on the skin. Optimizing wettability, fiber shape, and pore size of the fabric can direct excess sweat away from the skin, improve evaporation, and prevent sweat accumulation for improved breathability and cooling performance (*11*).

Integrating advanced textiles and intelligent wearable technologies could bring controllable cooling systems that dynamically adapt to changing environments. Although commercial products such as fan-

Wear smart, stay cool

Traditional wearables struggle to release heat and moisture efficiently. These fabrics quickly become drenched in sweat under extreme heat. Integrating advanced textiles with hierarchical structures and smart wearable devices could provide sustainable cooling of the body. These engineered materials can self-regulate heat and discharge sweat by dynamically responding to the changes in atmospheric temperature and humidity.



GRAPHIC: N. BURGESS/*SCIENCE*

equipped clothing and liquid-cooling vests, optionally with pocketed phase change materials, have demonstrated rapid cooling capabilities. However, they are often constrained by excessive weight and energy consumption. Alternatively, a lightweight wearable variable-emittance device with adjustable emissivity and a kirigami-enabled design (three-dimensional structure from cutting and folding of a flat material) dynamically regulates heat transfer, expanding the thermal comfort zone (conditions of temperature, humidity, and air circulation in which the body feels comfortable) by 4.9°C (*12*). Furthermore, wearable flexible organic photovoltaic modules have been paired with bidirectional electrocaloric devices that heat or cool by applying or removing an electric field to achieve thermoregulating clothes with an extended thermal comfort zone (12.5° to 37.6°C) (*4*). Despite these advancements, many wearable devices still face inherent limitations such as lack of flexibility, short battery life, and wiring complexity.

Conventional personal cooling technologies perform well under static conditions but often encounter challenges in dynamic environments. Developing responsive textiles and adaptive wearable devices is key to addressing this issue. The ability to autonomously regulate fiber diameter and shape, porosity, thickness, and thermal properties (such as conductivity, emissivity, and reflectivity) in response to temperature and humidity changes or through external stimuli such as electrical field is highly desired (see the figure). These functionalities could bring new cooling strategies such as mode-selective thermoregulation that modulates the predominant cooling mechanisms under specific temperature and humidity ranges. For example, evaporative cooling is more efficient in low-humidity and high-temperature environments, whereas convective cooling is more favorable under high-humidity conditions. Additionally, the concept of diode-like heat transfer—heat can exit but cannot enter—could bring efficient directed thermoregulation in harsh environments with extreme temperatures.

Artificial intelligence can predict personal cooling needs by analyzing real-time physiological and psychological data from wearable sensors that monitor body temperature, sweat, heart rate, humidity, and stress level. These intelligent systems can activate personalized cooling proactively while preventing overcooling by responding only to detected discomfort. In addition, integrating textile-based sensors, fiber-based coolers, and wearable flexible energy storage systems enables active cooling by harvesting energy from the body and the surrounding environment; converting mechanical, thermal, or moisture energy into electricity through use of triboelectric, piezoelectric, thermoelectric, and moisture-electric generators; as well as potentially using nutrients from sweat and blood. For example, a smart fabric embedded with nanoparticles exhibits high solar radiation reflectance and infrared emissivity, effectively enabling thermoelectric generation for self-sustainable cooling (*13*). Artificial intelligence, such as machine learning, can also be used to design all-weather cooling technologies by optimizing the four key cooling mechanisms. This approach leverages multiobjective optimization to systematically explore parameters such as fabric thickness, structural architecture, fiber diameter, porosity, and material composition while dynamically adapting to environmental variations and human activities to achieve customized cooling performance.

Human-centered design is essential to balance performance, durability, and wearability. Poor textile design can cause excessive sweating, which compromises heat transfer efficiency by lowering near-infrared reflectivity and increasing heat absorption. Sweat-drenched clothing not only becomes heavy but also turns transparent to visible light, undermining its ability to block sunlight. Further, trapped moisture decreases fabric permeability, impairing convective and evaporative cooling. To overcome these challenges, skin-mimicking fabrics that can transfer sweat as liquid droplets in a specific direction could remove sweat more efficiently than evaporative drying (*11*). Another emerging technology is a washable wearable electroosmotic system, which actively drives sweat across a fabric by applying a low voltage. Such an approach can maintain the skin dry and breathable even under intense perspiration (*14*). Hierarchical nano- and microstructured fabrics have an optimized surface area that improves moisture-wicking and radiative heat management through increasing radiation emissivity and solar reflectivity (*15*). These advanced textiles and intelligent wearables hold great potential to enhance cooling efficiency while reducing excessive perspiration and energy expenditure.

Next-generation cooling technologies should prioritize sustainability. Textile production contributes 5 to 10% of global greenhouse gas emissions (*3*), with 57% of discarded clothing ending up in landfills and 25% of them being incinerated. Using a single type of material, instead of blending different fibers, could improve recyclability. Many cooling technologies also rely on surface coating and chemical treatments, which may pose environmental risks. Alternatively, natural fibers such as silk and wool with inherent hierarchical microstructures could be optimized for solar reflection and evaporative cooling. Future efforts should focus on sustainable yet cost-effective methods that are compatible with existing systems to enable rapid adoption.

However, certain challenges need to be addressed to enable widespread adoption of robust, sustainable personal cooling technologies. Integrating different cooling mechanisms and approaches across disciplines such as textile engineering, artificial intelligence, thermodynamics, flexible electronics, and materials science remain to be improved. Standardized metrics including cooling power, coefficient of performance (efficiency of cooling relative to energy used), acceptance of end users, and user-specific indicators (such as thermal sensation, body temperature, and metabolic rate) must be widely established.

Personal cooling technologies not only transform daily life but also address critical needs in extreme environments. For example, construction workers, agricultural laborers, and firefighters are constantly exposed to intense heat, for which efficient cooling solutions are essential to prevent heat stroke and maintain safety and productivity. Likewise, outdoor enthusiasts benefit from smart wearable devices such as chest straps or active cooling vests equipped with biosensors and adaptive thermoregulation technologies to intelligently regulate body temperature. Sustainable personal cooling systems provide more than just comfort; they enhance the health and performance of humans under a rapidly changing climate. □

REFERENCES AND NOTES

1. World Health Organization, "Heat and health" (2024); https://www.who.int/news-room/fact-sheets/detail/climate-change-heat-and-health.
2. P.-C. Hsu *et al*., *Science* **353**, 1019 (2016).
3. L. Zhang, M. Leung, S. Boriskina, X. Tao, *Nat. Sustain.* **6**, 243 (2022).
4. Z. Wang *et al*., *Science* **382**, 1291 (2023).
5. S. Zeng *et al*., *Science* **373**, 692 (2021).
6. R. Wu *et al*., *Science* **384**, 1203 (2024).
7. O. I. Kalaoglu-Altan, B. K. Kayaoglu, L. Trabzon, *iScience* **25**, 103825 (2022).
8. X. Zhang *et al*., *Adv. Sci.* **11**, e2309605 (2024).
9. M. Wu *et al*., *Science* **382**, 1379 (2023).
10. X. Li *et al*., *Sci. Adv.* **7**, eabj7906 (2021).
11. L. Lao, D. Shou, Y. S. Wu, J. T. Fan, *Sci. Adv.* **6**, eaaz0013 (2020).
12. T. H. Chen *et al*., *Proc. Natl. Acad. Sci. U.S.A. Nexus* **2**, pgad165 (2023).
13. W. Jiang *et al*., *Device* **3**, 100564 (2025).
14. D. Shou, Y. Gu, U.S. patent serial no. 18/764,400 (2024).
15. D. Thiyagarajan, S. Li, B.-K. Lee, *Nano Lett.* **25**, 5609 (2025).

ACKNOWLEDGMENTS

The authors acknowledge support from the Research Grants Council of Hong Kong (grants PolyU152052/21E, PolyU152196/23E, PolyU152157/24E, and PolyU152103/25E), Innovation and Technology Fund of Hong Kong (grant ITS/288/22), and Endowed Young Scholar Scheme of the Hong Kong Polytechnic University (project 84CC).

10.1126/science.adt9536

Future Intelligent Wear Centre, School of Fashion and Textiles, the Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China. Email: dshou@polyu.edu.hk

Vegetation on meandering rivers alters the direction of sedimentation on the inner banks of river bends, causing them to expand outward.

PHOTO: SVEN ZACEK/NPL/MINDEN PICTURES

GEOLOGY

Vegetation steers a river's path

Evolution of plants could have modified ancient river trajectories

Jim Pizzuto

Vegetation in river valleys slows floodplain erosion rates (*1*), reduces river channel widths (*2*), and drives a transition from braided, multichannel rivers to single-channel, meandering rivers (*3*). These changes also control carbon sequestration and other biogeochemical processes along river channels (*4*). On page 915 of this issue, Hasson *et al.* (*5*) report a new way in which floodplain vegetation—the plant life found in flat land adjacent to a river—controls river dynamics. Meandering rivers in floodplains without vegetation tend to migrate downstream (parallel to the overall river flow direction), whereas meandering rivers with vegetated floodplains are likely to expand outward (across river valleys). The findings resolve conflicting interpretations of early Paleozoic (>419 million years ago) river deposits, which were formed before the evolution of land plants, and challenge current understanding of river meandering on Earth and other planetary bodies.

As a river flows, sand and gravel on the streambed are carried downstream. These sediments, which are collected in bars (elevated areas of sediment) and other deposits, move around and create features that record the directions of a river's flow. River deposits transform into sedimentary rocks that serve as preserved indicators of flow direction and character of ancient rivers. Braided, multichannel rivers tend to flow straight down their valleys. By contrast, winding, single-channel meandering rivers have variable flow directions that reflect their sinuous courses (see the photo). Analyses of ancient rocks have demonstrated that Paleozoic rivers without vegetated floodplains primarily ran straight downstream (*6*). These studies, coupled with observations of vegetation inhibiting the formation of multichanneled braided rivers (*3*), have led to a proposal that single-channel meandering rivers only became widespread on Earth after land plants developed in the Paleozoic Era.

This interpretation has been challenged by the discovery of well-developed meanders (or bends) in unvegetated arid regions on Earth, and the presence of unvegetated meandering channels on Mars and Titan. If vegetation is not a prerequisite for river meandering, meandering channels should have existed throughout Earth's history. How, then, can the uniform downstream flow directions of pre-Paleozoic rivers be explained?

Hasson *et al.* analyzed satellite images to document the migration directions of modern vegetated and unvegetated meandering rivers, which are not primarily influenced by climate, the variable erosion resistance of floodplain deposits, or the total amount of sediment carried by rivers. The authors hypothesized that vegetation modifies key river migration processes. Meanders migrate by eroding the outer banks and depositing sediments on inner banks that form point bars. When these two processes balance each other, meanders maintain a constant width as they migrate. Sediment accumulation on point bars (point bar accretion) is accelerated when near-bed river currents are deflected toward the inner bank, rather than following the downstream direction. Hasson *et al.* suggest that vegetation modifies the direction of point bar accretion, which causes river bends to mi-

grate across valley rather than downstream. Vegetation also stabilizes channels as a river migrates, inhibiting bend abandonment by avulsion (an abrupt jump of a river to a new location) (*7*) or bend cutoff (a sudden straightening of a bend by erosion). The authors further argue that preserved indicators of near-bed flow toward point bars, which is a diagnostic feature of meandering river deposits, has been overlooked in past studies. This resulted in overestimating the prevalence of braided rivers with straight paths on ancient Earth.

Although the hypotheses of Hasson *et al*. require additional testing, they can inspire further study to clarify the underlying mechanism of meander migration, which has been a topic of active debate among geomorphologists for decades. Some argue that meander migration first requires erosion of the outer bank, followed by point bar deposition (*8*). This so-called "bank pull" hypothesis is often countered by the "bar push" hypothesis in which point bar deposition causes erosion by "pushing" the river flow towards the outer bank (*9*). Others suggest that both mechanisms work jointly, with neither bank erosion nor point bar deposition in preeminent control of meander migration (*10*). However, if vegetation plays a central role in controlling the direction of point bar advance, then the bar push mechanism must be predominant.

The study of Hasson *et al*. also challenges existing mathematical theories of river meandering (*11*). These theories offer physics-based explanations that underlie the occurrence, characteristic forms, and temporal development of meanders by combining principles from geomorphology and fluid mechanics. Although some progress has been made (*12*), most models do not explicitly represent inner bank deposition processes. Thus, existing theories cannot explain the observations of Hasson *et al*. Incorporating inner bank deposition processes and vegetation into the mathematical models will require collaborations between ecologists, botanists, and theoreticians.

Improved understanding of vegetated and unvegetated meandering rivers should catalyze scientific advancement. For example, it could reveal river dynamics and biogeochemical cycling during the early history of Earth and of understudied rivers of arid regions. Beyond Earth, reconstructions of meandering water flows on Mars (*13*) and meandering methane flows on Titan (*14*) may be inferred. The growing field of river restoration and ecological engineering could also benefit because restoring rivers often involves planting vegetation around artificially constructed river meanders (*15*). The detailed interactions between river meanders and vegetation could help improve restoration designs to create self-sustaining environments with improved water quality and a flourishing animal and plant life surrounding a river. □

REFERENCES AND NOTES

1. A. Ielpi, M. G. A. Lapôtre, *Nat. Geosci.* **13**, 82 (2020).
2. R. D. Hey, C. R. Thorne, *J. Hydraul. Eng.* **112**, 671 (1986).
3. M. Tal, C. Paola, *Water Resour. Res.* **35**, 1014 (2010).
4. M. Repasch *et al*., *Nat. Geosci.* **14**, 842 (2021).
5. M. Hasson, A. Finotello, A. Ielpi, M. G. A. Lapôtre, *Science* **389**, 915 (2025).
6. M. R. Gibling *et al*., *Proc. Geol. Assoc.* **125**, 524 (2014).
7. J. H. Gearon *et al*., *Nature* **634**, 91 (2024).
8. W. I. van de Lageweg, W. M. van Dijk, A. W. Baar, J. Rutten, M. G. Kleinhans, *Geology* **42**, 319 (2014).
9. J. A. Constantine, T. Dunne, J. Ahmed, C. Legleiter, E. D. Lazarus, *Nat. Geosci.* **7**, 899 (2014).
10. J. Mason, D. Mohrig, *Geology* **47**, 1136 (2019).
11. G. Seminara, *J. Fluid Mech.* **554**, 271 (2006).
12. E. Eke, G. Parker, Y. Shimizu, *J. Geophys. Res. Earth Surf.* **119**, 1455 (2014).
13. M. G. A. Lapôtre, A. Ielpi, *AGU Adv.* **1**, e2019AV000141 (2020).
14. M. Malaska *et al*., *42nd Lunar Planet. Sci. Conf.* **1608**, 1562 (2011).
15. I. Costaz-Puyou, R. Williams, A. Black, C. Spray, C. MacDonell, *J. Environ. Manage.* **388**, 125672 (2025).

Published online 21 August 2025
10.1126/science.aea2382

Department of Earth Sciences, University of Delaware, Newark, DE, USA. Email: pizzuto@udel.edu

GENETICS

Penetrance and variant consequences—Two sides of the same coin?

To get more out of genome sequences, the effects of variants need to be quantified

Harper Raiken and **Amelie Stein**

Over the past few decades, sequencing a human genome has gone from a moonshot project to standard clinical practice, revealing thousands of variants in each individual. A crucial question in linking genome sequences to health outcomes is, which gene variants affect the risk of getting a particular disease, and if so, how many of the carriers go on to develop the disease? For example, a patient that carries a particular variant may benefit from more frequent screening for diabetes and breast cancer. In many such cases, early detection of disease onset can greatly improve overall outcomes for the patient. On page 894 of this issue, Forrest *et al*. (*1*) describe the integration of large-scale phenotype data, genetic sequence data, and artificial intelligence for approximating disease risk using machine-learning methods. This framework provides precedent for the synthesis of clinical and genetic data for personalized treatment over a variety of diseases.

A particularly useful metric for determining disease risk on a genetic level is penetrance. Penetrance represents the proportion of individuals carrying a particular variant who also show clinical signs of a particular disorder, typically expressed as a percentage (*2*). For example, many variants of the breast cancer gene 1 (*BRCA1*) gene are highly correlated with breast cancer incidence. Studies have illustrated that variant carriers on average have a 57% risk of developing the disease (*3*). Penetrance for individual variants is difficult to accurately assess because of small numbers of known carriers. It has traditionally been assigned depending on observed disease cases versus known cases in family clusters, which is likely to overestimate penetrance. More recently, large genome sequencing studies have enabled population-based penetrance estimates, which may underestimate owing to study recruitment biases (*4*). More accurate penetrance estimates are useful to clinicians to support the use of more frequent screenings (such as mammograms) and/or preventive treatments such as prophylactic surgery.

Forrest *et al*. created machine-learning models for 10 selected diseases that incorporate patient electronic health records to estimate disease risk scores on the basis of patient data from more than 1 million individuals and then compared these scores for carriers of different variants. Machine-learning penetrance aims to be a holistic single metric that integrates a variety of clinical factors that contribute to disease risk, captured in a value between 0 and 1. The authors purport that for a given individual, the relevant disease model can provide improved risk assessment that is based on their patient data to inform additional screening measures. They also suggest

that on a broader population scale, electronic health records can be combined with genetic sequencing data to detect variants with high risk of disease incidence.

For each of the 10 diseases, a gradient-boosted, tree-based machine-learning model was trained on curated cases and controls with both clinical and genetic data. Specifically, each model was trained and internally validated on clinical data from the Mount Sinai Health System then tested on data from BioMe, a biobank at Mount Sinai focused on cultivating patient diversity. The BioMe data was used to both generate disease scores and assess penetrance per variant. Analysis of the weight assigned to the 40-plus different features revealed that some models balance multiple features well, whereas others are dominated by single aspects such as age. Thus, careful consideration of the individual model may help to assess how much additional insight it can be expected to contribute. Machine-learning penetrance assessment was carried out for around 1600 variants in 31 genes, with established associations to the 10 selected diseases. Scores tended to agree with available data from conventional penetrance assessment but were less bimodally distributed, a challenge with conventional penetrance for rare variants that this more quantitative approach alleviates. Further, machine-learning penetrance scores also show reasonable agreement with experimental variant consequence data where available.

Accurate penetrance assessment is important for patient stratification and to optimize the use of screening resources.

Accurate penetrance assessment is important for patient stratification and to optimize the use of screening resources. Approaching the problem "top down," from observed phenotypes and clinical markers, offers a good chance of integrating all relevant factors, known and unknown. Conversely, for understanding the mechanistic underpinnings of how a given variant causes disease—and avenues toward treatment—starting "bottom up" from the individual components may be more promising (*5*). In the field of protein variant classification, a combination of oligonucleotide synthesis and sequencing techniques have enabled systematic screens of all possible single variants for dozens of proteins. A plethora of variant consequence predictors has been developed, partly fueled by the machine-learning revolution, building on large language models trained to capture evolutionary tolerance (essentially capturing which amino acids are tolerated at which positions), some further boosted by integration of high-resolution protein structures (*6*). One of the main findings across proteins has been that loss of cellular stability is a common molecular cause of observed loss of function and may account for about half of the pathogenic variants (*7*, *8*). For example, this is the underlying cause for several pathogenic variants in the cystic fibrosis transmembrane receptor, for which stabilizing molecules have been developed that largely alleviate the disease phenotype (*9*).

Conceptually, both variant consequences and penetrance aim to capture the chances of a variant having detrimental outcomes but start from opposite ends of the scale: one based on models of biomolecules, the other based on observations of whole organisms. Both aim to quantify and go beyond the traditional binary distinction of variants as pathogenic or benign. As the comparison by Forrest *et al*. of machine-learning penetrance and AlphaMissense (an artificial intelligence model that incorporates protein structure to predict the pathogenicity of genetic variants that change a single amino acid) scores illustrates, the correlation for 31 proteins examined is only moderate. This could be due to, for example, predicted effects from the buffering of the molecular level by regulatory layers. Thus, slightly lower activity of an enzyme might be compensated by increased expression. Likewise, lack of patient data could lead to misassessment of machine-learning penetrance, in particular because many of the variants in question are extremely rare.

Forrest *et al*. report that machine-learning penetrance agrees particularly well with prior data that show high penetrance, which may relate to high-penetrance variants being less dependent on other factors, especially those not captured by the model. A similar observation has been made for protein variant consequences, in which accuracy for prediction of pathogenic and benign variants often is substantially higher than for datasets that systematically profile all single-point variants in a protein. This is likely due to clinically classified pathogenic and benign variants being among the most clear-cut cases (*6*).

Both accuracy and mechanistic insight may benefit from combining the top-down and bottom-up approaches. Recent years have seen the integration of systematic protein variant screening data into clinical variant assessment (*10*). The least understood—and also currently least data-rich—layers are those in the middle, cells and tissues, where emergent properties arise from interactions between biomolecules, potentially affected by variants in unexpected ways (*11*–*13*). Pharmacogenomics approaches may offer further insights into treatment options by mapping drug efficacy for different variants.

On the other end of the scale from molecules to humans, another potential challenge pertains to the diversity of the patient dataset, which as Forrest *et al*. note is limited in its breadth of various minority demographics—in particular, Asian and Hispanic individuals (6.4 and 2.1% of the training data, respectively). This can result in health care inequality for underrepresented individuals because they may experience misinformed care when benign mutations are incorrectly predicted to cause disease. Genome-wide association studies are faced with the same issue, and differences in variant impact between ethnic groups have been illustrated for people of African descent in America and native populations (*14*, *15*). Ideally, this weakness will be overcome in time with the active, continual incorporation of more diverse training data from underrepresented demographics to ensure that accurate disease risk assessment will be equally helpful to all groups of people. □

REFERENCES AND NOTES

1. I. S. Forrest *et al*., *Science* **389**, eadm7066, (2025).
2. D. N. Cooper *et al*., *Hum. Genet.* **132**, 1077 (2013).
3. S. Chen, G. Parmigiani, *J. Clin. Oncol.* **25**, 1329 (2007).
4. C. F. Wright *et al*., *Nat. Genet.* **56**, 1772 (2024).
5. A. Stein *et al*., *Trends Biochem. Sci.* **44**, 575 (2019).
6. L. Gerasimavicius, S. A. Teichmann, J. A. Marsh, *Curr. Opin. Struct. Biol.* **92**, 103023 (2025).
7. M. H. Høie, M. Cagiada, A. H. Beck Frederiksen, A. Stein, K. Lindorff-Larsen, *Cell Rep.* **38**, 110207 (2022).
8. A. Beltran, X. 'er Jiang, Y. Shen, B. Lehner, *Nature* **637**, 885 (2025).
9. S. Pankow, C. Bamberger, D. Calzolari, *Nature* (2015).
10. A. E. McEwen, M. Tejura, S. Fayer, L. M. Starita, D. M. Fowler, *Nat. Rev. Genet.* 10.1038/s41576-025-00870-x (2025).
11. S. Larsen-Ledet, A. Panfilova, A. Stein, *PLOS Genet.* **21**, e1011681 (2025).
12. K. Jaganathan *et al*., *Cell* **176**, 535 (2019).
13. J. Schmit, K. A. Dill, *Annu. Rev. Biophys.* **52**, v (2023).
14. F. F. Stæger *et al*., *Nature* **639**, 404 (2025).
15. A. K. Manrai *et al*., *N. Engl. J. Med.* **375**, 655 (2016).

10.1126/science.aea0628

Department of Biology, University of Copenhagen, Copenhagen, Denmark.
Email: amelie.stein@bio.ku.dk

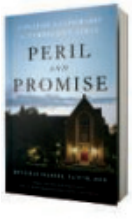
Students walk across Spelman College campus, where Beverly Daniel Tatum served as president from 2002 to 2015.

PHOTO: ELIJAH NOUVELAGE/BLOOMBERG VIA GETTY IMAGES

HIGHER EDUCATION

Academic leadership in uncertain times

A former college president offers advice for administrators navigating unprecedented challenges **Susan Elrod**



Beverly Daniel Tatum
Basic Books, 2025.
368 pp.

PERIL AND PROMISE | Higher education leaders in the United States are currently confronting some of the most difficult situations they have faced in decades, from dealing with revenue shortfalls because of declining enrollment to figuring out how to respond to the Trump administration's litany of demands. Leadership turnover is at an all-time high, and the tenures of executive leaders continue to shrink as administrators face difficult choices and no-win situations.

In her new book, *Peril and Promise: College Leadership in Turbulent Times*, Beverly Daniel Tatum—author of the seminal text *Why Are All the Black Kids Sitting Together in the Cafeteria?* (*1*)—offers a modern treatise on the current state of university leadership. Drawing on 45 years of scholarship and her experience as a professor, dean, acting college president, and college president, she chronicles the trials and triumphs of university leaders in the postpandemic era. Although the book was completed just as Donald Trump took office for the second time and therefore does not speak to the new challenges administrators have faced since that time, it covers in detail many of the issues that have drawn the president's ire, including diversity, equity, and inclusion (DEI) policies, campus protests, and free speech.

Tatum opens with an overview of her own leadership motivations and journey, describing her ascent from professor to administrator. Here, she attests "that wise, creative, and courageous leadership can mitigate even the most challenging problems of our time." The book is part memoir and part advice, offering insights and suggesting actions that academic leaders can take to mitigate such challenges. The ensuing chapters cover topics ranging from the role of deans, to shared governance, to managing risks, to the enrollment cliff, to college finances as well as other disruptions grappled with by those in higher education. She closes on a positive note, sharing reasons why university leadership is satisfying and can be filled with joy.

In the introduction, Tatum provides historical context regarding the demographics of students attending American higher education institutions, the diversity of institutional types and systems, and where students choose to attend college. I found this section particularly interesting, even as a leader in higher education myself—we rarely spend time thinking on such a big-picture level. It was nice to be reminded of the sheer number of degree-granting postsecondary institutions in the United States—3633 in the 2022–2023 school year—and the amazing diversity of institutional types, missions, and students who benefit from postsecondary opportunities.

The chapters that follow recount events that highlight Tatum's many career leadership challenges and accomplishments. In the chapter "Who's in Charge, Really?," for example, she describes a challenging hiring decision she faced early in her tenure as president of Spelman College and how she navigated it. The board chair in this situation did not agree with her decision to move forward with a hire and undermined her authority by suggesting that three board members help her "undo [her] 'hiring mistake.'" The incident, she recalls, represented an important test of her authority—the appointment proceeded, and she learned the political skills needed to navigate such conflicts.

Tatum also provides insights into how administrators might respond to concerns raised by faculty. Here, she describes the urgency she felt to make certain changes when she became the president of Spelman in 2002. Although the changes she instituted to improve cross-campus communication, internet infrastructure, and departmental leadership, among other efforts, appeared externally to be having a positive impact, her speedy actions were not met positively by a key constituency: the faculty. The lesson she learned in this situation is that leaders must engage all stakeholders to ensure that everyone is clear regarding goals, strategies, and decision-making processes.

The book will even be informative for those who are not experienced higher education leaders but wish to learn more about the nuances of college administration. For example, Tatum demystifies the role of governing boards in higher education—a much-misunderstood subject—writing that they are ultimately responsible for the health and vitality of institutions, delegating authority to the president of the college or university to ensure that institutions are meeting their missions. These boards are generally composed of unpaid volunteers, most of whom are not experts in higher education, she reveals, which means that an institution's president must ensure that they are communicating clearly and consistently to keep the board informed.

When I was invited to provide a review of this book, I was excited by the prospect that it might provide specific suggestions or recommendations to current college leaders. And, while Tatum's insights are embedded in the stories she tells, the book would have benefited from concise summaries of her advice at the end of each chapter so that leaders who find themselves in similar situations might more quickly find relevant sections. That said, I still found it to be an educational, insightful, and hopeful read. □

REFERENCES AND NOTES

1. B. D. Tatum, *Why Are All the Black Kids Sitting Together in the Cafeteria? And Other Conversations About Race* (Basic Books, 1997).

10.1126/science.aea3930

The reviewer is former chancellor and professor emerita, Indiana University South Bend, South Bend, IN, USA. Email: slelrod@iu.edu

SCIENCE AND SOCIETY

Mars mania and the making of scientific authority

Twentieth-century speculation about martian life shed light on the interplay of science and culture **Dov Greenbaum**

The Martians: The True Story of an Alien Craze That Captured Turn-of-the-Century America
David Baron
Liveright, 2025. 336 pp.

In *The Martians*, David Baron offers a detailed account of how socialites, inventors, and respectable astronomers became convinced that intelligent life thrived on Mars at the turn of the 20th century. Baron's engaging narrative examines science put on public trial and litigated through personalities, persuasion, and press rather than peer review. The result: public fascination that outpaced scientific fact, revealing more about earthlings than martians.

Set in the late 19th and early 20th centuries, the book spans four decades, tracing how Mars transformed from a distant astronomical object into a cultural obsession. Baron's vivid narrative reads as part history, part social study, and part literary drama, making complex scientific dynamics accessible without sacrificing depth.

The martian "canals"—optical illusions widely mistaken for infrastructure on the distant planet—became the foundation for an elaborate but inaccurate vision of an advanced civilization. Amplified by mass media, this idea fed a cultural feedback loop where entertainment, belief, and speculation blurred into perceived scientific fact. But society's fascination with Mars also reflected deeper cultural needs, argues Baron. Amid rapid industrialization, global tensions, and spiritual uncertainty, the red planet came to symbolize humanity's hopes for technological progress and the possibility of peaceful, intelligent life beyond Earth.

Science became spectacle in the Gilded Age, shaped by outsized personalities rather than cautious deliberation. Baron shows how the era's nascent scientific institutions lacked the mechanisms to counter charismatic advocacy—a dynamic he illustrates through compelling portraits of the very personalities who exploited this institutional vulnerability.

Chief among these figures was Percival Lowell, the wealthy "poet astronomer" whose resources, eloquence, and unwavering conviction made him especially influential. His telescopic observations, distorted by both atmospheric conditions and personal ambition, helped entrench the myth of canal-building martians in the public imagination. What set Lowell apart was his stubborn certainty, which turned a scientific hypothesis into an unyielding belief. As Baron writes, "millions read of Lowell's lectures, and as they did so, his ideas—his very language—filtered deep into the American consciousness."

Mars evangelism extended beyond Lowell to include other compelling figures: Nikola Tesla, the "inventor of strange brilliance"; Mabel Loomis Todd, a "spellbinding lecturer"; and Camille Flammarion, the "starry-eyed prophet" with a "cultlike following." These personalities gained crucial momentum through what Baron refers to as a "giddy" yellow press that successfully framed the debate as a choice between "the stodgy old guard and an inventive newcomer"—with the public consistently favoring the latter.

Perhaps most troubling was the scientific community's initial response to the Mars fervor—or lack thereof. Although figures such as Edward Holden voiced forceful opposition, the broader establishment failed to mount a consistent or unified rebuttal. Instead, hesitancy, internal divisions, and difficulty confronting the media spectacle enabled Lowell's narrative to dominate.

The scientific establishment eventually responded decisively. When Mount Wilson Observatory astronomers deployed the era's most advanced telescope, they found no evidence of Lowell's canals—delivering the empirical refutation that passive responses had failed to provide. Yet, confronted with contradictory evidence, Lowell "stubbornly refused to update his thinking when new evidence came in."

To his credit, Baron avoids caricature. Figures such as Lowell are not painted as charlatans but as complex characters—brilliant, flawed, and caught in the zeitgeist. In this way, *The Martians* does more than examine this episode in astronomical history; it also illuminates the very human processes by which science is often made, contested, and believed. As Baron notes, "Science eventually self-corrects, and Lowell's vigorous advocacy pushed other astronomers to refine their own ideas and evidence. The edifice of science rests on the remains of countless discarded theories; error forms an inevitable part of the constructive process." Baron wisely resists drawing heavy-handed lessons from his historical narrative, instead allowing the Mars episode to illuminate enduring patterns in how scientific authority is constructed and contested. □

ACKNOWLEDGMENTS

The reviewer used AI-assisted tools (ChatGPT) in the final phase of writing to improve the readability of this piece.

10.1126/science.aea2456

Animals of all sorts have varying understandings of death.

SCIENCE OF DEATH

What, if anything, plays through a possum's mind when—facing a predator—it engages in a convincing approximation of death? What does the predator make of this behavior? This week on the *Science* podcast, Susana Monsó, author of **Playing Possum**, invites listeners to consider how nonhuman animals approach and understand death.

https://scim.ag/PlayingPossum

10.1126/science.aeb1117

The reviewer is at the Zvi Meitar Institute for Legal Implications of Emerging Technologies and Harry Radzyner Law School, Reichman University, Herzliya, Israel, and the Department of Biomedical Informatics and Data Science, Yale University, New Haven, CT, USA.
Email: dov.greenbaum@aya.yale.edu

PHOTO: PANORAMIC IMAGES/ALAMY STOCK PHOTO

LETTERS

Edited by
Jennifer Sills

Shipyard workers spray copper-based antifouling paint onto a vessel's hull to prevent marine biofouling.

Copper threatens marine ecosystems

Antifouling paint prevents organisms such as barnacles from sticking to the hull of ocean vessels. Each year, more than 900 million liters of antifouling paint enter the ocean as these coatings degrade, releasing toxic metals into the water (*1*). Because tributyltin, one active ingredient in antifouling paint, has severe impacts on nontarget species, it was banned globally in 2008 (*2*). Copper-based biocides—primarily copper(I) oxide (Cu_2O) and copper pyrithione (CuPT)—then became the industry standard. However, mounting evidence suggests that these copper compounds exert substantial chronic and cumulative ecological impacts. Washington State has ordered an assessment of copper-based antifouling paints and may implement a ban after reviewing the results (*3*). Other states and countries should follow Washington's lead and revise the standards and regulations for antifouling paint components.

Dissolved copper in coastal waters routinely exceeds the chronic threshold determined by the US Environmental Protection Agency (EPA) (*4*). Although rarely acutely lethal at these concentrations, copper impairs key physiological functions in nontarget marine species. Exposure disrupts salmonid olfaction (*5*), hinders crustacean larval development (*6*), and suppresses bivalve immune responses (*7*). CuPT can cause mortality in juvenile rainbow trout (*8*) at concentrations below the accepted safety limits for dissolved copper (*4*). Both dissolved copper and CuPT move through marine food webs, magnifying their ecological harm (*1*).

Copper ultimately accumulates in marine sediments, where it remains bioavailable (*9*). Recorded copper concentrations in some marina environments exceed the Australia and New Zealand Environment and Conservation Council marine sediment guideline (*10*), indicating that copper represents a persistent, long-term ecological threat within sheltered coastal environments. When physical disturbances occur, these toxic sediments can be abruptly resuspended (*11*), transforming a latent hazard into an acute ecological crisis.

Ocean acidification markedly intensifies copper toxicity, compounding the effects of copper exposure (*12*). Under some climate scenarios, toxicity thresholds for sensitive marine organisms are projected to decrease by approximately 50% by 2099 (*12*), effectively doubling the ecological impact of existing copper contamination. These synergistic interactions between climate drivers and copper biogeochemistry constitute a substantial yet underrecognized mechanism that could amplify anthropogenic stress in coastal ecosystems, particularly in industrialized harbors and estuaries already burdened by copper pollution.

Current regulatory frameworks, which are still anchored to water column concentrations, fail to address copper's chronic and climate-intensified risks. Instead, regulatory standards should integrate bioavailability models and revise sediment quality criteria to reflect copper's persistent toxicity under evolving marine conditions. Climate-scenario modeling should further inform prospective risk assessments. Beyond regulation, governments worldwide must fund copper-free antifouling innovations and incentivize their adoption.

Luhua Yang[1], Shun Li[2], Ruiyong Zhang[1]

[1]State Key Laboratory of Advanced Marine Materials, Key Laboratory of Marine Environmental Corrosion and Bio-fouling, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China. [2]State Key Laboratory of Microbial Technology, Shandong University, Qingdao, China. Email: ruiyong.zhang@qdio.ac.cn

REFERENCES AND NOTES

1. Washington State Department of Ecology, "Toxicological effects of antifouling agents on non-target marine species," publication no. 23-04-066 (2023).
2. International Maritime Organization (IMO), "International Convention on the Control of Harmful Anti-fouling Systems on Ships" (2008).
3. Washington State Department of Ecology, "Antifouling boat paints: Update 2024," publication no. 24-04-004 (2024).
4. EPA, "Aquatic life ambient freshwater quality criteria—Copper 2007 revision," EPA-822-R-07-001 (2007).
5. D. H. Baldwin *et al.*, *Environ. Toxicol. Chem.* **22**, 10 (2003).
6. Z. Valeria *et al.*, *Environ. Toxicol. Chem.* **20**, 7 (2001).
7. T.V. Nguyen *et al.*, *Metallomics* **10**, 7 (2018).
8. C. Bourdon *et al.*, *Peer Community J.* **4**, e6 (2024).
9. A. Hartland *et al.*, *Sci. Rep.* **9**, 9573 (2019).
10. California Regional Water Quality Control Board, "Total maximum daily load for dissolved copper in Shelter Island Yacht Basin, San Diego Bay," Resolution No. R9-2005-0019 (2005).
11. J. Eggleton *et al.*, *Environ. Int.* **30**, 7 (2004).
12. L. Cui *et al.*, *Sci. Total Environ.* **927**, 172145 (2024).

10.1126/science.ady8442

Curtail North Korea's illegal wildlife trade

Although North Korea aims for economic self-sufficiency, it does not exclude the use of international, often illicit, trade to generate revenue (*1*, *2*), including the trade of legally protected native wildlife (*3*). Furthermore, in response to the country's economic and food crises (*4*), a black market in wildlife products has emerged that encompasses almost all native mammal species weighing more than 500 g, including those protected under North Korean law and Appendix I of the Convention on International

PHOTO: RUIYONG ZHANG

Trade in Endangered Species of Wild Fauna and Flora (CITES) (*3*, *5*). Illegal trade imperils North Korean biodiversity and threatens the recovery of metapopulations in these species across the Korean Peninsula (*6*).

Before 2020, animal products from highly protected species such as tiger (*Panthera tigris*), Asiatic black bear (*Ursus thibetanus*), and Eurasian otter (*Lutra lutra*) flowed from North Korea into China, breaching China's CITES commitments and, in the case of wild meat, UN Security Council Resolution 2397 (*3*). In response to the COVID-19 pandemic, North Korea closed its borders, which restricted cross-border smuggling (*4*, *7*). Since Russia's full-scale invasion of Ukraine in 2022, North Korea has moved to reengage with its historic allies, Russia and China (*8*).

North Korea's neighbors must act to minimize the trade of protected species. As North Korea's largest trading partner, China can mitigate the ecological, health, and reputational risks associated with illicit North Korean trade to Chinese buyers by curbing domestic illegal wildlife demand (*9*, *10*). Other countries should remain alert to the risk that even official, state-sponsored North Korean wildlife trade may include protected species (*11*).

Joshua Elves-Powell[1,2,3], Jan C. Axmacher[2], John D.C. Linnell[4,5], Sarah M. Durant[1]

[1]Institute of Zoology, Zoological Society of London, London, UK. [2]Department of Geography, University College London, London, UK. [3]Tiger and Leopard Conservation Fund in Korea, Seoul, Republic of Korea. [4]Norwegian Institute for Nature Research, Lillehammer, Norway. [5]Department of Forestry and Wildlife Management, University of Inland Norway, Koppang, Norway.
Email: joshua.powell@ucl.ac.uk

REFERENCES AND NOTES

1. J. V. Hastings, *Rev. Int. Polit. Econ.* **22**, 162 (2015).
2. J. Campbell, *Int. Aff.* **95**, 1470 (2019).
3. J. Elves-Powell *et al.*, *Biol. Conserv.* **308**, 111102 (2025).
4. M. Noland, *Asia Glob. Econ.* **2**, 100049 (2022).
5. J. Elves-Powell *et al.*, *IUCN Otter Spec. Group Bull.* **41**, 210 (2024).
6. A. Borzée *et al.*, *Science* **363**, 1161 (2019).
7. S. Han, M. Richey, *Asian Surv.* **63**, 235 (2023).
8. Y.-C. Ha, B.-S. Shin, *Asian Surv.* **62**, 893 (2022).
9. Q. Huang *et al.*, *Biol. Conserv.* **254**, 108937 (2021).
10. J. B. Rizzolo *et al.*, *Oryx* **57**, 747 (2023).
11. L. S. Wyler, D. K. Nanto, "North Korean Crime-for-Profit Activities," RL33885 (Congressional Research Service, 2008).

10.1126/science.ady8692

OUTSIDE THE TOWER

PHOTO: CHRISTINE ELLIOTT

The Cricket Quartet performs at Purdue University's annual Bug Bowl.

Creating connection with wings and strings

The lights fade to a soft, rosy glow as the chorus of crickets begins to swell—not in a forest, but inside a university lecture hall filled with 200 community members, listening intently. The crickets? A string quartet, in which each musician wears handmade, species-specific cricket wings and uses their instrument to mimic the rasp and trill of Indiana's native insects.

The Cricket Quartet debuted in 2024 at Purdue University's annual Bug Bowl, an insect outreach event that draws tens of thousands of visitors and offers everything from cockroach races to edible insect tastings (1). As cocreators of the performance, we worked to plan an event where classical music, entomology, and the science of sound converge. First, we explain the mechanics of sound and the reasons crickets sing. Then, one by one, the string quartet brings four crickets to life.

The musicians use advanced techniques such as double stops and saltando to evoke complex cricket calls. For classically trained performers, it's a rare opportunity to intentionally sound imperfect. That freedom is contagious and encourages the audience to answer questions, share stories, and join in the conversation, regardless of their background.

The Cricket Quartet serves a purpose beyond entertainment. Insects are essential to ecosystem function and food security, but they remain underrepresented in conservation funding. As public trust in science wanes and global insect populations plummet, insects desperately need better public relations. Music can help. By activating regions of the brain associated with emotion, memory, and reward (2), music can move people in ways that traditional science communication cannot. When the final note fades, the silence reveals a shift in the room. The experience has transformed the viewers' unease into affection for these insects, which we cannot afford to overlook.

Christine Harrison Elliott and Emily Jean Justus

Department of Entomology, Purdue University, West Lafayette, IN, USA. Email: ellio139@purdue.edu

REFERENCES AND NOTES

1. Bug Bowl, Purdue University; https://www.purduealumnus.org/bug-bowl-and-spring-fest/traditions/.
2. S. Koelsch, *Nat. Rev. Neurosci.* **15**, 170 (2014).

10.1126/science.adz9348

CALL FOR SUBMISSIONS

Outside the Tower is an occasional feature highlighting scientists' advocacy experiences. Submit your advocacy story at http://cts.sciencemag.org.

POLICY ARTICLE

PHARMACEUTICAL TRADE

Supply-chain vulnerabilities in critical medicines: A persistent risk to pharmaceutical security

Asymmetries in the global supply chain expose countries to systemic vulnerabilities that require investment in visibility, coordination, and resilience

Giona Casiraghi, Georges Andres, Frank Schweitzer, Luca Verginer

In 2025, the US administration announced broad tariffs against key trading partners, and pharmaceuticals have emerged as a potential next target. This would represent a marked shift from long-standing commitments under the 1994 Agreement on Trade in Pharmaceutical Products, which eliminated tariffs on a wide range of essential medicines. Although specific trade measures continue to evolve, the incident highlights a broader and persistent problem: The global supply chain for critical medicines is structurally fragile. Trade policies such as tariffs are not the cause of these vulnerabilities. Rather, they exacerbate them. Although short-term measures can alleviate immediate disruptions, they cannot address deeper structural vulnerabilities. The tensions spurred by US tariffs present an opportunity to reassess pharmaceutical supply-chain resilience. Rather than retreating into isolationism, the focus should shift to improving systemic resilience through collaboration, transparency, and targeted investment. Visibility into opaque global supply chains is the necessary foundation to build long-term resilience.

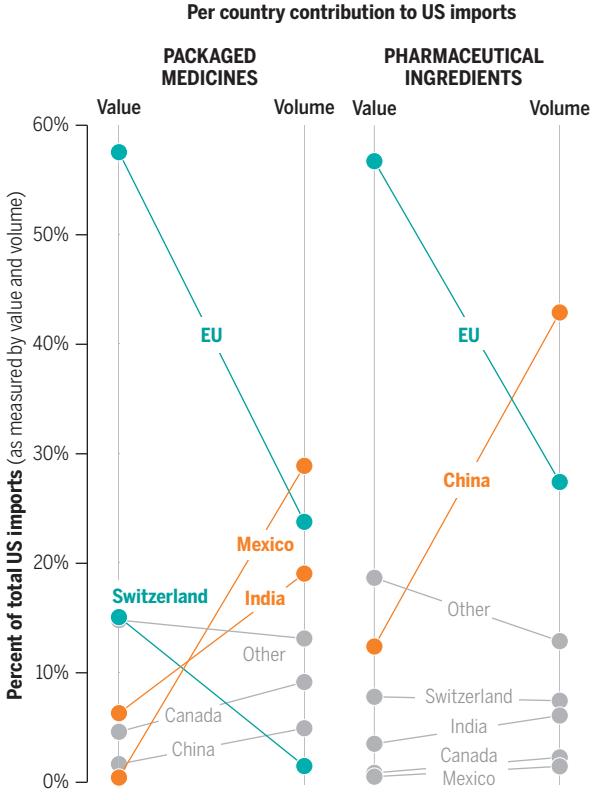
Some medicines are critical for patient care because of the harm their absence would cause. Critical medicines are mostly low-value, high-volume products that form the backbone of modern health care systems yet depend on fragile, global supply networks (*1*, *2*) that are especially vulnerable to disruptions. Fragility in the critical medicines supply chain stems primarily from geographic concentration. Production is increasingly consolidated in a handful of countries, creating strong dependencies: Many importing countries rely heavily on pharmaceutical ingredients and packaged medicines and lack the capacity to respond flexibly to disruptions. As a result, countries with production capacity can exert considerable leverage, whether intentionally or incidentally, over access to critical medicines.

By raising the cost of imported inputs, tariffs may increase production costs for domestic manufacturers. The Association for Accessible Medicines, representing US producers of generic medicines, has warned that many firms operate on narrow margins and cannot absorb these additional costs, thereby increasing the risk of shortages (*3*). Because most critical medicines are prescription drugs reimbursed by insurers or public payers, mechanisms such as fixed reimbursement rates and price ceilings limit how quickly cost increases can be passed on to consumers. As a result, even moderate tariffs can compromise the financial viability of production and destabilize supply. If sufficiently high, tariffs could effectively act as import restrictions, especially when the exporting country has pricing power (the ability to pass along costs to customers without diminishing sales) and the importing country bears the full cost. This risk is particularly acute in the pharmaceutical sector, where demand is price inelastic (demand is not very sensitive to price) and substitution options are limited (*4*).

Compounding these risks, the US, like many other countries, lacks the operational flexibility and buffer capacity to rapidly compensate for lost imports (*1*, *2*, *4*, *5*). Proponents of tariffs may argue that such measures help reduce dependency by discouraging imports and incentivizing domestic production. However, reshoring pharmaceutical manufacturing is a long-term objective, requiring sustained investment, infrastructure, lead times of several years, and regulatory approvals (*2*, *5*). In the short and medium term, tariffs can severely disrupt supply chains, worsening access to medicines and undermining national security and patient care.

Asymmetry of US critical medicine trade

Countries with low trade value but high trade volume, such as China, Mexico, and India (●), hold considerable leverage for potential trade retaliation. By contrast, trade partners with high trade value but low trade volume, such as the European Union (EU) and Switzerland (●), are more vulnerable to the economic impact of tariffs. Data span 2017–2023, with value standardized in USD and trade volume measured in tonnes. See supplementary materials for details on data.



GRAPHIC: M. HERSHER/*SCIENCE*

ASYMMETRIES IN SUPPLY CHAINS

In this article, we follow the definition and list of critical medicines provided by the Critical Medicines Alliance, which includes 288 substances (*6*). We track international trade of these substances using BACI-COMTRADE data (*7*), mapping Anatomical Therapeutic Chemical (ATC) codes, which classify medicines on the basis of their therapeutic use, to their corresponding trade codes (HS6), which are used in international trade statistics to categorize products. This enables consistent mapping across health and trade datasets (see supplementary materials for details).

Most critical medicines are off-patent generics, primarily small-molecule drugs, though some biologics are included. For instance, heparin (an anticoagulant), amoxicillin (an antibiotic), and morphine (an analgesic) all appear across national and international priority lists. Generic medicines (across all therapeutic categories) account for more than 80% of global prescriptions but only 20 to 30% of the total pharmaceutical market value; most critical medicines fall into this category (*8*). Such an imbalance between volume and commercial value has led to the concentration of production capacity in a few facilities, driven by economies of scale, which disincentivizes investment in redundant capacity. Over the past decade, this dynamic has caused critical medicines supply chains to become geographically concentrated and vulnerable to disruptions (*5*, *9*).

The US exemplifies this vulnerability. Its domestic production covers only a fraction of its demand for critical medicines. It relies on imports from key trade partners, including China, the European Union (EU), Switzerland, India, Mexico, and Canada (see the first figure) (*9*, *5*). This external dependence, especially when upstream production is geographically concentrated, creates strategic risk.

Since 2008, more than 60 export bans on antibiotics have been documented globally (*10*). These measures were primarily protective—triggered by domestic health emergencies such as H1N1 and COVID-19. Nevertheless, they underscore the willingness of governments to halt pharmaceutical exports when national priorities take precedence. In April 2025, China responded to US tariff escalation by imposing export controls on several rare earth elements, showing that export restrictions are used as geopolitical leverage. If applied to pharmaceuticals, such retaliatory measures would exploit the thin margins and concentrated production that characterize the sector, where even minor disruptions can cascade into severe shortages.

The economic calculus for export restrictions is starkly asymmetric. For some exporting countries, restricting pharmaceutical trade may inflict disproportionate harm on importers at minimal economic cost. Our analysis shows that the US sources 29% of its imported packaged critical medicines (by volume) from Mexico, yet these account for less than 0.1% of Mexico's total export value to the US (see the first figure). An export ban would degrade US supplies while minimally affecting Mexico's trade balance. Similarly, 43% (by volume) of US imports of pharmaceutical ingredients for critical medicines originate from China, yet these represent less than 0.2% of China's total export value to the US. This asymmetry transforms low-value generics into high-leverage geopolitical tools: Supplier nations may exploit pharmaceuticals to retaliate against tariffs on other sectors.

These same asymmetries also expose countries to self-inflicted disruption. A 125% tariff on Chinese pharmaceutical products, as the one proposed in the April 2025 US trade negotiations, would more than double the cost of critical inputs such as tetracycline active pharmaceutical ingredients (an antibiotic), which the US imports almost entirely (94%) from China. Although other suppliers like the EU contribute small volumes (2 to 3%), they are themselves over 90% reliant on Chinese intermediates and offer products at considerably higher prices. Similar structural dependencies apply to other essential inputs: streptomycins (another antibiotic, 85% from China), vitamin C (91%), steroidal hormones (76%), and even basic medical supplies such as wadding, gauze, and bandages (>78%).

SHORT-TERM EMERGENCY RESPONSE

Tariffs, export restrictions, and ordinary supply disruptions cause drops in supply. However, the demand for critical medicines, being inelastic (*4*), remains unchanged. This mismatch leads to shortages that require swift mitigations.

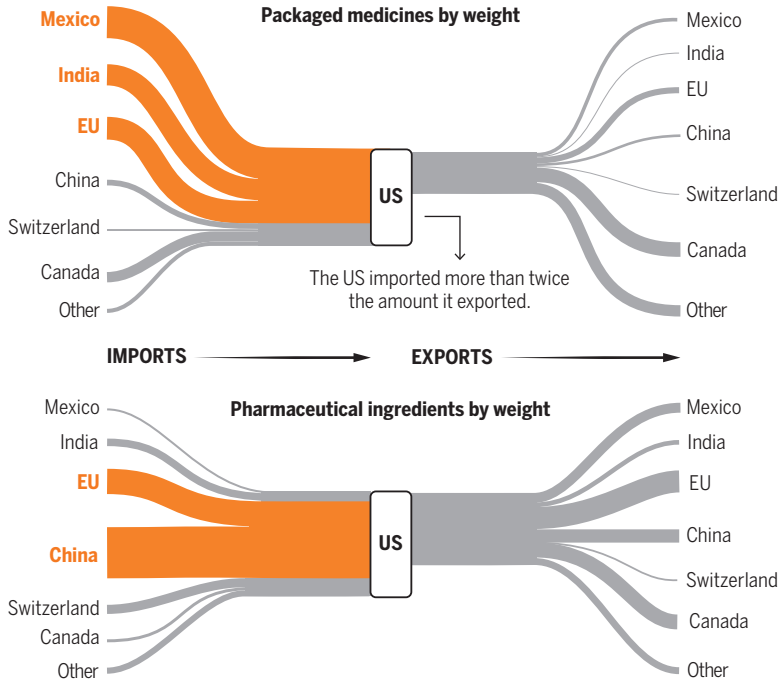
As we explored previously (*11*, *12*), a short-term response is to redistribute stockpiles along existing distribution networks, reallocating doses from areas of surplus to areas of need. This approach can help mitigate vulnerabilities not only in retaliation scenarios but also in the event of other acute disruptions, such as foreign plant closures or logistic bottlenecks, as observed during the COVID-19 pandemic or the 2021 Suez Canal blockage. Short-term redistribution does not solve the root causes of supply-chain fragility but buys valuable time to implement medium-term measures such as initiating trade negotiations, securing emergency imports, or issuing alternative treatment guidelines.

To assess the operational feasibility and impact of such redistribution, in our previous study (*11*), we simulated a disruption to US imports, causing a 70% reduction in the availability of pharmaceutical ingredients for analgesics. By modeling thousands of manufacturers, distributors, and dispensers connected by millions of routes, our work demonstrated how redistributing existing stockpiles along pre-existing, but underutilized, distribution paths can temporarily bridge supply gaps.

Those results revealed three key insights. (i) Time sensitivity: Without intervention, shortages escalate rapidly. A 6% shortage (3 million doses) materialized within 35 days as the disruption propagated through the supply chain. (ii) Hidden capacity: Through informed re-

Import-export balance of critical medicines

The diagram reveals the heavy dependence of the United States on imports from a few major trading partners (●) for most of its pharmaceutical ingredients and packaged medicines. The width of flow lines is proportional to the quantities traded. Data span 2017—2023. See supplementary materials for details on data.



GRAPHIC: M. HERSHER/*SCIENCE*

distribution, the same 6% shortage was delayed to 50 days—a 43% gain in response time. (iii) Systemic visibility: Uninformed redistribution may exacerbate the shortage.

Unlocking this time buffer requires granular, real-time visibility into stockpiles, inventories, and distribution networks. This systemic visibility is technically feasible. In fact, the ongoing opioid crisis in the US has led to the release of more than 10 years of distribution data (*11*, *12*), albeit retrospective in nature. Unfortunately, further efforts are hindered by regulatory fragmentation, inconsistent data standards, and industry resistance to data sharing.

Our simulations confirm both the feasibility and the complexity of large-scale redistribution efforts. In practice, rerouting will need to be implemented in tandem with targeted rationing or triage to maximize its effectiveness and extend the window for coordinated response. Rerouting supplies during shortages does not eliminate disruptions, but it provides time to pivot from reactive crisis management to coordinated emergency response and long-term resilience planning.

LONG-TERM RESILIENCE PLANNING

The US remains a net importer of critical medicines, heavily reliant on pharmaceutical ingredients and packaged medicines from abroad, as demonstrated by comparing imports (inflow) and exports (outflow) (see the second figure). Yet this dependence is not specific to the US; it reflects the inherently complex and globalized nature of the pharmaceutical industry. Production hinges on specialized suppliers concentrated in a few regions (notably China, India, and Europe) with minimal redundancy.

Long-term resilience is often equated with reshoring pharmaceutical production. However, full-scale reshoring is not economically viable for low-margin products owing to cost constraints and global specialization. Even selective reshoring faces considerable technical and institutional barriers: Many medicines require complex inputs, specialized infrastructure, skilled labor, and sustained investment. Establishing domestic capacity would demand long-term commitments, including subsidies, tax incentives, and public–private partnerships.

Critically, feasibility depends on economies of scale and existing industrial infrastructure. Moreover, expanding domestic production would require sufficient regulatory capacity, including inspection, certification, and market authorization. When successful, reshoring can also contribute to global resilience by increasing geographic diversification and strengthening regional production capacity. However, for nonproducing countries, it may simply result in a shift in dependencies, offering limited protection against upstream shocks. Further, although reshoring a single active ingredient may be prohibitively expensive, strategically colocating the production of related drugs could create efficiencies through shared infrastructure, expertise, and supply networks.

Resilience could be improved by selectively reshoring a subset of critical medicines and complementing these efforts with mitigation strategies such as strategic stockpiling, dual sourcing, and improved supply-chain agility (*1*), though these measures require substantial coordination and investment. Given these constraints and trade-offs, policy-makers must answer a central question: Which medicines warrant reshoring investments?

Answering this question requires balancing medical necessity, supply chain fragility, geopolitical risk, and feasibility. Achieving this balance hinges on greater visibility into global supply chains, which remain opaque and challenging to assess (*13*). Although data infrastructures such as the US Food and Drug Administration's electronic drug registration and listing system (eDRLS), the European Medicines Verification System (EMVS), and other initiatives offer a starting point for mapping the global supply chain for critical medicines, these systems are not integrated across borders and do not capture upstream sourcing or dependency structures.

Visibility depends not only on technical capabilities, but also on institutional incentives and governance. Data sharing between firms, regulators, and international partners is hindered by commercial secrecy, fragmented oversight, and inconsistent data standards. Technical solutions to this challenge already exist. For example, federated learning platforms can enable pooled risk estimation across jurisdictions or firms while preserving confidentiality, an approach already trialed in health care contexts where patient data privacy is fundamental (*14*). Developing these tools at scale requires not only investment but also shared governance frameworks that establish common data standards, provide a legal basis for industry participation, and align incentives for both public and private actors to contribute high-quality data. Without reliable information on production and sourcing, policy-makers risk misallocating resources, overlooking critical vulnerabilities, and failing to build resilient pharmaceutical supply chains.

The importance of this issue is reflected in recent legislative initiatives, such as the US Mapping America's Pharmaceutical Supply (MAPS) Act and the EU Critical Medicines Act. Both underscore the national security imperative of mapping pharmaceutical supply chains and applying advanced data analytics to identify vulnerabilities. The vulnerabilities are structural, embedded in the way global production and sourcing of critical medicines have evolved. Trade disputes will only exacerbate them, triggering disruptions that expose and intensify the underlying fragilities. Without visibility into global supply chains, policy-makers operate blindly, unable to see vulnerabilities, prioritize intervention, or allocate resources effectively. □

REFERENCES AND NOTES

1. Health Emergency Preparedness and Response Authority, "Assessment of the supply chain vulnerabilities for the first tranche of the union list of critical medicines," Technical Report (2024); https://health.ec.europa.eu/publications/assessment-supply-chain-vulnerabilities-first-tranche-union-list-critical-medicines-technical-report_en.
2. National Academies of Sciences, Engineering, and Medicine, *Building Resilience into the Nation's Medical Product Supply Chains* (National Academies Press, 2022).
3. Association for Accessible Medicines, "Association for Accessible Medicines Comments on New Tariffs" (2025); https://accessiblemeds.org/resources/press-releases/aam-comments-new-tariffs/.
4. K. Yeung, A. Basu, R. N. Hansen, S. D. Sullivan, *Health Econ.* **27**, 1788 (2018).
5. USFDA, *"Drug Shortages: Root Cause and Potential Solutions,"* Technical Report (US Food and Drug Administration, 2019).
6. European Medicines Agency, "First version of the Union list of critical medicines agreed to help avoid potential shortages in the EU" (2024); www.ema.europa.eu/en/news/first-version-union-list-critical-medicines-agreed-help-avoid-potential-shortages-eu.
7. G. Gaulier, S. Zignago, "BACI: International trade database at the product-level. The 1994-2007 version," Working Papers 2010-23, CEPII, October 2010; www.cepii.fr/CEPII/fr/publications/wp/ abstract.asp?NoDoc=2726.
8. G. Siotis, C. Ornaghi, M. Castanheira, *Market definition and competition policy enforcement in the pharmaceutical industry* (European Center for Advanced Economic Research, 2020).
9. A. Mulcahy *et al*., *"Assessing relationships between drug shortages in the United States and other countries,"* Technical Report (RAND Corporation, 2021).
10. Global Trade Alert, Global trade alert database (accessed 21 February 2025); globaltradealert.org/.
11. A. Amico, L. Verginer, G. Casiraghi, G. Vaccario, F. Schweitzer, *Sci. Adv.* **10**, eadj1194 (2024).
12. A. Amico, L. Verginer, F. Schweitzer, *Sci. Data* **11**, 774 (2024).
13. A. Pichler *et al*., *Science* **382**, 270 (2023).
14. J. Zhou *et al*., *Sci. Adv.* **10**, eadh8601 (2024).

ACKNOWLEDGMENTS

The authors thank F. Bruni for useful discussions.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adx0871

10.1126/science.adx0871

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

REVIEW SUMMARY

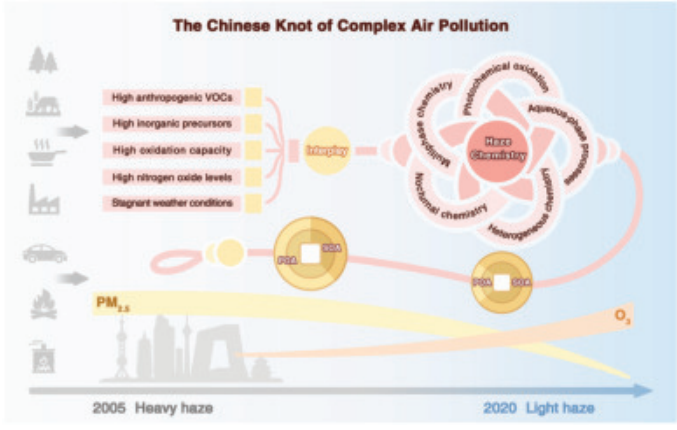
AIR POLLUTION

Secondary organic aerosol in urban China: A distinct chemical regime for air pollution studies

Ru-Jin Huang*, Yong Jie Li, Qi Chen, Yanli Zhang, Chunshui Lin, Chak K. Chan, Jian Zhen Yu, Joost de Gouw, Shengrui Tong, Jingkun Jiang, Weigang Wang, Xiang Ding, Xinming Wang, Maofa Ge, Weijian Zhou, Doug Worsnop, Michael Boy, Merete Bilde, Ulrike Dusek, Annmarie G. Carlton, Thorsten Hoffmann, V. Faye McNeill, Marianne Glasius

Full article and list of author affiliations: https://doi.org/10.1126/science.adq2840

BACKGROUND: Air pollution is a global environmental problem with far-reaching implications for air quality, climate, ecosystems, and human health. Over the recent decades, China has experienced severe haze pollution, also called "complex air pollution." These haze extremes are characterized by high concentrations of fine particulate matter (smaller than 2.5 μm, or $PM_{2.5}$) and occur with extensive temporal and spatial coverage, although the situation was alleviated after 2013, owing to regulatory efforts. Distinct from the historical London fog caused mainly by coal combustion and the Los Angeles smog mainly formed by the photochemical oxidation of vehicular exhaust, haze pollution in China stems from high primary emissions and efficient secondary formation. One distinct feature of China at the current stage is a blend of agricultural and industrial societies, leading to high emissions of secondary aerosol precursors from diverse sources. These include NH_3 from agricultural activities and anthropogenic volatile organic compounds (VOCs) from residential coal combustion and straw burning for heating and cooking (features of an agricultural society), as well as NO_x ($NO + NO_2$), SO_2, and anthropogenic VOCs from vehicular exhaust and industrial activities (features of an industrial society). The mixture of these abundant inorganic and organic precursors can largely change the yields, chemical speciation, and formation pathways of secondary organic aerosol (SOA).



The distinct chemical regime for SOA formation in urban China. The interplay of high anthropogenic VOCs, NO_x, NH_3, and SO_2 from diverse emission sources results in high SOA formation that involves photochemical oxidation, aqueous-phase processes, nocturnal chemistry, heterogeneous chemistry, and multiphase chemistry, especially under stagnant weather conditions. There was a large increase of SOA relative to primary organic aerosol (POA) after the Clean Air Act.

ADVANCES: Numerous studies have revealed that air pollution in urban China is not only of enormous magnitude but also represents a distinct chemical regime less commonly observed elsewhere. The efficient formation of secondary aerosol in winter is distinct, which to a large extent drives the haze formation, particularly in recent winters after large reduction of primary aerosol. Higher-than-expected atmospheric oxidizing capacity in winter China is a combined result of photolysis of elevated HONO, alkene ozonolysis in the presence of NO_x, anthropogenic halogen radicals, and the O_2/H_2O-involved interfacial oxidation and hydrolysis processes. This represents a previously unidentified chemical regime to describe the explosive growth of SOA and secondary inorganic aerosol in complex air pollution. In addition, the elevated NO_x levels can lead to the formation of a variety of nitrogen-containing multifunctional oxidation products. For example, secondary organic nitrate was reported to account for more than 40% of organic aerosol mass during haze events in urban China, which is important for SOA enhancement and altered physicochemical properties. Also, the formation of peroxyacyl nitrates may worsen the regional air pollution by prolonging the effective lifetime of peroxyl radicals. Moreover, the increasing fractional contribution of nitrate over sulfate in recent years results in enhanced aerosol liquid water content, promoting aqueous-phase SOA formation, as revealed through multiple field studies during humid haze events.

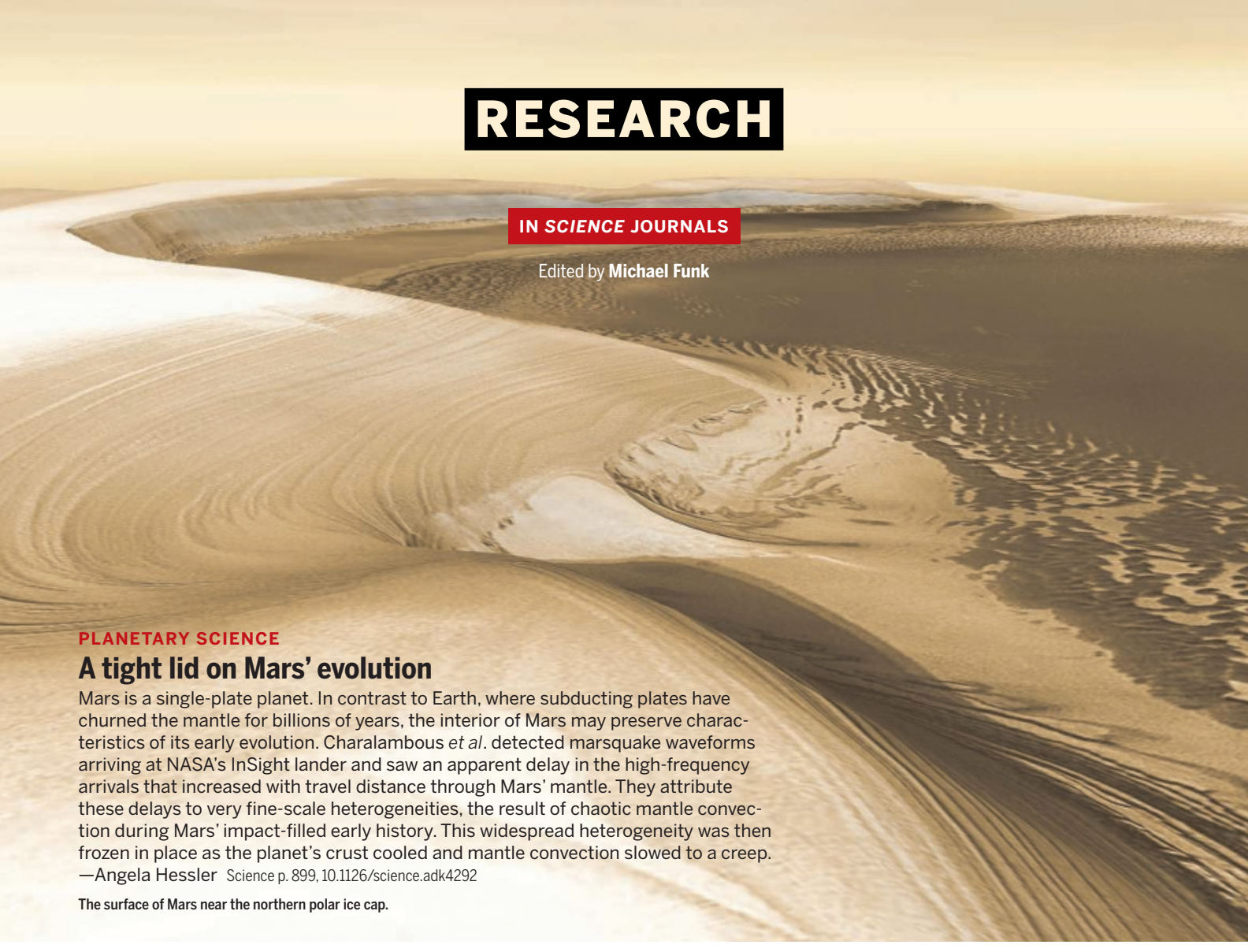
OUTLOOK: Despite considerable advances over the past decade, the precursors, formation, and transformation of SOA in urban China and their impacts on the radiative budget and human health are still very uncertain. Molecular-level speciation of SOA precursors and multigeneration products are essential to elucidate the formation and fate of SOA, and particular focus should be given to less-explored precursors, such as semivolatile and intermediate-volatility organic compounds and volatile chemical products. Noticeably, the sharp decrease of $PM_{2.5}$ concentrations in urban China over the past decade, which has not been reported anywhere else, has led to a substantial increase of surface O_3 concentrations, which facilitates the formation of SOA. Quantitative understanding of the unrecognized sources responsible for the enhanced atmospheric oxidizing capacity in winter urban China is therefore critical. With accurate parameterizations of the above aspects, an improved SOA simulation is expected. As for mitigation, diagnosing the O_3-NO_x-VOC sensitivity at the city or regional scale is essential for a cost-effective strategy to prioritize control measures on precursors that lead to both high SOA and high O_3 concentrations. □

*Corresponding author. Email: rujin.huang@ieecas.cn Cite this article as R.-J. Huang *et al.*, *Science* **389**, eadq2840 (2025). DOI: 10.1126/science.adq2840

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**



PLANETARY SCIENCE

A tight lid on Mars' evolution

Mars is a single-plate planet. In contrast to Earth, where subducting plates have churned the mantle for billions of years, the interior of Mars may preserve characteristics of its early evolution. Charalambous *et al*. detected marsquake waveforms arriving at NASA's InSight lander and saw an apparent delay in the high-frequency arrivals that increased with travel distance through Mars' mantle. They attribute these delays to very fine-scale heterogeneities, the result of chaotic mantle convection during Mars' impact-filled early history. This widespread heterogeneity was then frozen in place as the planet's crust cooled and mantle convection slowed to a creep. —Angela Hessler Science p. 899, 10.1126/science.adk4292

The surface of Mars near the northern polar ice cap.

RESTORATION ECOLOGY

Migrating bison enhance plant nutrition

Bison once migrated across the western United States in enormous herds, shaping the landscape in ways that they currently cannot because smaller populations are now mainly constrained to specific properties. The population in Yellowstone National Park is an exception because these bison are able to migrate across hundreds of square kilometers. Geremia *et al*. performed grazing exclusion experiments to examine the effects of bison on soil and vegetation across their migration area. Grazed areas had higher soil microbe density and nitrogen content in both soils and plants. When given large areas to roam, bison create habitat heterogeneity and areas with enhanced nutrition that can support denser populations than are deemed sustainable by current management recommendations. —Bianca Lopez
Science p. 904, 10.1126/science.adu0703

PROTEOSTASIS

A conserved arena for ubiquitination

Protein homeostasis in eukaryotes is managed in part by serial attachment of the protein ubiquitin to target proteins. The resulting polyubiquitin chains target the protein to the proteasome for degradation. So-called E4 ligases extend a nascent ubiquitin chain to reinforce the degradation signal. Grabarczyk *et al*. determined the cryo–electron microscopy structures of the human E4 ligase UBR4 with two protein cofactors that govern specificity. The protein forms a massive ring with a central active site. A cross-kingdom structural study revealed broad similarities and specific differences in E4 complexes. —Michael A. Funk
Science p. 909, 10.1126/science.adv9309

ROBOTICS

Bubble power

Rapid motions such as jumping require the fast conversion of a large amount of potential energy into kinetic energy. For a robotic system, this can tax the limits of the onboard power supply. Wang *et al*. designed a method to launch a robot that uses focused near-infrared heating, ultrasound, or electrical sparks to remotely initiate the growth and catastrophic collapse of cavitation bubbles. Key to the design is the suppression of bubble release until a stability limit is reached, allowing the energy contained therein to be harnessed upon the bubbles' violent collapse. The authors show that, in addition to jumping, the same approach can be used to power the swimming of a robot. They also demonstrate that this concept works for a range of materials, liquids, and operating environments. —Marc S. Lavine
Science p. 931, 10.1126/science.adu8943

PHOTO: NASA/JPL-CALTECH/ASU

QUANTUM NETWORKS

Classically decisive quantum networks

The development of a quantum network requires the ability to propagate fragile quantum states over large distances, ideally using existing long-distance optical fiber infrastructure. Zhang *et al*. demonstrate the integration of quantum information into advanced photonic technologies. Using a hybrid networking protocol, they showed that the classical signal can be used to efficiently guide entanglement distribution while preserving the integrity of quantum information. This hybrid approach enables dynamic routing of high-fidelity entanglement over existing fiber networks and provides a practical path toward a scalable quantum internet. —Ian S. Osborne

Science p. 940, 10.1126/science.adx6176

ALZHEIMER'S DISEASE

Assembly of prefibrillar amyloids

The amyloid hypothesis is one of the leading explanations for the cause of Alzheimer's disease. It proposes that the accumulation of amyloid-β (Aβ) peptide oligomers and fibrils in the brain is the primary initiating event that triggers a cascade of neurodegenerative processes. Liang *et al.* conducted time-resolved imaging of the prefibrillar assembly of Aβ using a combination of transmission electron microscopy, cryo–electron tomography, atomic force microscopy, and single-particle cryo–electron microscopy. The collection of recorded images proposes a universal mechanism of assembly in which all observed structures, from oligomers to curvilinear protofibrils and ring-shaped annular assemblies, represent a continuum. Patch-clamp conductance measurements suggested that the observed annular structures form channels that span lipid membranes and increase their ion permeability. —Vadim Cherezov

Sci. Adv. (2025) 10.1126/sciadv.adx9030

NANOMATERIALS

Unraveling icosahedra formation

The morphology of nanoparticles is the origin of many of their interesting properties. However, understanding their formation process remains a grand challenge because of complex reaction pathways and the small size of evolving species. Through careful design of ligands and reducing agents, Hu *et al*. synthesized two giant silver icosahedral nanoclusters containing 213 and 429 silver atoms, which serve as model systems for studying icosahedra formation. X-ray diffraction studies revealed multilayered configurations, supporting a layer-by-layer evolution from nuclei to seeds. The emergence of surface plasmon resonance confirms that these silver nanoclusters are metallic. —Jack Huang

Science p. 921, 10.1126/science.adx6639

MICROBIOLOGY

Big roles for small proteins

Bacterial small proteins regulate diverse cellular activities and responses to environmental stimuli. Vellappan *et al*. identified small proteins induced by magnesium ion starvation in *Escherichia coli* and investigated their mechanisms of induction, subcellular localization, and effects on cell growth. One of these proteins, YoaI, mediated cross-talk between the two-component signaling systems PhoR-PhoB and EnvZ-OmpR, which control cellular responses to phosphate starvation and osmotic stress, respectively. These findings expand the known repertoire of stress-induced small proteins that regulate bacterial adaptive signaling. —Annalisa VanHook

Sci. Signal. (2025) 10.1126/scisignal.adu7253

IN OTHER JOURNALS

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

MARINE BIOLOGY

Men-o'-war sailing by

In the 18th century, naturalists discerned that the appearance of the large, floating, jellyfish-like Portuguese man-o'-war, *Physalia*, differed from region to region. Since then, it has been assumed that *Physalia* is a single species and that surface, or neuston, populations are globally well mixed among the connected oceans. In response to efforts to remove plastic that inadvertently capture surface fauna, Church *et al.* wanted to test this assumption of uniformity and discovered that *Physalia* shows substantial genetic variation. Ocean circulation modeling coupled with a citizen-science effort to collect photographs of beached *Physalia* identified at least four species across the world's oceans. The neuston has the potential for distinct regional variation, which needs to be recognized to conserve the biodiversity of these vulnerable populations. —Caroline Ash

Curr. Biol. (2025) 10.1016/j.cub.2025.05.066

Image showing the colonial, marine siphonophore *Physalia* spp. topped by its carbon monoxide–filled float that catches the wind while long stinging tentacles trail underwater.

PHOTO: JURGEN FREUND/MINDEN IMAGES

IMMUNOLOGY

Preventative measures

When the human body detects invasion by a pathogen, it triggers immune responses, which include the activation of innate lymphoid cells. But could the brain help the immune system get a head start by warning of likely infection? Trabanelli *et al*. exposed healthy participants to virtual reality avatars that showed clear signs of infection and measured the resulting immune responses. Compared with exposure to "healthy" avatars, exposure to the "infectious" avatars was associated with an increase in the amount and activation of lymphocytes in the blood, similar to the response induced by a flu vaccine (a model of actual contact with pathogens). This exploratory finding suggests that

PLANT ECOLOGY

Climate shifts captured in herbaria

Many plant and animal species are shifting the timing of events (phenology) such as flowering in response to climate change. Yet in the Arctic, where climate change is proceeding most rapidly, the difficulty in studying phenology means that little is known about the extent of these shifts. Panchen *et al.* examined 17,000 digitized specimens of plants from the Canadian Arctic to determine when flowering occurred for 97 species and how flowering time changed in response to temperature from 1899 to 2019. Most species shifted to earlier flowering, especially late-flowering species, causing a shorter flowering season overall, with potential implications for pollination and food webs. This work shows the importance of digitizing specimens to aid in research and inform conservation. —Bianca Lopez *New Phytol.* (2025) 10.1111/nph.70386

Herbarium specimens of Arctic plants like the moss campion (*Silene acaulis*) reveal the impact of climate change on key traits such as flowering time.

a neuroimmune reaction can prepare the body for potential infection. —Sarah Lemprière

Nat. Neurosci. (2025) 10.1038/s41593-025-02008-y

CARDIOLOGY

Straight to the heart

Nucleic acid therapeutics may eventually be able to treat many diseases, but they must be delivered to the correct tissue. These approaches work best for liver disorders because nucleic acids delivered in lipid nanoparticles get trapped in the liver through a mechanism mediated by apolipoprotein E (ApoE). Shuvaev *et al*. developed lipid nanoparticles optimized for cardiac delivery, and then used ApoE-knockout mice to demonstrate that they could modulate cardiac function with small interfering RNA targeting a component of the cardiomyocytes' contractile mechanism. This method is still far from human application, but it provides a research tool for targeted modulation of cardiac function and shows what might be possible if we can overcome the delivery limitations caused by ApoE. —Yevgeniya Nusinovich

Proc. Natl. Acad. Sci. U.S.A. (2025) 10.1073/pnas.2409266122

STRONG COUPLING

Symmetry rules in polariton chemistry

Symmetry plays a critical role in chemical reactivity, particularly vibrational symmetry, which can substantially influence reaction pathways. Vibrational strong coupling (VSC) with optical cavities represents a new frontier of research where much remains to be understood, including the role of symmetry, because VSC can selectively modify vibrational symmetry. Jayachandran *et al.* explored this effect within donor–acceptor charge transfer complexes in cavities, comparing isomers of trimethylbenzene with different symmetries. Their findings suggest a phenomenological symmetry rule that accounts for reactivity variations across the different point groups. However, further theoretical research is essential to elucidate the general mechanisms by which symmetry influences chemical reactivity under VSC. —Yury Suleymanov

Angew. Chem. Int. Ed. (2025) 10.1002/anie.202503915

GALAXIES

A plane of dwarfs from a galaxy merger

Most large galaxies are orbited by numerous satellite dwarf galaxies, as predicted by cosmological simulations. Several nearby large galaxies have satellite dwarfs orbiting on a single plane, which was not predicted and has an unknown origin. Jerjen *et al.* examined the dwarf galaxies around NGC 5713 and NGC 5719, a pair of large galaxies in the early stages of merging. The authors argue that the relative motion of the merger has arranged the two sets of satellite dwarfs into a single plane rotating around the merger's center of mass. Once the merger is complete, this would leave a plane of satellite dwarfs orbiting a single large galaxy. —Keith T. Smith

Mon. Not. R. Astron. Soc. (2025) 10.1093/mnras/staf1235

BATTERIES

Making liquid sulfur at room temperature

Liquid sulfur droplets generated in a two-step reaction can boost the cell capacity of lithium-sulfur batteries. Subramaniam-Venkatesh *et al.* studied the formation of liquid sulfur in planar cells with platinum cathodes, lithium anodes, and lithium polysulfide salts as the electrolytes. Dispersed liquid sulfur droplets formed by two-electron oxidation of the polysulfide at 3.5 volts at room temperatures or even –15°C, well below the sulfur's melting point of 115°C. This oxidation also occurred with carbon-based electrodes. The droplets effectively enlarged the electrode surface, and initial results showed that their presence led to higher capacity and faster kinetics. —Phil Szuromi

ACS Nano (2025) 10.1021/acsnano.5c09160

PHOTO: PATRICIO ROBLES GIL/MINDEN PICTURES

RESEARCH ARTICLE SUMMARY

PAIN

Somatotopic organization of brainstem analgesic circuitry

Lewis S. Crawford†, Fernando A. Tinoco Mendoza†, Rebecca V. Robertson, Noemi Meylakh, Paul M. Macey, Kirsty Bannister, Tor D. Wager, Vaughan G. Macefield, Kevin A. Keay, Luke A. Henderson*

Full article and list of author affiliations: https://doi.org/10.1126/science.adu8846

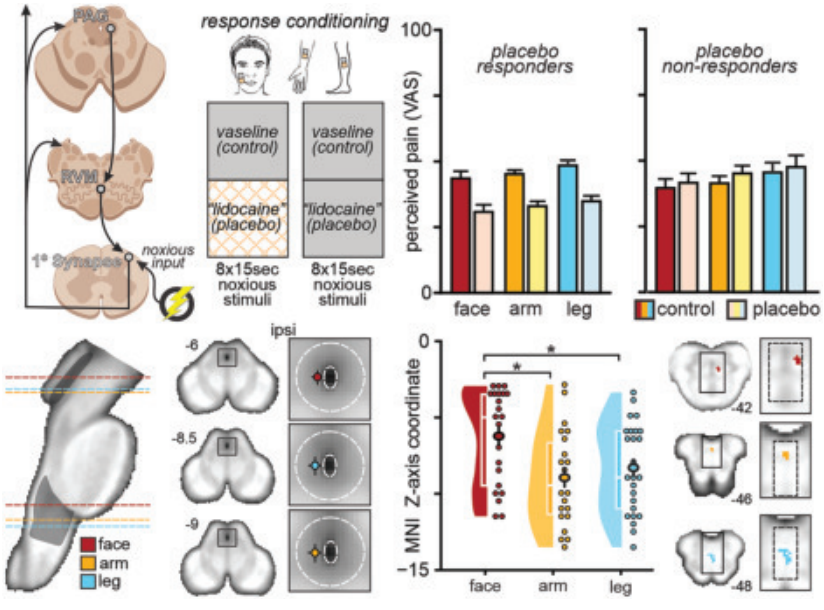
INTRODUCTION: One of our most important survival features is the ability to detect danger and to confront or escape from it. Fight-or-flight behaviors can be instantly initiated by painful stimuli and are mediated by a specific area in the core of the brainstem, the periaqueductal gray matter (PAG). An important aspect of these behavioral reactions is the dampening of the initiating painful stimulus—i.e., analgesia—so that once fight-or-flight behaviors begin, they can be executed without the interference of pain. Experimental studies have shown that the PAG is somatotopically organized—one part receives pain inputs from the face and produces fight behaviors, whereas another part receives pain inputs from the body and produces flight behaviors. Whether the PAG is also organized in this manner with respect to analgesia is not known. If it is, it would be possible to activate specific parts of the PAG to produce pain relief in a restricted part of the head or body. One way to explore the organization of PAG analgesic responses is using a placebo analgesia paradigm.

RATIONALE: By conducting a placebo analgesia conditioning paradigm using a falsely labeled and described cream, 93 pain-free participants believed that this cream could reduce the intensity of painful heat applied to either their face, arm, or leg. In reality, this "lidocaine" cream was a placebo, and the temperature of a thermode was surreptitiously lowered whenever placed onto this cream relative to a control Vaseline petroleum jelly site. During the collection of images of brainstem activity using a 7-T magnetic resonance imaging (MRI) scanner, both the placebo and control cream sites received identical temperatures, with a decrease in perceived pain intensity during the placebo cream site stimulation being a placebo analgesia response. We analyzed activity changes in the PAG and an area that the PAG contacts to inhibit pain, the rostral ventromedial medulla (RVM), during analgesia responses on the face, arm, and leg. Further, using the hidden application of a third cream to a separate body site, we explored whether analgesia was restricted to the area of stimulation or whether it was more widespread.

RESULTS: Our findings show that analgesia evoked on the face, arm, and leg changes activity in a somatotopically organized fashion—i.e., face analgesia is toward the rostral (top) of the PAG and RVM, whereas arm and leg analgesia represented more caudally (lower). These somatotopically organized analgesic responses occurred on opposite sides of the PAG but at the same locations as pain activity changes, and the analgesia was restricted to the site where conditioning was applied.

CONCLUSION: These data show that the PAG regulates analgesic responses in a highly spatially localized manner and has the ability to mediate body site–selective control over pain. The finding that both pain and analgesia evoked PAG activity changes in the same somatotopic fashion suggests that painful inputs to the PAG can not only evoke spatially appropriate behavioral responses, but that these responses are coupled to spatially appropriate analgesic responses. □

*Corresponding author. Email: luke.henderson@sydney.edu.au †These authors contributed equally to this work. Cite this article as L. S. Crawford *et al.*, *Science* **389**, eadu8846 (2025). DOI: 10.1126/science.adu8846



Somatotopic organization of analgesic brainstem circuitry. Placebo analgesia induced on the face, arm, and leg evoked somatotopically organized activity changes within the lateral PAG and RVM revealed through 7-T functional MRI. This analgesic circuitry organization suggests that stimulating discrete brain circuits can produce pain relief in specific parts of the body or face and provides a platform for targeted analgesic treatment development. ipsi, ipsilateral; MNI, Montreal Neurological Institute; VAS, visual analog scale.

RESEARCH ARTICLE SUMMARY

MEDICAL GENOMICS

Machine learning–based penetrance of genetic variants

Iain S. Forrest, Ha My T. Vy, Ghislain Rocheleau, Daniel M. Jordan, Ben O. Petrazzini, Girish N. Nadkarni, Judy H. Cho, Mythily Ganapathi, Kuan-Lin Huang, Wendy K. Chung, Ron Do*

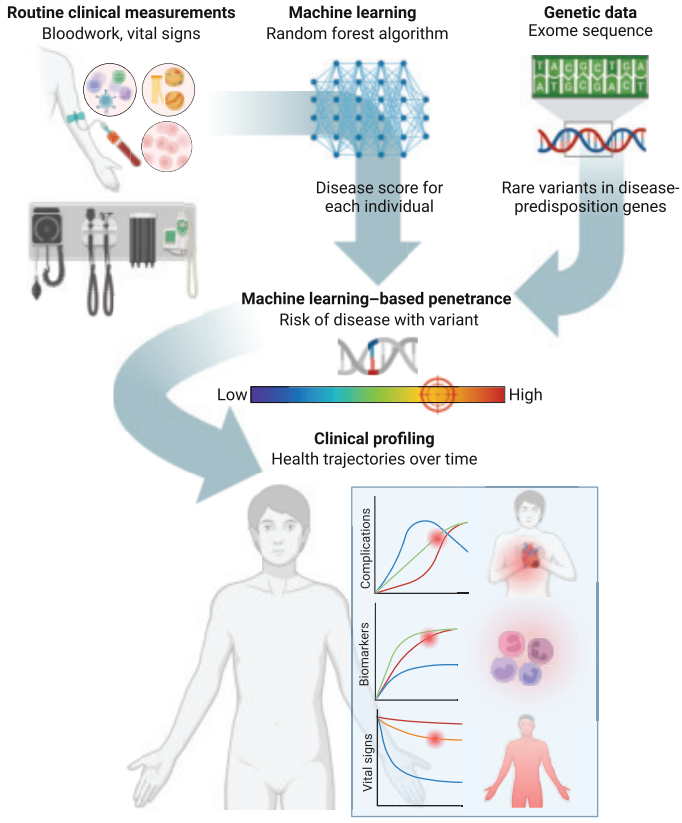
Full article and list of author affiliations: https://doi.org/10.1126/science.adm7066

INTRODUCTION: Accurately estimating the penetrance of genetic variants—the probability that an individual with a variant develops disease—is essential for risk assessment and clinical decision-making. Traditional approaches rely on disease-enriched families or cohorts, which are limited by small sample sizes and ascertainment bias. Moreover, case-versus-control classifications oversimplify diseases that exist on a spectrum, further reducing the accuracy of risk estimates. Machine learning (ML) offers a scalable solution by integrating large-scale electronic health record (EHR) and genetic data to assess penetrance in a data-driven, quantitative, and precise manner.

RATIONALE: Sequencing advances have facilitated the discovery of rare variants in disease-associated genes, many of which are submitted to repositories such as ClinVar and classified based on predicted pathogenicity. However, reliance on laboratory and expert review as well as the absence of large variant datasets measuring real-world disease risk can lead to discordant variant classifications. Emerging population-based analyses have revealed that some variants previously classified as pathogenic (P) exhibit low or variable penetrance, whereas variants of uncertain significance (VUS) remain challenging to interpret clinically. To address these challenges, we developed an ML-based approach to estimate penetrance by leveraging routine clinical laboratory tests, which are widely available in health systems, and intersecting them with genetic data.

RESULTS: We constructed ML models for 10 genetic conditions—arrhythmogenic right ventricular cardiomyopathy, familial breast cancer, familial hypercholesterolemia (FH), hypertrophic cardiomyopathy (HCM), adult hypophosphatasia, long QT syndrome, Lynch syndrome, monogenic diabetes, polycystic kidney disease (PKD), and von Willebrand disease—using 1,347,298 participants with EHR data and applied them to an independent exome-sequenced cohort. Using disease probability scores from these models, we computed ML penetrance for 1648 rare variants across 31 autosomal dominant disease-predisposition genes, spanning P, benign (B), VUS, and previously unknown loss-of-function (LoF) variants. ML penetrance was highest for P and LoF variants, followed by VUS, and lowest for B variants, providing refined quantitative estimates compared with traditional case-versus-control methods.

Notably, ML penetrance correlated with disease-relevant clinical outcomes, such as risk of end-stage renal disease for PKD variants and heart failure for HCM variants. ML penetrance also aligned with experimentally derived measures of variant function, reinforcing its biological relevance. Importantly, ML penetrance aided in the evaluation of VUS and previously unknown LoF variants by delineating clinical trajectories—individuals with highly penetrant variants showed perturbed vital signs, electrocardiogram measures, and disease biomarkers over time. For example, individuals with highly ML penetrant FH variants exhibited 119 mg/dl higher low-density lipoprotein cholesterol and those with highly ML penetrant PKD variants had a 40 ml/min lower glomerular filtration rate.



ML-based penetrance estimation of genetic variants. Routine clinical measurements analyzed by ML generate continuous disease scores, which are integrated with genetic data to assess the penetrance of rare variants in disease-predisposition genes. ML-based penetrance correlates with an individual's clinical outcomes and biomarkers and refines risk assessment of variants beyond traditional pathogenicity categories, presenting a scalable strategy for precision medicine. [Figure created with BioRender.com]

CONCLUSION: This study presents an ML-based blueprint to systematically evaluate penetrance at scale, integrating genomic and clinical phenotype data. By providing refined, individualized disease risk estimates, ML penetrance has the potential to improve variant assessment, guide clinical decision-making, and enhance precision medicine approaches. □

*Corresponding author. Email: ron.do@mssm.edu Cite this article as I. S. Forrest *et al.*, *Science* **389**, eadm7066 (2025). DOI: 10.1126/science.adm7066

RESEARCH ARTICLE SUMMARY

CELL BIOLOGY

A hypoxia-responsive tRNA-derived small RNA confers renal protection through RNA autophagy

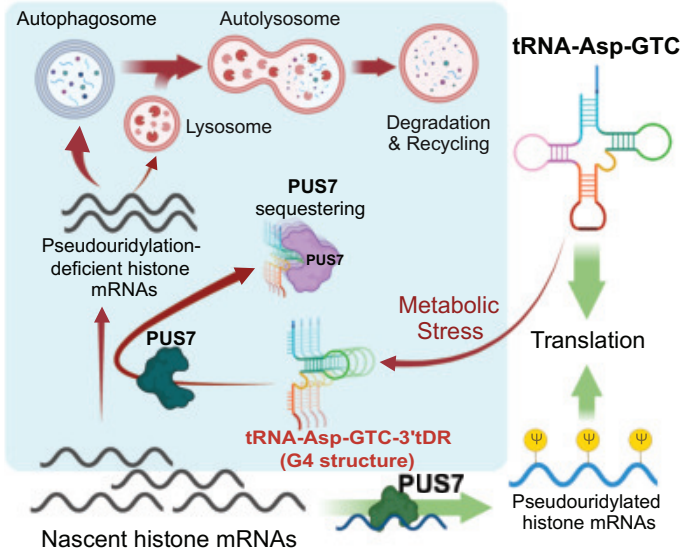
Guoping Li *et al.*

Full article and list of author affiliations: https://doi.org/10.1126/science.adp5384

INTRODUCTION: Cleavage products of transfer RNAs (tRNAs) called tRNA-derived small RNAs (tsRNAs or tDRs), are an evolutionarily conserved class of noncoding small RNAs that play important roles in cellular stress responses. Derived from cleavage by endonucleases such as angiogenin at the anticodon loop, the 5′ halves of tRNA-Ala and tRNA-Cys have been implicated in the regulation of stress granules and global translation. Two tDRs derived from each end of tRNA-Asp-GTC, tRNA-Asp-GTC-3′tDR and tRNA-Asp-GTC-5′tDR, have been shown to be the most highly up-regulated hypoxia-responsive tDRs in multiple cell types. However, their functions in regulating the cellular stress response are not known.

RATIONALE: Although 5′tDRs have been extensively investigated, the regulation and function of 3′tDRs remain poorly characterized. We observed significant levels of tRNA-Asp-GTC-3′tDR in metabolically active tissues, notably the kidney. Validated murine models of kidney diseases and human tissue samples allowed us to investigate the regulation of this tDR in disease, as well as its clinical relevance. Reagents that we developed to modulate the levels of tRNA-Asp-GTC-3′tDR in cells and in vivo allowed us to determine the function of this tDR in kidney disease pathogenesis at a detailed molecular level.

RESULTS: We validated the biogenesis of hypoxia-induced tRNA-Asp-GTC-3′tDR in human embryonic kidney (HEK) cells using ultrasensitive Northern blotting and an engineered fluorescent reporter for tRNA-Asp-GTC-3′tDR, and further confirmed a critical role for the endonuclease angiogenin in its biogenesis. Using overexpression (with synthetic mimics) or specific silencing (using locked nucleic acid–modified antisense oligonucleotides, ASOs), we demonstrated that tRNA-Asp-GTC-3′tDR (but not the 5′tDR) was necessary and sufficient for driving autophagic flux in HEK cells. We found high basal levels of this tDR in most primary kidney cells, correlating with high basal autophagy in these cells. Silencing this tDR in primary kidney cells led to an inhibition of autophagy and cell death, suggesting a homeostatic role for tRNA-Asp-GTC-3′tDR. In several murine kidney disease models, as well as in human tissue samples, tRNA-Asp-GTC-3′tDR was rapidly up-regulated early after kidney injury. Using in vivo delivery of our ASOs, we demonstrated that silencing of tRNA-Asp-GTC-3′tDR led to inhibited autophagy, exacerbated kidney cell injury, and increased inflammation and fibrosis in two different kidney disease models, supporting a compensatory role for this tDR. Conversely, increasing the levels of this tDR using polymer nanoparticle–based delivery of the synthetic mimic was renoprotective, maintaining autophagy and markedly decreasing markers of kidney injury, inflammation, and fibrosis. Mechanistically, the regulation of autophagic flux by tRNA-Asp-GTC-3′tDR was critically dependent on binding to the RNA-modifying enzyme pseudouridine synthase 7 (PUS7). This interaction between tRNA-Asp-GTC-3′tDR and PUS7 relied on both structural oligo-guanine motifs, which allowed for stable intermolecular G-quadruplex (G4) structure formation, and a binding motif for PUS7. Binding and sequestration of PUS7 by tRNA-Asp-GTC-3′tDR prevented the pseudouridylation of target mRNAs, notably histone mRNAs. The pseudouridine-deficient histone mRNAs were targeted to the autophagosome-lysosome pathway for degradation, triggering RNA autophagy as a likely stress-adaptive response.



tRNA-Asp-GTC-3′tDR induces RNA autophagy in kidney cells. Under metabolic stress, kidney cells activate endonucleases such as angiogenin to cleave tRNA-Asp-GTC and generate tRNA-Asp-GTC-3′tDR. tRNA-Asp-GTC-3′tDRs form intermolecular G4 structures and sequester PUS7, thus preventing histone mRNA pseudouridylation. The resulting pseudouridine-deficient histone mRNAs entered the autophagosome-lysosome pathway, triggering RNA autophagy. This process serves as a protective stress response in kidney cells. [Figure created with Biorender.com]

CONCLUSION: This work adds to emerging studies on the role of tDRs in cellular homeostasis and the stress response. We found that the hypoxia-responsive tRNA-Asp-GTC-3′tDR maintains cellular homeostasis in kidney cells by regulating autophagic flux and plays a key role in the stress response. The levels of tRNA-Asp-GTC-3′tDR increased acutely in murine and human kidney diseases to enhance autophagic flux and protect against cellular injury, inflammation, and fibrosis. Mechanistically, key structural motifs in tRNA-Asp-GTC-3′tDR, including the oligo-guanine motif and T-arm region, are essential for both tDR stability and sequestration of the RNA-modifying enzyme PUS7. The sequestration subsequently drives RNA autophagy by preventing the pseudouridylation and stabilization of PUS7-targeted histone mRNAs. Targeting tRNA-Asp-GTC-3′tDR to maintain autophagic flux in the setting of kidney diseases may thus represent a promising therapeutic strategy. □

Corresponding authors: Guoping Li (gli21@mgh.harvard.edu); Saumya Das (sdas@mgh.harvard.edu) Cite this article as G. Li *et al.*, *Science* **389**, eadp5384 (2025). DOI: 10.1126/science.adp5384

RESEARCH ARTICLE SUMMARY

CELL BIOLOGY

Transcription factors SP5 and SP8 drive primary cilia formation in mammalian embryos

Yinwen Liang*, Richard Koche, Ravindra B. Chalamalasetty, Daniel N. Stephen, Mark W. Kennedy, Zhimin Lao, Yunong Pang, Ying-Yi Kuo, Moonsup Lee, Francisco Pereira Lobo, Xiaofeng Huang, Anna-Katerina Hadjantonakis, Terry P. Yamaguchi, Kathryn V. Anderson, Alexandra L. Joyner*

Full article and list of author affiliations: https://doi.org/10.1126/science.adt5663

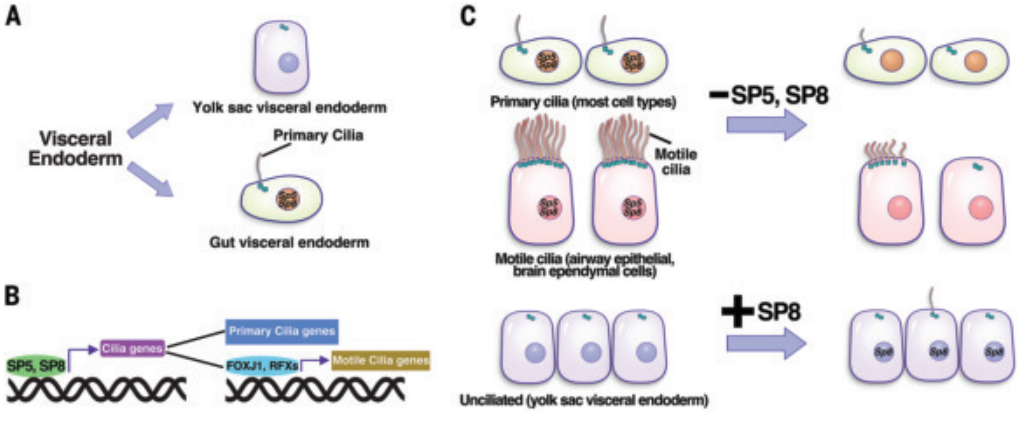
INTRODUCTION: Cilia are thin hairlike structures that project from most cells and provide critical functions during embryogenesis. There are two types of cilia: Primary cilia are present in one copy per cell and involved in signal transduction and sensation, whereas motile cilia are present in multiple copies on terminally differentiated cells and specialized for cell motility or regulating fluid flow. Cilia consist of hundreds of distinct proteins, and most are shared by both cilia types, though there are additional proteins specific to motile cilia. Human mutations in genes involved in cilia structure and function (cilia genes) cause developmental abnormalities, termed ciliopathies, that include hydrocephalus and situs inversus. Despite the importance of primary cilia, whether there are specific transcription factors (TFs) that regulate cilia gene expression has not been determined.

RATIONALE: We reasoned that there is a TF family that coordinately regulates expression of many cilia genes and that these TFs are up-regulated in ciliated cells. Yolk sac visceral endoderm (YsVE) is one cell type in the mouse embryo that does not have cilia. We took a multiomics approach to identify the putative cilia TFs by comparing single-cell RNA and transposase-accessible chromatin sequencing data from YsVE to three cell types with primary cilia in early mouse embryos. We then analyzed mouse gain- and loss-of-function mutants and similar embryonic stem cells (ESCs) to examine whether one TF family promotes cilia formation.

RESULTS: A shared set of cilia genes were found to be up-regulated and to have open chromatin in ciliated gut visceral endoderm and mesoderm cells versus in YsVE cells. TFs SP5 and SP8 were found to preferentially bind to the shared cilia genes, and many cilia genes were down-regulated in mouse ESC–derived gastruloids upon loss of *Sp5* and *Sp8* and up-regulated in YsVE upon misexpression of SP8. The SP5 and/or SP8 target genes included TF families previously shown to induce motile cilia–specific genes. Mouse mutant embryos lacking *Sp5* and *Sp8* had a decrease in the proportion of cells with primary cilia across cell types, and the remaining cilia were shortened. Similarly, motile cilia frequency, number per cell, and length were decreased in the developing lungs and brain. Reflecting the broad defect in cilia formation, the mutant embryos had morphological defects seen in ciliopathy patients. Lastly, we found that misexpression of SP8 was sufficient to induce primary cilia in YsVE and unciliated endoderm SCs in culture.

CONCLUSION: Our study of the transcriptional regulation of cilia genes reveals that SP5 and SP8 bind to and induce many primary cilia genes and genes encoding TFs that regulate motile cilia–specific genes. In addition, *Sp5* and *Sp8* are required for proper primary and motile cilia formation across cell types in mouse embryos. A single TF, SP8, is sufficient to induce primary cilia in unciliated endoderm cells. These findings indicate that SP5 and SP8 are at the top of a transcriptional network that activates primary and motile cilia genes but are only sufficient to induce primary cilia. *SP5* and *SP8* are therefore candidate human ciliopathy genes. □

*Corresponding author. Email: jollylyw@163.com (Y.L.); joynera@mskcc.org (A.L.J.)
Cite this article as Y. Liang *et al.*, *Science* **389**, eadt5663 (2025). DOI: 10.1126/science.adt5663



SP5 and SP8 are necessary and sufficient to drive primary cilia formation. (**A**) In the visceral endoderm lineage, YsVE cells lacks primary cilia, whereas gut visceral endoderm and yolk sac mesoderm cells form cilia and express SP5 and SP8. (**B**) SP5 and SP8 directly activate many cilia genes, including those for TF family members FOXJ1 and RFX, which regulate motile cilia-specific genes. (**C**) SP5 and SP8 are required for the formation of primary and motile cilia in embryos (top). Ectopic expression of SP8 is sufficient to induce cilia in normally unciliated cells (bottom).

FIGURE CREDIT: M3 (XIAO MAN SAN)

RESEARCH ARTICLE SUMMARY

AIR POLLUTION

Heavily polluted Tijuana River drives regional air quality crisis

Benjamin Rico, Kelley C. Barsanti, William C. Porter, Karolina Cysneiros de Carvalho, Paula Stigler-Granados, Kimberly A. Prather*

Full article and list of author affiliations: https://doi.org/10.1126/science.adv1343

INTRODUCTION: The Tijuana River watershed is the epicenter of an environmental and public health crisis. For decades, millions of gallons per day of untreated sewage, industrial waste, and contaminated runoff have been released into the river, subsequently flowing through the Tijuana River Estuary and into the Pacific Ocean. These heavily contaminated waters recently caused nearly 1300 consecutive days of Southern California beach closures, negatively affecting both public health and the economy. Whereas health concerns have typically focused on direct exposure to contaminated waters, this study addresses the often-ignored impact of polluted waters on air quality. Pollutants are transferred from water to air—a process that is enhanced by turbulent flows—where they can be dispersed over tens of kilometers. As water pollution increases globally, understanding the atmospheric consequences of contaminated water bodies is critical to safeguarding public health.

RATIONALE: In summer 2024, even in the absence of rain, wastewater flows surged to millions of gallons per day, enhancing water-to-air transfer of hydrogen sulfide (H_2S)—a toxic gas and key sewage tracer—and other hazardous gases. Despite community concerns, the pollutants causing malodors and adverse health effects had never been identified or quantified. In this study, we combined advanced measurements and trajectory modeling to demonstrate that poor water quality can substantially affect air quality, highlighting inhalation as a major exposure pathway while validating long-dismissed community voices.

RESULTS: Neighborhood-based nighttime H_2S levels near a riverine hotspot located at the Saturn Boulevard river crossing peaked at 4500 parts per billion (ppb)—with hourly averages orders of magnitude above typical urban levels (<1 ppb) and California's 1-hour ambient air quality standard (30 ppb). A sudden wastewater diversion abruptly reduced peak nighttime flows from between 40 and 80 million gallons per day (MGD) to <5 MGD. H_2S levels and odor reports were strongly correlated (correlation coefficient $r = 0.92$), both sharply decreasing (~95%) after the wastewater diversion. The strong correlation supports that the turbulent riverine hotspot was the dominant source of malodors and H_2S. Air samples collected before and after the diversion also identified numerous compounds with known toxicity that decreased in abundance after the wastewater diversion, underscoring that exposure is not limited to a single toxic gas. An air trajectory model was able to reproduce the observed H_2S trends only when meteorology and river flows were represented, further supporting the riverine hotspot as the primary source. The model demonstrated the potential for widespread dispersion of airborne pollutants well beyond the edge of the contaminated Tijuana River.

CONCLUSION: Communities along the Tijuana River Valley have endured decades of environmental pollution, with a public health crisis now clearly linked to airborne exposure of wastewater-derived pollutants. Such prolonged exposure without adequate protection or timely intervention highlights an environmental injustice that demands action. Immediate actions—such as diverting wastewater away from the Tijuana River and implementing localized air monitoring—can sharply reduce toxic exposures, but long-term solutions require upstream treatment infrastructure, strengthened regulatory oversight, and binational collaboration to reduce pollutant discharges at their source. As the number of heavily polluted water bodies grows with ever-increasing inputs, ignoring water-to-air contaminant emissions in air quality management allows hidden risks to persist. Recognizing and addressing this airborne exposure pathway is critical to protecting vulnerable populations and advancing global health equity. □

*Corresponding author. Email: kprather@ucsd.edu Cite this article as B. Rico *et al.*, *Science* **389**, eadv1343 (2025). DOI: 10.1126/science.adv1343

Turbulence in polluted waters transfers contaminants to the air. Culverts at the Saturn Boulevard river crossing generate high turbulence, enhancing water-to-air transfer of toxic wastewater pollutants.

RESEARCH ARTICLE SUMMARY

TISSUE IMMUNOLOGY

Septal LYVE1^{+} macrophages control adipocyte stem cell adipogenic potential

Xiaotong Yu†, Yanan Hu[1]†, *et al.*

Full article and list of author affiliations: https://doi.org/10.1126/science.adg1128

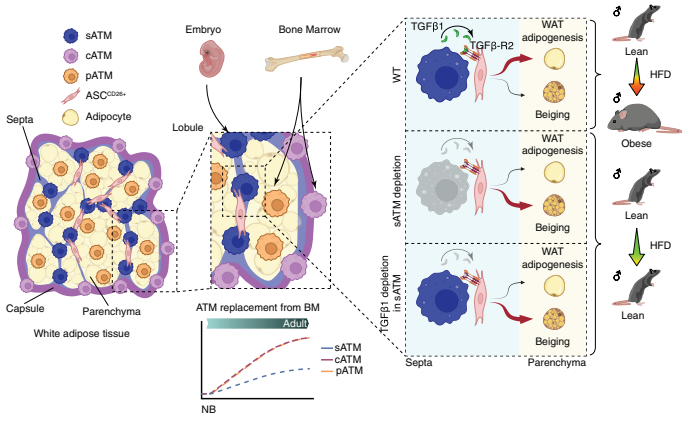
INTRODUCTION: Adipose tissue (AT) is a dynamic organ composed of lipid-storing adipocytes, adipocyte stem cells (ASCs), and resident immune cells. Macrophages are among the most abundant immune cells in AT and, depending on the physiological context, can maintain tissue function or drive inflammation. Prior studies have characterized inflammatory macrophages in obesity, but the identity and role of resident macrophages under steady state and their interaction with stromal populations is poorly understood. Recent evidence indicates that macrophage specialization may extend beyond tissue-level niches into more refined "subtissular" compartments. Whether such localization impacts AT macrophage identity and function has remained unresolved.

RATIONALE: We hypothesized that subtissular macrophage niches within AT serve as instructive hubs for ASC fate specification. Using spatial, transcriptomic, and functional analyses in mice and humans, we investigated whether specific macrophage subsets localized to defined AT compartments, interacted with progenitor cells, and regulated adipogenesis.

RESULTS: We defined three distinct AT macrophage (ATM) populations based on their subtissular localization and molecular signatures: parenchymal (pATMs), capsular (cATMs), and septal (sATMs). The sATMs, characterized by the expression of CD209b and LYVE1 (CD209b^{+}LYVE1^{+}), were selectively enriched in intralobular septa, a dense collagen- and hyaluronan-rich structure traversing adipocyte lobules, in close proximity to early CD26 expressing ASCs (ASCs^{CD26+}). These sATMs were primarily embryonic-derived, long-lived, and distinct from monocyte-derived, inflammation-associated ATMs. In response to a high-fat diet (HFD) challenge, sATMs remained spatially restricted and resisted monocyte replacement whereas pATMs increased in number through both local proliferation and monocyte influx. Selective genetic depletion of sATMs using a Cd209b-driven DTR model enhanced thermogenesis, WAT beiging, increased oxygen consumption, and protection from HFD-induced obesity, independently of food intake. These mice also exhibited improved glucose tolerance and insulin sensitivity, along with reduced liver steatosis.

Single-cell and bulk RNA-seq of AT stromal fractions identified ASCs^{CD26+} as highly proliferative, early progenitors localized in the septa. Upon sATM depletion, ASC^{CD26+} frequency declined, and their transcriptional profiles shifted toward a thermogenic, brown/beige fate. Ligand-receptor interaction modeling pinpointed TGFβ1 as a signal derived from sATMs that acted on ASCs via TGFβR1/2. Specific deletion of *Tgfb1* in LYVE1^{+} or TIM4^{+} macrophages in vivo recapitulated sATM depletion phenotypes. In vitro, ASCs^{CD26+} from TGFβ1-deficient mice showed impaired white adipogenic differentiation and greater capacity for thermogenic conversion.

We identified a conserved population of CD206^{+}LYVE1^{+} human septal ATMs (hsATMs) in obese patients, expressing high levels of TGFB1 and residing adjacent to CD26^{+}CD55^{+} early human ASCs (hASCs^{CD26+}). Multiplex imaging confirmed close physical proximity between hsATMs and hASCs^{CD26+} within septal zones of human WAT.



Spatially distinct macrophage niches govern adipogenesis. White fat is organized into three anatomical compartments—septum, capsule, and parenchyma—each containing distinct populations of adipose tissue macrophages (ATMs). Septal macrophages (sATMs) are positioned in close contact with a population of early fat stem cells expressing CD26 (ASCs^{CD26+}) within the septal niche. This spatial association enables sATMs to guide stem cell fate through TGFβ1 signaling, promoting the formation of energy-storing white fat cells. [Figure created using BioRender.com]

CONCLUSION: Subtissular localization may be a critical determinant of macrophage function and positions the adipose septum as a discrete immunometabolic niche essential for tissue remodeling. Our findings define a spatially confined niche containing resident macrophages and stem cells within the septum of white adipose tissue that governs adipose tissue plasticity and energy balance. Septal ATMs provided localized TGFβ1 signals that direct early adipocyte progenitors toward white adipocyte differentiation while restraining thermogenic potential. Depletion or functional inactivation of this sATM niche shifted the ASC fate toward beige adipogenesis, enhanced energy expenditure, and conferred systemic metabolic benefits in obesity. This regulatory axis between septal resident macrophages and adipocyte progenitors develops understanding of immune-stromal cross-talk in metabolic tissues. Targeting the sATM–TGFβ1–ASC axis may offer therapeutic strategies to boost WAT beiging and counteract obesity and insulin resistance without triggering inflammation. □

Corresponding author: Svetoslav Chakarov (svetoslav_chakarov@shsmu.edu.cn); Florent Ginhoux (florent_ginhoux@immunol.a-star.edu.sg) †These authors contributed equally to this work. Cite this article as X. Yu *et al.*, *Science* **389**, eadg1128 (2025). DOI:10.1126/science.adg1128

PLANETARY SCIENCE

Seismic evidence for a highly heterogeneous martian mantle

Constantinos Charalambous[1]*, W. Thomas Pike[1], Doyeon Kim[2], Henri Samuel[3], Benjamin Fernando[4,5], Carys Bill[2,6], Philippe Lognonné[3], W. Bruce Banerdt[7]†

A planet's interior is a time capsule, preserving clues to its early history. We report the discovery of kilometer-scale heterogeneities throughout Mars' mantle, detected seismically through pronounced wavefront distortion of energy arriving from deeply probing marsquakes. These heterogeneities, likely remnants of the planet's formation, imply a mantle that has undergone limited mixing driven by sluggish convection. Their size and survival constrain Mars' poorly known mantle rheology, indicating a high viscosity of $10^{21.3}$ to $10^{21.9}$ pascal-seconds and low temperature dependence, with an effective activation energy of 70 to 90 kilojoules per mole, suggesting a mantle deforming by dislocation creep. The limited mixing, coupled with ubiquitous, scale-invariant heterogeneities, reflects a highly disordered mantle, characteristic of the more primitive interior evolution of a single-plate planet, contrasting sharply with the tectonically active Earth.

A planet's mantle, situated between its crust and core, holds clues to planetary origin and evolution. On Earth, ancient mantle heterogeneities, identified through the isotopic signatures of lava erupted at hot spot volcanoes and fed by mantle plumes, may be remnants of an early magma ocean (*1*, *2*) or core-mantle interactions (*3*). However, widely distributed passive mantle heterogeneities are unlikely to have survived 4.5 billion years of convective stirring (*4*), unless their physical properties differ considerably from the convecting mantle (*5*–*7*). Today, the dominant source of mantle heterogeneity instead arises from ongoing subduction of oceanic lithosphere driven by plate tectonics, which, although observable seismically (*8*, *9*), offers limited insight into Earth's early mantle history (*4*, *8*, *10*, *11*).

By contrast, Mars is a smaller, single-plate planet (*12*). It likely convects less vigorously with more limited mixing of its primordial components (*13*) beneath a crust decoupled from mantle dynamics (*12*). As a result, Mars is expected to retain a more complete record of its early interior processes compared with Earth (*14*), potentially shedding light on the evolution of terrestrial planets (*15*). Although the heterogeneity in the martian mantle can be locally inferred from chemical and isotopic variability in achondritic meteorites (*14*), seismic studies offer the opportunity to examine present mantle heterogeneity at the planetary scale.

Martian seismic observations

Seismic data from NASA's InSight mission (*16*), gathered from marsquakes, has greatly advanced our understanding of Mars' interior. Whereas initial analyses led to foundational interior models of Mars' large-scale velocity structure (*17*), recent studies have also uncovered smaller-scale velocity perturbations in the planet's lithosphere—the rigid outer shell of crust and shallow mantle—offering new insights into its scattering and absorption properties (*18*, *19*). As seismic waves traverse a planet's interior, they scatter upon encountering compositional and structural heterogeneities on scales comparable to their wavelengths. When finer-scale heterogeneities are prevalent and densely distributed, this scattering intensifies at higher frequencies, causing the seismic wavefront to distort and lose coherence, manifested as a systematic time delay in the wavefront's arrival (*20*–*22*). Owing to Mars' smaller circumference, the seismic wavefront remains detectable at higher frequencies over greater epicentral distances and at relatively greater depths within the planet as compared with Earth (*23*), allowing the identification of finer-scale heterogeneities preserved across the planet's mantle.

We report that seismic wavefronts exhibit such distortion when propagating through Mars' mantle—defined hereafter as its portion below the lithosphere—with an apparent delay in high-frequency arrivals that systematically increases with event epicentral distance. This allows us to place new constraints on the fine scattering structure of the martian mantle. The time delay is only apparent in distant broadband events that traverse longer mantle raypaths, compared with lithospheric paths from those located nearer to the InSight lander (Fig. 1). Eight seismic events—S0264e, S0421a, S0424c, S1000a, S1094b, S1102a, S1153a, and S1415a, labeled by mission Sol (martian day) of detection—preserved their wavefronts up to 8 Hz across epicentral distances ranging from ~57° to 126.7° (~3400 to 7500 km) (*24*).

We focused our analysis for each event on a short time window around the arrival of the *P* wavefront. After this window, the seismic energy profile is increasingly affected by poorly constrained, intrinsic attenuation of the martian mantle [(*23*) and materials and methods sections 1 and 2]. The nearest and farthest events correspond to two large meteorite impacts with precise locations determined from observed craters (*25*, *26*). Mid-distance events were estimated with differential travel times of the first-arriving *P* and *S* waves, using existing Mars velocity models [(*27*) and materials and methods section 1].

On the basis of raypaths for these events, we found that the apparent *P*-wave time delay correlates with the path length traversed beneath the lithosphere, reaching a maximum depth of ~1400 km for the deepest-probing event (S1415a; Fig. 2). By contrast, marsquakes with epicentral distances ~<50°, consistent with paths largely confined to the lithosphere, show no notable delay attributable to this layer, despite traveling longer paths through the lithosphere compared with deep-diving events (Figs. 1 and 2). Although *S*-wavefront arrivals for these events also exhibit apparent delays at lower frequencies compared with *P* wavefronts, as expected, interference from extended *P*-coda energy prevents their robust estimation (figs. S7 and S8 and materials and methods section 1.2).

Diffusion modeling of seismic-wavefront propagation

We modeled the apparent delay using a simple diffusion theory (*28*), attributing it to multiple scattering by mantle heterogeneities that disrupt the *P* wavefront (Fig. 3A). For the frequency-dependent delay observed from 2 to 8 Hz for *P*-wave mantle velocities [v_p ~ 8 to 9.5 km s^{-1} (*17*)], the dominant wavelength λ corresponds to a heterogeneity length scale, δ, of 1 to 4 km, which is also consistent with the observed *S*-wave delay (materials and methods section 1.2). We characterized the heterogeneity number density distribution using a power law $n(\delta) \sim \delta^{-D}$ (*29*), where *D*, the fractal dimension (ranging from 0 to 3), quantifies the disorder and spatial-filling properties of the heterogeneity (Fig. 3A). The frequency-dependent scattering mean free path (μ) then varies as $\mu(f) \sim \delta^{D-2}$.

[1]Department of Electrical and Electronic Engineering, Imperial College London, London, UK. [2]Department of Earth Science and Engineering, Imperial College London, London, UK. [3]Université Paris Cité Institut de Physique du Globe de Paris, CNRS, Paris, France. [4]Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, MD, USA. [5]Department of Physics, University of Oxford, Oxford, UK. [6]Department of Earth Sciences, University of Oxford, Oxford, UK. [7]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA. ***Corresponding author. Email: constantinos.charalambous@imperial.ac.uk** †Retired.

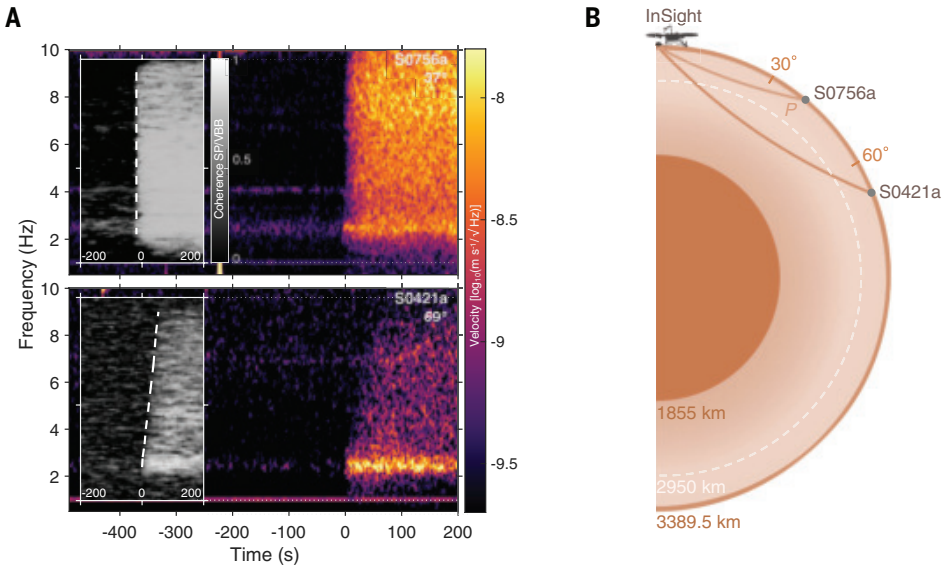
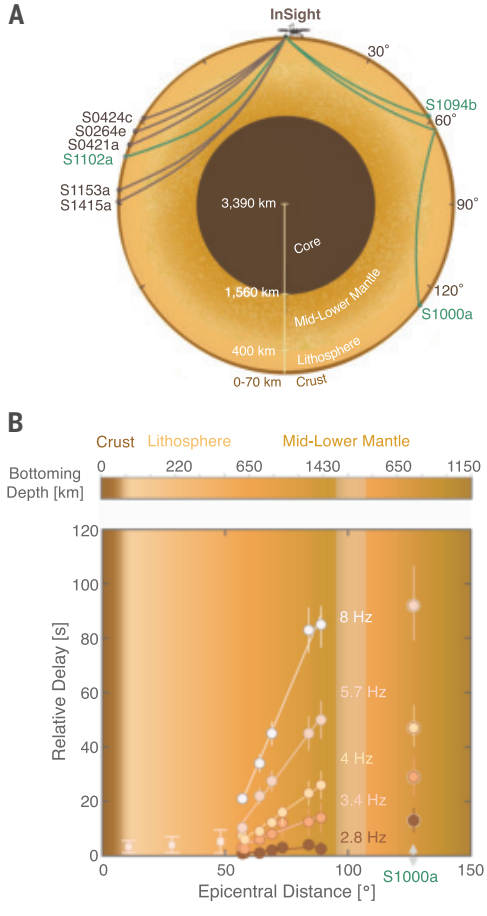


Fig. 1. Contrasting seismic *P* wavefronts for propagation paths in the mantle versus lithosphere. (**A**) Close-up spectrogram (total power) comparison of the *P* wavefront from events S0756a and S0421a (color-scale matched), with corresponding raypaths shown in (**B**). S0756a, a broadband-energy event located at ~37°, exhibits a *P* wavefront with no frequency-dependent delay, representative of all events with seismic wave propagation confined to the lithosphere. By contrast, S0421a at ~68° exhibits a *P* wavefront with an increasing delay at higher frequencies, representative of broadband-energy events at larger teleseismic distances beyond 50°, with raypaths predominantly traversing the mid- to lower mantle. Data up to 8 Hz are derived from the InSight lander's very-broadband (VBB) seismometer; the short-period (SP) seismometer extends the frequency range to 10 Hz. The grayscale inset displays the corresponding coherograms, showing coherence as a function of time and frequency between the SP and VBB signals (20 samples s^{-1}), with seismic energy rising above the largely incoherent background noise. The dashed white line indicates the *P*-wave frequency–dependent arrival, determined independently through noise analysis (materials and methods section 1.1).

Fig. 2. Seismic event raypaths and frequency-dependent *P*-wavefront delays. (**A**) Ray propagation paths for all analyzed events (epicentral distance >50°). Green raypaths represent events with confirmed [S1094b, S1000a (*25*,*26*)] or estimated [S1102a (*24*)] back azimuths, with S1102a located nearly on the opposite side of the planet from the others. Black paths represent the events that lack back-azimuth estimates, depicted in a 2D plane based on Mars' interior structure (*17*). Heterogeneity is depicted only for the mid- to lower mantle, as constrained by this study. (**B**) Observed relationships of the apparent *P*-wavefront delay across different frequencies, plotted relative to the 2 to 2.4 Hz band, as a function of epicentral distance. For epicentral distances ~≤50°, propagation remains lithosphere-bound with minor delays within the error bars, marking a transition where delay trends converge. The background color transition corresponds to depth-layer transitions in (A). Error bars represent delay uncertainties derived from various complementary approaches (materials and methods section 1.1). The lighter vertical strip at ~100° epicentral distance delineates the *P*-wave shadow zone, where rays repropagate through the lithosphere by reflecting midway off the surface as *PP* waves. This boundary extends to the point where *PP* waves begin to penetrate the mid-lower mantle, with only the S1000a impact located in this region. The top bar indicates bottoming depths corresponding to epicentral distances.



Seismic wavefront propagation through a scattering mantle can be parameterized by a frequency-dependent diffusivity, $d(f) = \nu_p \mu(f)/3$. For a mantle propagation length (L), the diffusion time scales as $L^2/d(f)$, characterizing the time required for the diffusive wavefront to traverse this scattering medium, in which the mantle-specific contribution to the delay is isolated. The relative time delay, $\Delta t(f)$, compared with the first observed wavefront arrival, reflects the additional diffusion time induced by increased scattering of higher-frequency energy within the mantle wavefront due to smaller mean free paths and is given by $\Delta t(f) \sim L^2 f^{D-2}$ (materials and methods sections 2.1 and 2.4). Following (*30*), we attribute this additional, frequency-dependent diffusion time of our deeply diving events solely to scattering in the mantle, because for these events only a small portion of the raypath is in the strongly scattering crust. Our diffusion model then gives a best fit to all events when $D = 2.9 \pm 0.1$ (Fig. 3, B and C, and fig. S9). That consistent patterns in the apparent delayed wavefronts are observed from different azimuths across these event locations (*24*–*26*), and are described by this single value of D, implies that mantle heterogeneities are not regional but rather uniformly distributed throughout the planet's interior.

Our simple diffusion model treats Mars as a single heterogeneous medium, a simplification that requires justification if applied at the planetary scale. Thus, to confirm the results for a more realistic model of the planet that incorporates Mars' boundary conditions and existing velocity profiles, we compared the diffusion-model results to particle-based numerical simulations (*31*) (Fig. 3C and materials and methods section 2). Agreement between both models confirms that the high-frequency seismic wavefront in the martian mantle can be treated as primarily diffusive, driven by small-scale heterogeneities that were previously unaccounted for in studies modeling the mantle as homogeneous (*18*, *19*). Notably, scattering in the lithosphere alone cannot account for the large frequency-dependent delay times observed in the *P* wavefront (materials and methods section 2.2). This good agreement also corroborates the weak intrinsic attenuation of the martian mantle beneath the lithosphere—which is consistent with geodynamic constraints [(*2*) and materials and methods section 2]—and aligns with both the high-entropy characteristics (*32*) and highly diffuse nature observed in high-frequency marsquake waveforms (*17*, *18*).

Heterogeneity in the martian mantle

The self-similar nature of the heterogeneity, characterized by a spatial power spectrum with $D \approx 3$, implies that Mars' mantle scatterers occupy nearly equal volumes over a range of length scales, giving a single-valued root-mean-square (RMS) velocity perturbation of ~0.6 to 1% with a correlation length of 1 km (materials and methods section 2). On Earth, RMS velocity perturbations of similar magnitude, inferred from scattered waves probing short-scale mantle heterogeneity, are typically associated with subducted lithospheric remnants with sampling wavelengths >8 to 10 km (*8*, *9*, *31*, *33*–*35*), reflecting the recycling of material in a convecting mantle.

By contrast, the absence of recycling in Mars' mantle implies the preservation of a source that predates the sealing of the planet's interior under a single plate. The lack of a preferred scale length, coupled with the uniformity and ubiquity of the observed heterogeneities as the fractal dimension approaches its upper physical limit, corresponds to a martian mantle that has evolved toward near-maximum entropy [e.g., (*36*–*38*)]. This sustained state of high disorder and complexity [e.g., (*36*)] likely reflects the planet's early processes and subsequent evolution as a closed system.

Considering its long-term survival beneath a stagnant lithosphere, the present heterogeneity is almost certainly compositional in nature because short-scale thermal anomalies would have rapidly diffused away without regeneration [e.g., (*33*)]. With a thermal diffusivity of $\kappa \sim 10^{-6}$ m^2 s^{-1} (*39*), the diffusion length scale over geological times ($t \sim 4.5$ billion years) is given by $\sqrt{\kappa t}$, corresponding to several hundreds of kilometers, a scale that is at least one order of magnitude larger than the characteristic size of thermal anomalies induced by solid-state convection in the martian mantle (*15*). From a geodynamic perspective, the uniform distribution of these compositional heterogeneities throughout the

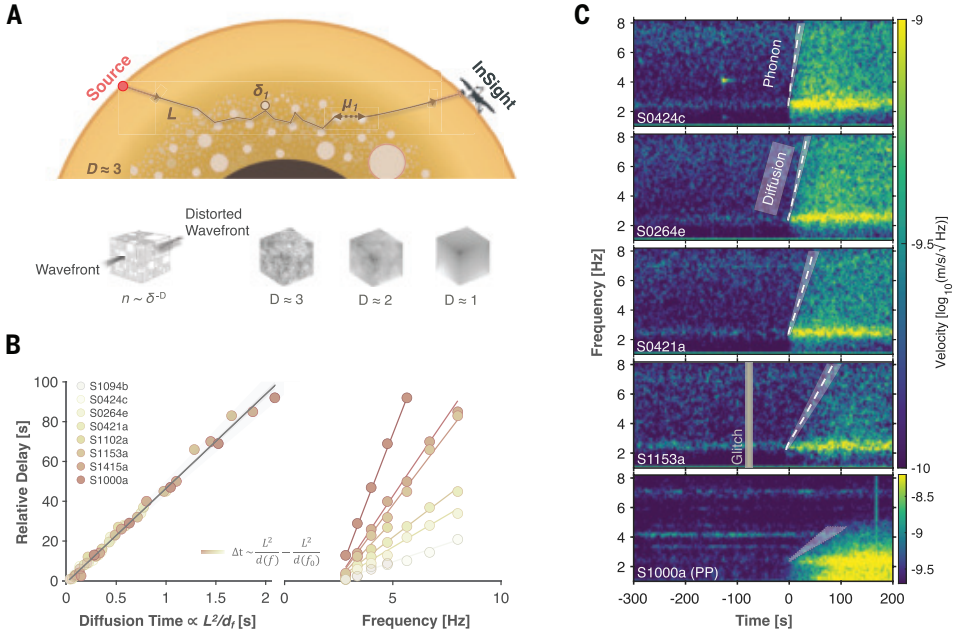


Fig. 3. Modeling seismic-wavefront delays: fractal heterogeneity, diffusive propagation, and *P*-wavefront fits. (**A**) Schematic (not to scale) of the diffusive propagation model with a fractal size distribution of scatterers (depicted only in the mantle), with D approaching 3, indicative of a highly fragmented medium (*29*). The wavefront is multiply scattered, predominantly at heterogeneity size δ_1 with mean free path μ_1. L is the path length within the mantle. (Bottom left) Paradigmatic model of fractal fragmentation (*29*) retaining approximately one block or less at each length scale, yielding $D \sim 2.9$, indicative of volume-filling fragmentation. (Bottom right) Synthetic random fields of a heterogeneous medium with spatial fractal dimensions of $D = 3$, $D = 2$, and $D = 1$, where $D = 3$ exhibits maximum complexity. Lower D values indicate insufficient small-scale structure to explain the increasing high-frequency wavefront delay (materials and methods section 2), because for $D \leq 2$, larger-scale structures dominate, resulting in a smoother medium. (**B**) The frequency-dependent relative delay, Δt, of *P*-wave arrivals scales linearly with increasing diffusion time, driven by lower diffusivity at higher frequencies relative to the reference frequency f_0, here set by the best-fit model at $D = 2.9$. The left panel shows ensemble data collapsing onto a master linear curve for diffusion-time scaling, and the right panel presents individual model fits. (**C**) Total-power spectrograms for five high signal-to-noise events, ranging from ~57° to 126.7° epicentral distance, capturing increasing wavefront distortion. White dashed lines show *P*-wave frequency–dependent delayed arrivals averaged from best-fit particle-based simulations, and the shaded areas represent the best-fit diffusion model where $D = 2.9$, accounting for variations in propagation lengths L across different interior models (materials and methods sections 1 and 2).

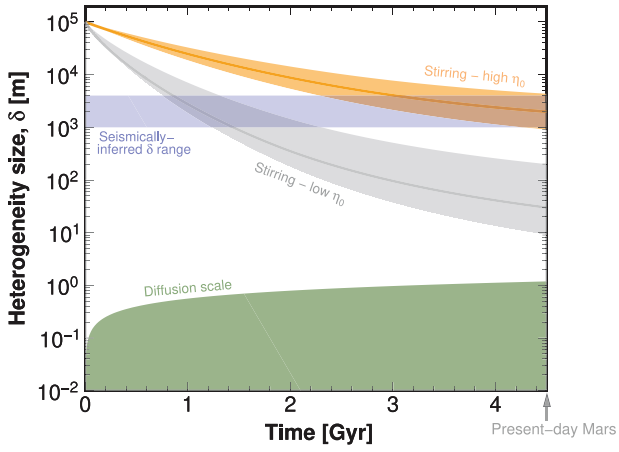


Fig. 4. Evolution of mantle heterogeneity size range (δ) under varying convective stirring conditions. The seismically inferred size range of present heterogeneity is highlighted in lavender. Convective stirring progressively reduces heterogeneity size over time (Gyr, billions of years) but remains well above the much slower regime of the solid-state diffusion limit ($\sqrt{\kappa t} \sim 1$ m; $\kappa_c = 1 \times 10^{-17}$ m^2 s^{-1}, the solid-state mantle chemical diffusivity, shown in green), below which chemical diffusion would homogenize heterogeneities (*11*). For high mantle viscosity ($\eta_0 = 10^{21.9}$ Pa s, shown in orange), heterogeneity sizes overlap with the present seismically inferred range. By contrast, a mantle that is 10 times less viscous ($\eta_0 = 10^{20.9}$ Pa s, shown in gray) undergoes stronger convective stirring, resulting in heterogeneity sizes that are substantially smaller than those observed seismically. See materials and methods section 3 for further modeling details.

mid-lower mantle implies near-neutral buoyancy; even a modest (a few percent) density difference would result in rapid vertical segregation (*4*).

Origin of the heterogeneities

The scaling exponent of the heterogeneity that we have identified, just <3, is characteristic of the injection of energy sufficient to disrupt the coherency of its parental material (*29*). This energy could originate from catastrophic events, such as giant impacts [e.g., (*40*)] with either additional or subsequent intensive mixing of magmas driven by melting and crystallization [e.g., (*36*, *41*)]. Both processes would initially generate self-similar fractal distributions (*29*, *36*, *41*). Once established, chaotic mantle convection plays a crucial role in maintaining these distributions through the self-similar cascade induced by repeated cycles of stretching and folding (*29*, *36*, *42*). Although these convective forces alter their scale, the persistence of the fractal exponent indicates that the mantle's evolutionary processes have neither homogenized these structures nor reduced their dimensionality through mechanisms such as fine-scale layering. Instead, convective stretching has progressively reduced the length scale of the heterogeneities, while folding has reorganized them volumetrically within the mantle, thereby preserving the fractal structure (*43*).

Consequently, the formation and present-day ubiquity of these heterogeneities can likely be attributed to a combination of giant impacts and planet-scale dispersal driven by convective mixing. A giant impact from a large, differentiated projectile, as proposed for Earth (*7*), would initially produce either regional or global heterogeneity [e.g., (*12*, *40*, *44*)], with induced episodic subduction [e.g., (*45*)] entraining fragments of the projectile and early crustal materials into the primordial mantle. The extreme energy of such impacts is expected to produce fragments with a scaling exponent close to 3 (*29*), which would be dispersed throughout the mantle by convection [e.g., (*11*, *44*) and materials and methods section 3]. Additionally, solidification of the impact-induced magma ocean would introduce marked compositional variations into the mantle, even if the impactor itself were undifferentiated (*12*). Such early sources of heterogeneity are consistent with achondritic meteoritic compositions (*14*) arising from magma-ocean crystallization and solidification during the first tens of millions of years of the Solar System, an epoch associated with crust formation on Mars [e.g., (*12*, *40*)]. The global signature of mantle heterogeneity thus reflects the combined effects of giant exogenic impact–driven sources and subsequent endogenic mantle processes, underscoring the pivotal role that these events played in shaping Mars' mantle and its geological history.

Implications for evolution of Mars' interior

Regardless of their origin, the extent to which remnants of primordial compositional heterogeneity have been preserved beneath Mars' stagnant lid quantifies the convective vigor of its mantle [e.g., (*13*)]. Convective motion progressively reduces heterogeneity size through deformation (*42*) with a vigor inversely proportional to mantle viscosity; a more viscous mantle mixes heterogeneities less efficiently, allowing them to persist longer. Unlike Earth, where indirect approaches such as postglacial rebound provide viscosity constraints (*46*), Mars' mantle viscosity is poorly understood, with no constraining measurements to date. We can exploit the Arrhenius relationship governing mantle creep (*47*) to constrain key rheological parameters: reference mantle viscosity, η_0, effective activation energy, E^*, and activation volume, V^*, allowing for the persistence of such heterogeneity. To do this, we also considered Mars' thermal history by exploring a broad range of martian mantle rheological and thermochemical histories through thermal evolution modeling [(*2*, *48*) and materials and methods sections 3 and 4]. We found that the survival of nearly neutrally buoyant primordial heterogeneities over 4.5 billion years of evolution (materials and methods section 4), at scales of ~1 to 4 km (as inferred from our seismic scattering analysis), requires sluggish convection in a highly viscous martian mantle (Fig. 4 and materials and methods section 3). Specifically, this requires a high reference viscosity ($\eta_0 = 10^{21.3}$ to $10^{21.9}$ Pa s), a high pressure dependence ($V^* = 8$ to 9 cm^3 mol^{-1}), and a small temperature dependence of viscosity: $E^* = 70$ to 90 kJ mol^{-1}. The low effective activation energy (E^*) implies that the martian mantle deforms (at a strain rate) within a dislocation creep regime, where viscosity exhibits a power-law dependence with deformation: $\eta^n \propto \dot{\varepsilon}^{1-n}$ with a power-law index $n = 3.5$ (*47*), which corresponds to an intrinsic value of activation energy $E \cong n\,E^* = 245$ to 315 kJ mol^{-1} (*49*). This sluggish mantle rheology is consistent with estimates constrained by the thermal-orbital evolution of Mars-Phobos, which suggest $\eta_0 = 10^{22.2 \pm 0.5}$ Pa s and $E^* = 80 \pm 20$ kJ mol^{-1} (*48*). These estimates correspond to present-day effective viscosities near 10^{23} Pa s (fig. S13B), indicating a martian mantle considerably more sluggish than Earth's upper mantle, whose estimated viscosity ranges between 10^{20} and 10^{21} Pa s (*46*, *50*). Because the viscosity of silicates decreases with increasing water content (*47*), such a difference suggests drier conditions prevailing in Mars' mantle compared with Earth's.

Whereas Earth's early geological records remain elusive (*1*), the identification of preserved ancient mantle heterogeneity on Mars offers an unprecedented window into the geological history and thermochemical evolution of a terrestrial planet under a stagnant lid, the prevalent tectonic regime in our Solar System (*51*); this evolution holds key implications for understanding the preconditions for habitability of rocky bodies across our Solar System and beyond.

REFERENCES AND NOTES

1. A. N. Halliday, R. M. Canup, *Nat. Rev. Earth Environ.* **4**, 19–35 (2023).
2. H. Samuel *et al.*, *Nature* **622**, 712–717 (2023).
3. A. Mundl-Petermeier *et al.*, *Geochim. Cosmochim. Acta* **271**, 194–211 (2020).
4. P. J. Tackley, in *Mantle Dynamics*, vol. 7 of *Treatise on Geophysics*, G. Schubert, Ed. (Elsevier, ed. 2, 2015), pp. 521–585.

5. M. Manga, *Geophys. Res. Lett.* **23**, 403–406 (1996).
6. A. Davaille, *Nature* **402**, 756–760 (1999).
7. Q. Yuan *et al.*, *Nature* **623**, 95–99 (2023).
8. L. Stixrude, C. Lithgow-Bertelloni, *Annu. Rev. Earth Planet. Sci.* **40**, 569–595 (2012).
9. M. A. Hedlin, P. M. Shearer, P. S. Earle, *Nature* **387**, 145–150 (1997).
10. I. S. Puchtel, J. Blichert-Toft, M. F. Horan, M. Touboul, R. J. Walker, *Chem. Geol.* **594**, 120776 (2022).
11. L. Kellogg, D. Turcotte, *J. Geophys. Res.* **95**, 421–432 (1990).
12. K. Mezger, V. Debaille, T. Kleine, *Space Sci. Rev.* **174**, 27–48 (2013).
13. W. S. Kiefer, *Meteorit. Planet. Sci.* **38**, 1815–1832 (2003).
14. A. Udry *et al.*, *J. Geophys. Res. Planets* **125**, e2020JE006523 (2020).
15. S. E. Smrekar *et al.*, *Space Sci. Rev.* **215**, 3 (2019).
16. W. B. Banerdt *et al.*, *Nat. Geosci.* **13**, 183–189 (2020).
17. P. Lognonné *et al.*, *Annu. Rev. Earth Planet. Sci.* **51**, 643–670 (2023).
18. F. Karakostas *et al.*, *Bull. Seismol. Soc. Am.* **111**, 3035–3054 (2021).
19. S. Menina *et al.*, *Geophys. Res. Lett.* **50**, e2023GL103202 (2023).
20. K. McLaughlin, L. Anderson, *Geophys. J. Int.* **89**, 933–963 (1987).
21. L. Margerin, *Seismic Waves, Scattering* (Springer Netherlands, 2011), pp. 1210–1223.
22. H. Sato, M. C. Fehler, T. Maeda, *Seismic Wave Propagation and Scattering in the Heterogeneous Earth* (Springer Science & Business Media, 2012).
23. P. M. Shearer, *Deep Earth Structure – Seismic Scattering in the Deep Earth* (Elsevier, 2007).
24. InSight Marsquake Service, Mars Seismic Catalogue, InSight Mission; V14 2023-04- 01, ETHZ, IPGP, JPL, ICL, Univ. Bristol (2023); https://doi.org/10.12686/a21.
25. L. V. Posiolova *et al.*, *Science* **378**, 412–417 (2022).
26. D. Kim *et al.*, *Science* **378**, 417–421 (2022).
27. C. Charalambous *et al.*, *Geophys. Res. Lett.* **52**, e2024GL110159 (2025).
28. E. Akkermans, G. Montambaux, *Mesoscopic Physics of Electrons and Photons* (Cambridge Univ. Press, 2007).
29. D. L. Turcotte, *Fractals and Chaos in Geology and Geophysics* (Cambridge Univ. Press, 1992).
30. U. Wegler, *Geophys. J. Int.* **163**, 1123–1135 (2005).
31. P. M. Shearer, P. S. Earle, *Geophys. J. Int.* **158**, 1103–1117 (2004).
32. W. Eaton, C. Haindl, T. Nissen-Meyer, *Geophys. J. Int.* **237**, 1109–1128 (2024).
33. G. R. Helffrich, B. J. Wood, *Nature* **412**, 501–507 (2001).
34. T. Zhang, C. Sens-Schönfelder, M. Bianchi, K. Bataille, *Geophys. Res. Lett.* **51**, e2024GL109348 (2024).
35. H. Bentham, S. Rost, M. Thorne, *Earth Planet. Sci. Lett.* **472**, 164–173 (2017).
36. D. Perugini, *The Mixing of Magmas: Field Evidence, Numerical Models, Experiments* (Springer Nature, 2021).
37. O. Zmeskal, P. Dzik, M. Vesely, *Comput. Math. Appl.* **66**, 135–146 (2013).
38. Y. Chen, *Physica A* **553**, 124627 (2020).
39. M. Chai, J. Brown, L. Slutsky, *Phys. Chem. Miner.* **23**, 470–475 (1996).
40. S. Marchi, R. J. Walker, R. M. Canup, *Sci. Adv.* **6**, eaay2338 (2020).
41. R. Deguen, M. Landeau, P. Olson, *Earth Planet. Sci. Lett.* **391**, 274–287 (2014).
42. P. Olson, D. A. Yuen, D. Balsiger, *Phys. Earth Planet. Inter.* **36**, 291–304 (1984).
43. D. H. Kelley, N. T. Ouellette, *Nat. Phys.* **7**, 477–480 (2011).
44. K. W. Cheng *et al.*, *Geophys. Res. Lett.* **51**, e2023GL105910 (2024).
45. A. Yin, *Lithosphere* **4**, 553–593 (2012).
46. J. X. Mitrovica, *J. Geophys. Res.* **101**, 555–569 (1996).
47. S. Karato, P. Wu, *Science* **260**, 771–778 (1993).
48. H. Samuel, P. Lognonné, M. Panning, V. Lainey, *Nature* **569**, 523–527 (2019).
49. U. R. Christensen, *Earth Planet. Sci. Lett.* **64**, 153–162 (1983).
50. D. F. Argus, W. R. Peltier, G. Blewitt, C. Kreemer, *J. Geophys. Res. Solid Earth* **126**, e2020JB021537 (2021).
51. N. Tosi *et al.*, *Astron. Astrophys.* **605**, A71 (2017).
52. InSight Mars SEIS Data Service, SEIS raw data, InSight Mission (2019); https://doi.org/10.18715/SEIS.INSIGHT.XB_2016.
53. B. Knapmeyer-Endrun *et al.*, *Science* **373**, 438–443 (2021).
54. A. Khan *et al.*, *Science* **373**, 434–438 (2021).
55. S. C. Stähler *et al.*, *Science* **373**, 443–448 (2021).
56. C. Charalambous *et al.*, *J. Geophys. Res. Planets* **126**, e2020JE006538 (2021).
57. C. Charalambous, ConstantinosCharalambous/comodulation, GitHub (2020); https://github.com/ConstantinosCharalambous/comodulation.
58. C. Charalambous, ConstantinosCharalambous/comodulation: Archive for comodulation paper, version v1.0, Zenodo (2020); https://doi.org/10.5281/zenodo.4292720.
59. H. P. Crotwell, T. J. Owens, J. Ritsema, *Seismol. Res. Lett.* **70**, 154–160 (1999).
60. H. P. Crotwell, The TauP Toolkit, version 2.6.1, *Zenodo* (2022); https://doi.org/10.5281/zenodo.10794857.
61. N. Mancinelli, P. Shearer, Q. Liu, *J. Geophys. Res. Solid Earth* **121**, 3703–3721 (2016).
62. N. Mancinelli, nmancinelli/psphoton, GitHub (2018); https://github.com/nmancinelli/psphoton.
63. J. Connolly, *Geochem. Geophys. Geosyst.* **10**, 2009GC002540 (2009).
64. J. A. D. Connolly, Perple_X (2025); https://www.perplex.ethz.ch.

ACKNOWLEDGMENTS

This paper is InSight contribution no. 292. We thank helpful discussions with T. Garth on mantle scattering on Earth. We acknowledge the entire Marsquake Service (MQS) frontline team for their comprehensive analysis of InSight seismic data, which made the MQS catalog possible. We also acknowledge NASA, CNES, their partner agencies and Institutions (UKSA, SSO, DLR, JPL, IPGP-CNRS, ETHZ, IC, and MPS-MPG) and the flight operations team at JPL, SISMOC, MSDS, IRIS-DMC, and PDS for providing SEED SEIS data (*52*). **Funding:** C.C. is funded by the UK Space Agency Fellowship in Mars Exploration Science under ST/Y005600/1. W.T.P. is funded by the UKSA under grant number ST/V00638X/1. B.F. is funded by the Blaustein Fellowship of Johns Hopkins University. W.B.B. was supported by the NASA InSight mission and funds from the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004). H.S. and P.L. acknowledge the support of CNES for SEIS operation and science analysis, with additional support from ANR (MAGIS, ANR-19-CE31-0008-08). H.S. and P.L. additionally acknowledge support from the IdEx Université Paris Cité (ANR-18-IDEX-0001). H.S. was granted access to the HPC resources of CINES under the allocation A0110413017, made by GENCI. Numerical computations were partly performed on the S-CAPAD/DANTE platform, IPGP, France. **Author contributions:** Conceptualization: C.C., W.T.P.; Data curation: C.C., D.K., W.T.P., P.L., W.B.B.; Instrumentation: P.L., W.T.P., W.B.B.; Interpretation: C.C., W.T.P., H.S., B.F., P.L., D.K.; Investigation: C.C., W.T.P, P.L., W.B.B.; Methodology: C.C., W.T.P., H.S., D.K.; Mission management: W.B.B.; Project administration: C.C.; Seismic and diffusion analysis: C.C., W.T.P., P.L.; Seismic numerical analysis (phonon global): D.K.; Supervision: C.C.; Thermochemical modeling: H.S.; Visualization: C.C., H.S., W.T.P., D.K.; Writing – original draft: C.C., W.T.P., H.S.; Writing – review & editing: C.C., W.T.P., D.K., H.S., B.F., C.B., P.L., W.B.B. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The InSight seismic data presented and analyzed here are available from the InSight Mars SEIS Data Service at IPGP, IRIS-DMC, and NASA PDS (*52*). The InSight seismic catalog V14 is available at ETH Zurich (*24*). Interior models used in this study are available from studies cited in (*53*–*55*). All the computations made in this article are based on codes described in published papers that are cited in the reference list. Comodulation is described in (*56*) and available at GitHub (*57*), and version v1.0 is archived at Zenodo (*58*). The travel time and ray tracing code TauP is described in (*59*), and version 2.6.1 is archived at Zenodo (*60*). The code for Monte Carlo simulation of elastic-wave scattering processes is described in (*31*, *61*) and is available at GitHub (*62*). The Perple_X Gibbs free energy minimization code is described in (*63*), and version 6.9.1 is available at ETH Zurich (*64*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adk4292
Materials and Methods; Figs. S1 to S16; Tables S1 to S3; References (*65*–*143*)

Submitted 5 September 2023; resubmitted 11 November 2024; accepted 20 June 2025
10.1126/science.adk4292

RESTORATION ECOLOGY

Yellowstone's free-moving large bison herds provide a glimpse of their past ecosystem function

Chris Geremia[1]†, E. William Hamilton[2]*†, Jerod A. Merkle[3]

Although momentum is building to restore bison across North America, most efforts focus on small, managed herds, leaving unclear how large, migrating herds shape landscapes and whether their effects enhance or degrade ecosystems. We assessed carbon and nitrogen dynamics across the northern Yellowstone ecosystem, where one of the last remaining large migratory populations resides. Bison stabilized net aboveground production while accelerating nitrogen turnover, increasing aboveground nitrogen pools while carbon pools remained stable, which improved landscape nutritional quality. Effects were strongest in wet, nutrient-rich habitats that received higher bison densities and grazing than is recommended in rangeland management, while soil and plant conditions suggested landscape resilience. Restoration should embrace heterogeneity in densities and effects across habitats and spatial scales beyond those guiding most current recovery efforts.

Migratory bison are, for all practical purposes, functionally extinct across their former range in North America (*1*, *2*). Bison were driven to near extinction in the late 1800s by market hunting, US military campaigns, and policies aimed at undermining Indigenous communities by destroying their primary food source. It is difficult to imagine the scale of the migrating herds before their decline. Historical accounts provide some illustration—Lewis and Clark estimated a herd of 20,000 bison in South Dakota in 1806, Thomas Farnham described a herd covering 3500 km^2 on the Santa Fe Trail in 1839, Colonel Richard H. Dodd encountered a herd of 3 million animals stretching 24 km wide by 40 km long along the Arkansas River in 1871, and Lakota oral histories measured herd size by the number of days it took them to pass (*3*, *4*). Today, there are about 400,000 bison, 95% of which are privately owned and raised for production, with an average herd size of 175 animals (*5*). The remainder exist in conservation herds averaging around 300 individuals, almost universally managed in constrained areas with strict limits on numbers and movements.

Research focused specifically on bison demonstrates that they can increase habitat heterogeneity, shape plant communities, and influence ecosystem functions such as nutrient cycling and productivity (*6*–*10*). But their role in large, migrating herds across broader landscapes remains poorly understood because today's bison are largely confined to ranches, refuges, reservations, and parks. Restoring this role would require recognizing that migratory bison systems operate differently than conventional grazing management frameworks. These frameworks regulate the timing and intensity of grazing to improve soil carbon storage and maintain desired plant communities (*11*–*15*). By contrast, historic bison migrations likely acted as a major ecological force, interacting with fire to create spatial and temporal heterogeneity (*7*). A bison-dominated North America was likely a heterogeneous landscape with highly concentrated impacts that might be perceived as degraded when viewed through the lens of modern grazing management (*16*, *17*).

Research from the Serengeti ecosystem, supported by plot-level experiments in North American tall and shortgrass prairies, African savannahs, and Tibetan alpine meadows, shows that grasslands are resilient to intensive grazing by native herbivores (*18*–*25*). This resilience arises from complex interactions among grazers, plants, and soil microbes that alter plant stoichiometry, primary production, nutrient cycling, and plant composition (*26*–*33*). Moreover, migratory animals move intentionally to exploit ephemeral foods (*34*, *35*), and the timing and amount of grazing in these areas may play a critical role in shaping animal feedbacks on soil and plant conditions (*36*). One notable trait of aggregate grazers, such as bison and wildebeest, is their propensity to create grazing lawns—heavily grazed areas characterized by short, nutritious vegetation that supports a diverse array of plant and animal species (*7*, *37*).

Restored bison migration enables analysis of grazing and nutrient dynamics

Restored bison migrations in the northern Yellowstone ecosystem now offer a rare opportunity to study how large herbivores shape grazing patterns and nutrient dynamics across a heterogeneous landscape. For more than 150 years, the Yellowstone area has served as a model for ecosystem restoration amid geographical, political, and societal constraints. It spans approximately 90,000 km^2 in the western United States. Through a series of evolving management strategies—including extirpation, reintroductions, control, and land use practices such as cultivation, hay production, and feeding wildlife—top predators and large herbivores such as wolves and bison (now occupying about 10% of the area) were restored to this landscape. Large herbivore populations, particularly elk and bison, fluctuated in response to these strategies (*38*). Since the 1960s, management within the park emphasized minimal human intervention, culminating in the restoration of a full predator guild by the late 1990s. Over recent decades, large herbivore biomass shifted from elk to bison (Fig. 1A). Bison numbers recently reached a stable winter population size because seasonal migrations (Fig. 1, B to D) out of the park have exposed them to hunting and federal removal efforts aimed at limiting range expansion (*39*).

We assessed grazing and nutrient dynamics across the northern migratory landscape from 2015 to 2022 by monitoring 16 sites representing three main habitats: valley bottoms (lawn-forming), adjacent hillsides (dry), and high-elevation topographically wet habitats (tables S1 and S2). Dry habitats were grazed in spring, fall, and winter depending on snowpack; lawn-forming habitats were grazed throughout the growing season; and high-elevation habitats were grazed from mid-season through season-ending frost. At each site, we compared natural grazing with grazing exclusion treatments using a repeated measures analysis of variance (ANOVA) design (*40*). Camera-trap data confirmed that bison were the primary grazers at these sites, accounting for more than 99% of relative abundance in lawn-forming habitats, 93% in high elevation habitats, and 77% in dry habitats (fig. S1).

Bison intensively graze lawn-forming habitats without depleting plant nitrogen pools

Bison numbers averaged near 3500 across the northern migratory landscape, utilizing roughly 852 km^2 annually at an average density of 4.2 bison km^{-2} $year^{-1}$. However, bison densities demonstrated significant variation across the migratory landscape during the growing season (Fig. 2A). The landscape produced an estimated 68 million kg

[1]Yellowstone National Park, National Park Service, Mammoth, WY, USA. [2]Department of Biology, Washington and Lee University, Lexington, VA, USA. [3]Department of Zoology and Physiology, University of Wyoming, Laramie, WY, USA. ***Corresponding author. Email: hamiltone@wlu.edu** †These authors contributed equally to this work.

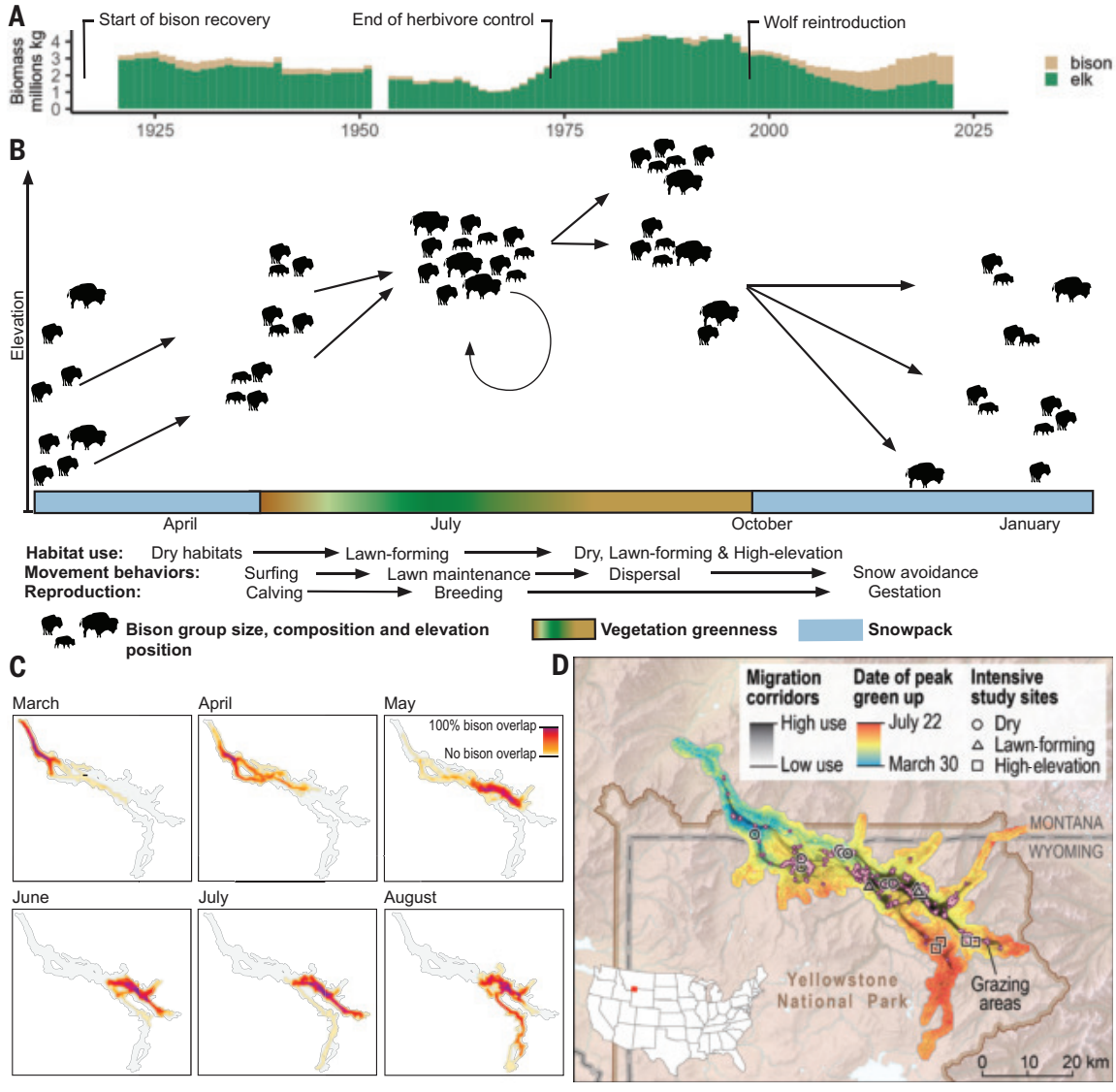


Fig. 1. Bison migration is a medley of moving in tempo with plant emergence, sustaining grazing lawns and escaping snow. (**A**) Bison now account for the majority of herbivore biomass on Yellowstone's northern range, which has a complete guild of migratory herbivores and their carnivores. (**B**) Beginning in spring at lower elevations in the West, bison move in dispersed groups, calving on the move and following the "green wave" of emerging vegetation (*36*). (**C**) As they ascend, they gather in larger, coordinated groups on floodplains and wet grasslands, forming grazing lawns crucial for nurturing calves (*36*). The longevity of high-quality food on these lawns depends on snowmelt and early summer moisture, and as they dry, many bison shift to higher elevations and subalpine meadows, exploiting the last of the emerging spring vegetation in Yellowstone. By fall, bison move westward to escape accumulating snow, pausing at productivity hotspots dictated by terrain. They typically migrate only as far as necessary, often not reaching the lowest elevations and delaying their arrival in the westernmost areas until late winter (*58*). (**D**) The main migratory route in Yellowstone covers about 100 km from west to east in the northern part of Yellowstone, influenced by varying moisture, productivity, and phenology gradients shaped by the region's topography.

of net aboveground production (NAP) per year, or ~19,000 kg of NAP per bison. Bison foraged at varying intensities across their migratory landscape, because consumption increased linearly with density [table S3; coefficient of determination (R^2) = 0.77]. Growing season net consumption was notably higher in lawn-forming habitats (table S4), averaging 135 ± 20.8 grams per square meter per season (n = 12) compared with 19.4 ± 3.1 (n = 20) in dry habitats and 66.9 ± 20.5 (n = 7) in high-elevation habitats. We defined "month" as the number of 30-day intervals since snowmelt at each site, thereby approximating the phenological stage of plant development. Using this definition, we found a significant interaction between habitat and month when comparing daily consumption rates across habitats (table S5). Pairwise comparisons revealed that daily consumption rates were similar across habitats in the first month, were higher in lawn-forming and high-elevation habitats in the second month, and remained higher in lawn-forming compared with dry habitats thereafter (fig. S2 and table S5). Bison consumed more in areas with greater NAP (grams per square meter per year), according to a mixed-effects model relating seasonal net consumption to NAP (Fig. 2B and table S6). The slope of this relationship was higher in lawn-forming habitats [0.64 ± 0.09; credible interval (CrI): 0.48 to 0.83] than in dry habitats (0.39 ± 0.19; CrI: 0 to 0.74) and high-elevation (0.44 ± 0.14; CrI: 0.20 to 0.73) habitats.

Higher bison densities and consumption per unit NAP for lawn-forming habitats suggest use of lawn-forming habitats beyond expectations from plant productivity. To understand why, we assessed carbon

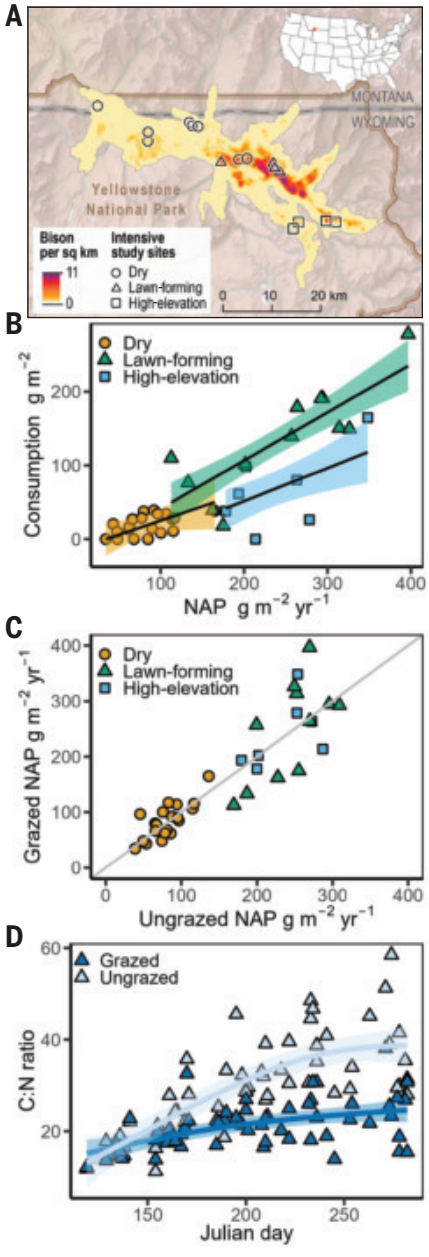


Fig. 2. Bison did not alter plant growth rate through their grazing but decreased whole-plant leaf C:N ratios. (**A**) Bison densities during the growing season were concentrated in lawn-forming habitats ranging from 5 to more than 10 bison per square kilometer per season. Bison densities were estimated from GPS collar and aerial count data. Growing season net consumption (grams per square meter per season) increased linearly with density (table S3; $R^2 = 0.77$). (**B**) Bison consumed more seasonally per unit net aboveground production (NAP; grams per square meter per year). A mixed-effects model showed that the slope of the consumption-NAP relationship was significantly higher in lawn-forming habitats (0.64 ± 0.09) than in dry (0.39 ± 0.19) or high-elevation habitats (0.44 ± 0.14). The 95% credible intervals are shown, based on a random intercept and slope linear model based on habitat. (**C**) NAP in grazed and ungrazed areas followed a 1:1 relationship across all production levels ($R^2 = 0.81$). In (B) and (C), points show annual values for each site (*40*). (**D**) Whole-plant carbon to nitrogen ratios were consistently lower in grazed treatments ($F_{1,52} = 18.52$, $P < 0.001$; table S11). Points show subsamples from each site across years for lawn-forming habitats, where effects were the strongest. Trend line was fitted using a generalized linear model assuming a Gamma distribution.

and nitrogen pools across different grazing areas. We calculated aboveground plant pools by multiplying total aboveground biomass by nutrient percentages. A significant three-way interaction among habitat, grazing treatment, and month affected the aboveground carbon pool size (table S7). This interaction was primarily driven by lawn-forming habitats, where high grazing significantly reduced the aboveground carbon pool size starting in the second month of the growing season (fig. S3 and table S7). Other habitats remained unaffected. However, grazing did not similarly impact the aboveground nitrogen pool size (table S8). Despite substantial bison consumption of plant material in these habitats, the nitrogen pool size in the remaining tissue of grazed plants was comparable to that of ungrazed plants throughout the summer in all habitats (fig. S4).

Grazing accelerates nitrogen cycling through plant regrowth and microbial feedbacks

To understand how plants maintained aboveground nitrogen pools despite intense, repeated grazing, we assessed growth rates and energy and nutrient flow in both grazed and ungrazed plants across the growing season. Grazed plants grew at the same rate as ungrazed plants (Fig. 2C and table S9). Growth rates varied by habitat and month of the growing season, with faster growth in lawn-forming and high-elevation habitats and slower growth in dry habitats (fig. S5 and table S9). Water availability is the primary limiting factor for plant growth in C3 ecosystems such as Yellowstone (*41*). Soil moisture levels were similar between grazed and ungrazed treatments (table S10), allowing soils to sustain essential growth processes such as exudation, microbial activity, and nutrient and water absorption (*42*, *43*). Consequently, as grazed and ungrazed plants grew at similar rates, the tissue removed by grazing was replaced with nitrogen-rich tissue (*41*, *44*). This led to an increase in the proportion of nitrogen in aboveground plant tissue in grazed treatments (Fig. 2D and table S11).

To understand the source of additional nitrogen, we examined microbial processes that mineralize organic matter into plant-available forms (*30*, *31*, *45*) for at least one 30-day period before peak biomass for a subset of sites. Grazing significantly increased microbial mass and ammonia-oxidizing bacteria (Fig. 3, A and B, and tables S12 and S13). This was accompanied by higher nitrogen availability, as measured by ion-exchange membranes that integrate nitrogen flux over time (Fig. 3C and table S14). However, mineralized nitrogen in soil cores was similar between grazing treatments at the start and end of the 30-day period (table S15), indicating rapid plant uptake. Total soil nitrogen—including both organic and mineralized forms—and soil carbon pools were also similar between treatments (tables S16 and S17). Taken together, these findings suggest that grazing enhanced nitrogen cycling without altering total soil nutrient stocks.

Bison grazing enhances crude protein availability across the migratory landscape

By considering the cumulative carbon and nitrogen consumed and reintegrating these nutrients into aboveground plant pools, we found that bison increased the aboveground plant nitrogen pool (Fig. 3D and table S18), while the concentration of carbon in the aboveground pool remained stable (table S19). The increase in the aboveground nitrogen pool was most pronounced at the peak of the growing season in lawn-forming habitats, as indicated by a significant three-way interaction among grazing treatment, habitat, and month (table S18). Established correlations between aboveground plant nitrogen and dietary crude protein for herbivores (*46*) show that when the aboveground plant nitrogen pool peaked, grazing provided 156% more crude protein in lawn-forming habitats, 119% more in dry habitats, and 155% more in high-elevation habitats.

Classic work by McNaughton and others demonstrated that herbivores could enhance nutrient cycling and create nutritional hotspots

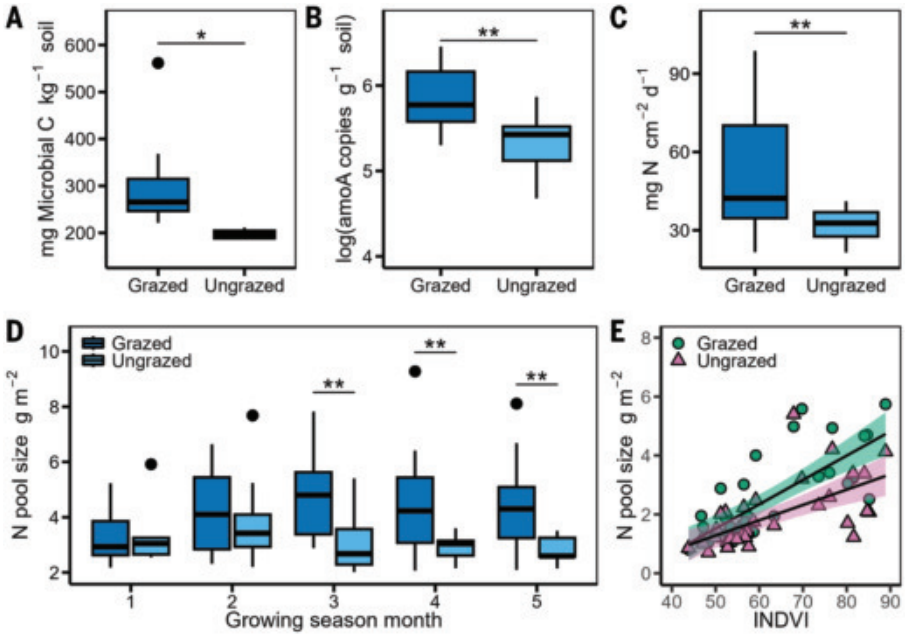


Fig. 3. Bison increased the size of the aboveground nitrogen pool through belowground linkages between microbes and plants. (**A**) Microbial mass increased in grazed treatments after periods of bison grazing (milligrams of microbial carbon per kilogram of soil; $F_{1,10} = 6.702$, $P = 0.027$; table S12). (**B**) Grazing increased the abundance of nitrogen cycle ammonia-oxidizing bacteria, which are responsible for converting soil ammonium into nitrite ($F_{1,20} = 10.678$, $P = 0.004$; table S13). Abundance was measured as the number of gene copies per gram of soil on the log scale of the ammonia monooxygenase (amoA) gene in *Eubacteria* quantified by quantitative polymerase chain reaction (*59*). (**C**) This, in turn, led to an increase in soil-plant available inorganic nitrogen (milligrams of nitrogen per square centimeter per day; $F_{1,18} = 8.208$, $P = 0.010$; table S14). (**D**) Plants in grazed treatments exhibited elevated nitrogen concentration in aboveground tissue ($F_{1,52} = 18.52$, $P < 0.001$; Fig. 2D and table S11), which was then consumed by bison (Fig. 2B and fig. S2). Reintegrating consumed nitrogen into aboveground plant pools resulted in higher aboveground plant nitrogen in grazed treatments as summer progressed, indicated by a three-way interaction among grazing treatment, habitat, and month ($F_{3.04,157.83} = 8.063$, $P = <0.001$; table S14). Adjusted N pool size is shown for lawn-forming habitats only. (**E**) Adjusted N pool size per unit of landscape greenness measured with integrated normalized difference vegetation index (INDVI) was higher in grazing treatments [table S20 (*40*)]. The relationship in (D) suggested that crude protein provided by the migratory landscape was doubled under grazing. In (A) to (D), the boxes represent the interquartile range, and asterisks (*$P < 0.05$, **$P < 0.01$) indicate significant differences based on post hoc tests of treatment (grazed, ungrazed) effects from ANOVA analysis (*40*).

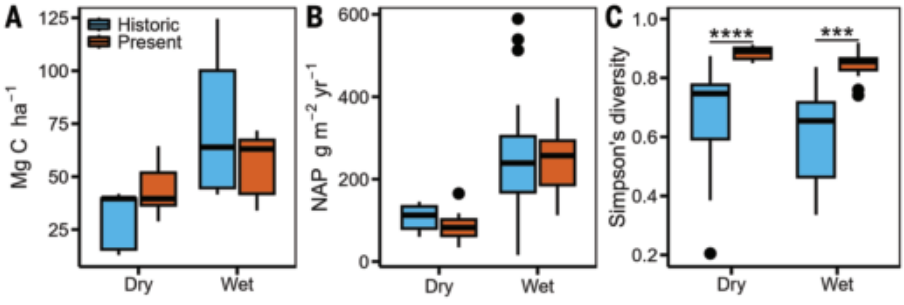


Fig. 4. The migratory landscape maintained soil carbon levels and herbaceous plant production, while plant communities became more diverse over several decades. (**A**) Soil carbon levels per hectare measured in the 1990s to 2000s (*45*, *48*) and (**B**) net aboveground plant production (NAP) measured in the 1980s to 1990s (*41*, *49*) were consistent with the conditions observed during 2015–2021. (**C**) However, plant diversity increased from the 1960s (*50*) to 2017–2021 (table S21). Soil and plant measurements were taken at different sites among periods generally reflecting dry and wet habitats across the migratory landscape. In each panel, the box represents the interquartile range, and asterisks (***$P < 0.001$, ****$P < 0.0001$) indicate significant differences based on pairwise comparisons.

through grazing, particularly in ecosystems that support grazing lawns (*7*, *18*, *19*, *37*, *47*). We extended this concept to estimate how such effects scale across a large migratory landscape by developing a linear model relating aboveground plant nitrogen pools to the Landsat-derived integrated normalized difference vegetation index (INDVI) and grazing treatment. We fitted the model to plot-level data (Fig. 3E and table S20) and applied it to predict aboveground nitrogen pool size across the landscape under grazed and ungrazed conditions. When converted to crude protein (*46*), grazing increased crude protein availability from approximately 3.45 million to 6.67 million kg.

Our findings suggest that the return of freely moving bison is actively stimulating and transforming Yellowstone grasslands. Landscape heterogeneity governs these dynamics. Wet, nutrient-rich soils foster intense grazing and nutrient acceleration, whereas drier areas receive less grazing and exhibit weaker nutrient effects. This produces strong spatial variation in grazing patterns and ecosystem processes, beyond what climate or topography alone would generate. Our work shows that spatial heterogeneity is an intrinsic feature of the migratory bison system. As this system continues to change—with bison now dominating the large herbivore guild (Fig. 1A)—cross-sectional data across multiple decades indicate that the landscape remains resilient. Soil carbon and NAP measured in the 1990s to 2000s (*45*, *48*) and 1980s to 1990s (*41*, *49*), respectively, were consistent with observations from 2015 to 2021 (Fig. 4, A and B). Plant diversity increased from levels recorded in the 1960s (*50*) to those measured from 2017 to 2021 (Fig 4C). Bison appear to be sustaining a dynamic grassland system that challenges the conventional emphasis for managing toward fixed, desired plant communities.

We expect that in grazing lawn habitats where herbivore engineering is the strongest, additional influences may exist, such as top-down control over plant composition, fire suppression, and increased diversity of smaller animals. Such effects are well documented in systems with large herbivores, which exert outsized effects on ecosystems (*24*, *25*). Additionally, owing to the substantial overlap of bison with other migrating herbivores (fig. S6) and increased available crude protein, we suspect that interspecific interactions such as facilitation are occurring in Yellowstone, similar to those identified in Serengeti National Park (*51*), but this has not yet been determined.

The Serengeti provides a valuable reference point, where a fully protected migratory route supports wildebeest near carrying capacity (*21*) and relatively stable predator populations (*52*). By contrast, only part of the Yellowstone bison migration route is protected, and removals outside the park limit bison densities (*39*, *53*). Yet bison are reshaping nutrient flow, producing effects absent or weaker in more constrained systems where herbivores are tightly regulated or fenced (*20*, *54*–*56*).

The claim that large herbivores have outsized effects on ecosystems is true (*24*). However, our findings reinforce that it is not just body size that matters. Large numbers, high densities, and the freedom to move across expansive landscapes are all critical (*57*). Whether it is 1.2 million wildebeest that shape the 25,000-km^2 Serengeti ecosystem, or a half-million caribou migrating >1200 km annually in the Alaskan tundra, the ecological function of migratory herbivores depends fundamentally on their ability to move across landscapes.

To move forward with conserving migratory herbivores and grassland ecosystems, we must embrace landscape-scale heterogeneity—not at the scale of individual ranches or pastures, but at sizes that allow for thousands of migrating large herbivores to move freely across the landscape. Our findings emphasize that ecosystems with large native herbivores, such as bison, can function successfully in today's world. Although challenges remain because of habitat fragmentation and sociopolitical constraints, restoring bison at this scale is not only possible—it works.

REFERENCES AND NOTES

1. C. H. Freese *et al.*, *Biol. Conserv.* **136**, 175–184 (2007).
2. E. W. Sanderson *et al.*, *Conserv. Biol.* **22**, 252–266 (2008).
3. W. T. Hornaday, *The Extermination of the American Bison* (Smithsonian Institution Press, reissue ed., 2002).
4. A. C. Isenberg, *The Destruction of the Bison* (Cambridge Univ. Press, 2000).
5. National Bison Association, "Bison by the numbers—From plains to plates: Insights from bison data & stats" (2025); https://nationalbison.org/bison-by-the-numbers/.
6. D. C. Hartnett, K. R. Hickman, L. E. Walter, *J. Range Manage.* **49**, 413–420 (1996).
7. A. K. Knapp *et al.*, *Bioscience* **49**, 39–50 (1999).
8. B. Ling, E. J. Raynor, D. G. Goodin, A. Joern, *Remote Sens.* **11**, 1364 (2019).
9. Z. Ratajczak *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2210433119 (2022).
10. N. V. Anguiano, K. M. Freeman, J. D. Figge, J. H. Hawkins, L. H. Zeglin, *Biogeochemistry* **167**, 759–773 (2024).
11. E. J. Dyksterhuis, *J. Range Manage.* **2**, 104–115 (1949).
12. M. Westoby, B. Walker, I. Noy-Meir, *J. Range Manage.* **42**, 266–274 (1989).
13. J. L. Holechek, H. Gomez, F. Molinar, D. Galt, *Rangelands* **21**, 12–16 (1999).
14. D. A. Pyke, J. E. Herrick, P. Shaver, M. Pellant, *J. Range Manage.* **55**, 584–597 (2002).
15. D. D. Briske, S. D. Fuhlendorf, F. E. Smeins, *Rangeland Ecol. Manag.* **58**, 1–10 (2005).
16. S. D. Fuhlendorf, D. M. Engle, *Bioscience* **51**, 625–632 (2001).
17. S. D. Fuhlendorf, D. M. Engle, R. D. Elmore, R. F. Limb, T. G. Bidwell, *Rangeland Ecol. Manag.* **65**, 579–589 (2012).
18. D. G. Milchunas, W. K. Lauenroth, *Ecol. Monogr.* **63**, 327–366 (1993).
19. D. A. Frank, S. J. McNaughton, B. F. Tracy, *Bioscience* **48**, 513–521 (1998).
20. S. E. Koerner *et al.*, *Ecology* **95**, 808–816 (2014).
21. J. G. C. Hopcraft *et al.*, in *Serengeti IV: Sustaining Biodiversity in a Coupled Human-Natural System*, A. R. E. Sinclair, K. L. Metzger, S. A. R. Mduma, J. M. Fryxell, Eds. (Univ. of Chicago Press, 2015), pp. 125–174.
22. T. D. Mipam, L. L. Zhong, J. Q. Liu, G. Miehe, L. M. Tian, *Front. Plant Sci.* **10**, 925 (2019).
23. C. Yang *et al.*, *Plant Soil* **444**, 239–250 (2019).
24. R. M. Pringle *et al.*, *Curr. Biol.* **33**, R584–R610 (2023).
25. J. Trepel *et al.*, *Nat. Ecol. Evol.* **8**, 705–716 (2024).
26. S. J. McNaughton, *Science* **191**, 92–94 (1976).
27. S. J. McNaughton, *Ecol. Monogr.* **55**, 259–294 (1985).
28. R. W. Ruess, S. J. McNaughton, *Oikos* **49**, 101–110 (1987).
29. R. W. Ruess, S. W. Seagle, *Ecology* **75**, 892–904 (1994).
30. D. A. Frank, R. D. Evans, *Ecology* **78**, 2238–2248 (1997).
31. D. A. Frank, P. M. Groffman, *Ecology* **79**, 2229–2241 (1998).
32. R. D. Bardgett, D. A. Wardle, *Ecology* **84**, 2258–2268 (2003).
33. D. A. Wardle *et al.*, *Science* **304**, 1629–1633 (2004).
34. J. M. Fryxell, A. R. E. Sinclair, *Trends Ecol. Evol.* **3**, 237–241 (1988).
35. J. A. Merkle *et al.*, *Proc. Biol. Sci.* **283**, 20160456 (2016).
36. C. Geremia *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 25707–25713 (2019).
37. S. J. McNaughton, *Am. Nat.* **124**, 863–886 (1984).
38. D. E. Huff, J. D. Varley, *Ecol. Appl.* **9**, 17–29 (1999).
39. P. J. White, R. L. Wallen, C. Geremia, J. J. Treanor, D. W. Blanton, *Biol. Conserv.* **144**, 1322–1334 (2011).
40. Materials and methods are available as supplementary materials.
41. D. A. Frank, S. J. McNaughton, *Ecology* **73**, 2043–2058 (1992).
42. E. W. Hamilton III, D. A. Frank, *Ecology* **82**, 2397–2402 (2001).
43. E. W. Hamilton III, D. A. Frank, P. M. Hinchey, T. R. Murray, *Soil Biol. Biochem.* **40**, 2865–2873 (2008).
44. D. A. Frank, R. L. Wallen, E. W. Hamilton III, P. J. White, J. D. Fridley, *J. Ecol.* **106**, 434–443 (2018).
45. D. J. Augustine, D. A. Frank, *Ecology* **82**, 3149–3162 (2001).
46. N. Sriperm, G. M. Pesti, P. B. Tillman, *J. Sci. Food Agric.* **91**, 1182–1186 (2011).
47. S. J. McNaughton, F. F. Banyikwa, M. M. McNaughton, *Science* **278**, 1798–1800 (1997).
48. P. Chuckran, D. A. Frank, *Environ. Res. Lett.* **8**, 044013 (2013).
49. D. A. Frank, M. M. Kuns, D. R. Guido, *Ecology* **83**, 602–606 (2002).
50. D. B. Houston, *The Northern Yellowstone Elk: Ecology and Management* (MacMillan, 1982).
51. T. M. Anderson *et al.*, *Science* **383**, 782–788 (2024).
52. M. E. Craft, K. Hampson, J. O. Ogutu, S. M. Durant, in *Serengeti IV: Sustaining Biodiversity in a Coupled Human-Natural System*, A. R. E. Sinclair, K. L. Metzger, S. A. R. Mduma, J. M. Fryxell, Eds. (Univ. of Chicago Press, 2015), pp. 419–447.
53. G. E. Plumb, P.J . White, M. B. Coughenour, R. L. Wallen, *Biol. Conserv.* **142**, 2377–2387 (2011).
54. C. de Mazancourt, M. Loreau, L. Abbadie, *Ecol. Appl.* **9**, 784–797 (1999).
55. E. S. Forbes *et al.*, *Funct. Ecol.* **33**, 1597–1610 (2019).
56. A. K. Knapp *et al.*, *J. Plant Ecol.* **5**, 357–365 (2012).
57. K. Joly *et al.*, *Sci. Rep.* **9**, 15333 (2019).
58. C. Geremia *et al.*, *Ecol. Appl.* **24**, 346–362 (2014).
59. K. L. Adair, E. Schwartz, *Microb. Ecol.* **56**, 420–426 (2008).
60. C. Geremia, E. W. Hamilton, J. A. Merkle, Dryad (2025); https://doi.org/10.5061/dryad.5tb2rbpdx.

ACKNOWLEDGMENTS

We thank R. Wallen and P. J. White, National Park Service (retired), for establishing this long-term monitoring research; D. A. Frank for conceptualization and initial methodology; I. Freeman with the Wyoming Migration Initiative for cartography work; M. Caldwell for extracting migration sequences; the InterTribal Buffalo Council for collecting plant and soil samples; undergraduate students with Washington and Lee University for collecting and analyzing soil and plant samples; students as citizen scientists from local, regional, and indigenous schools with Ecology Project International for collecting soil and plant data; and National Park Service staff for data collection and installation and maintenance of grazing exclosures. **Funding:** National Park Service (USA) and Washington and Lee University Summer Lenfest Grant (USA) to E.W.H. **Author contributions:** Conceptualization: C.G., E.W.H.; Funding acquisition: C.G.; Investigation: C.G., E.W.H., J.A.M.; Methodology: C.G., E.W.H.; Project administration: C.G.; Supervision: C.G.; Visualization: C.G., E.W.H., J.A.M.; Writing – original draft: C.G., E.W.H., J.A.M.; Writing – review & editing: C.G., E.W.H., J.A.M. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data and code are available at Dryad (*60*). **License information:** Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adu0703
Materials and Methods; Figs. S1 to S7; Tables S1 to S23; References (*61*–*70*); MDAR Reproducibility Checklist

Submitted 22 October 2024; accepted 24 July 2025

10.1126/science.adu0703

PROTEOSTASIS

Architecture of the UBR4 complex, a giant E4 ligase central to eukaryotic protein quality control

Daniel B. Grabarczyk[1]*, Julian F. Ehrmann[1,2]†‡, Paul Murphy[1], Woo Seok Yang[3], Robert Kurzbauer[1], Lillie E. Bell[1,2], Luiza Deszcz[1], Jana Neuhold[4], Alexander Schleiffer[1], Alexandra Shulkina[2,5,6], Juyeon Lee[3], Jin Seok Shin[3], Anton Meinhart[1], Gijs A. Versteeg[5,6], Eszter Zavodszky[7], Hyun Kyu Song[3], Ramanujan S. Hegde[7], Tim Clausen[1,8]*

Eukaryotic cells have evolved sophisticated quality control mechanisms to eliminate aggregation-prone proteins that compromise cellular health. Central to this defense is the ubiquitin-proteasome system, where UBR4 acts as an essential E4 ubiquitin ligase, amplifying degradation marks on defective proteins. Cryo–electron microscopy analysis of UBR4 in complex with its cofactors KCMF1 and CALM1 reveals a massive 1.3-megadalton ring structure, featuring a central substrate-binding arena and flexibly attached catalytic units. Our structure shows how UBR4 binds substrate and extends lysine-48–specific ubiquitin chains. Efficient substrate targeting depends on both preubiquitination and specific N-degrons, with KCMF1 acting as a key substrate filter. The architecture of the E4 megacomplex is conserved across eukaryotes, but species-specific adaptations allow UBR4 to perform its precisely tuned quality control function in diverse cellular environments.

Protein quality control (PQC) relies on the ubiquitin-proteasome system (UPS) to tag misfolded or mislocalized proteins for destruction before they aggregate and damage cells (*1*). Most ubiquitin E3 ligases recognize well-defined degrons (*2*), but PQC ligases must detect a broad spectrum of aberrant states (*3*) and often cooperate with E4 enzymes (*4*) that extend ubiquitin chains to reinforce the degradation signal (*5*, *6*). Alteration of this surveillance network underlies neurodegeneration, myopathies, and cancer (*7*, *8*). Aneuploidy, for example, disrupts subunit stoichiometry, generates orphan proteins, and imposes proteotoxic stress, highlighting the importance of PQC ligases as potential therapeutic targets (*9*, *10*).

Among the >600 human ubiquitin ligases, UBR4 plays a particularly important PQC role, being an essential and conserved E3/E4 ligase found across eukaryotes (*11*). UBR4 is critical for maintaining proteostasis in long-lived cells with high metabolic demands, such as neurons and muscle cells, where proteotoxic stress is particularly harmful (*12*–*15*). It recognizes several stress signatures, including mitochondrial targeting sequences (MTSs) on mislocalized mitochondrial precursors, orphan proteins resulting from unassembled complexes, and aggregation-prone proteins (*16*–*19*). Additionally, UBR4 plays a key role in autophagic processes, bridging proteasomal and lysosomal degradation pathways to ensure efficient clearance of aberrant proteins (*15*). The diverse functions of UBR4 underscore its role as a central hub in the PQC network, capable of integrating multiple stress response pathways. Aside from teaming up with various ubiquitin ligases, UBR4 activity is linked to two specific cofactors: calcium-binding protein calmodulin 1 (CALM1) and potassium channel modulatory factor 1 (KCMF1) (*13*, *16*, *17*, *20*, *21*).

How CALM1 and KCMF1 together modulate UBR4 selectivity and its reported E4 chain-extension activity remains poorly understood. To resolve these points, we reconstituted human, nematode, and plant UBR4 assemblies, defined their architectures, and characterized their degradation labeling function, providing a comparative framework for UBR4's central role in PQC and disease.

Human UBR4 assembles a ring-shaped ubiquitination arena

We reconstituted the human E4 complex by coexpressing UBR4, KCMF1, and CALM1 in insect cells and purifying the ternary complex through a tag on UBR4. Mass photometry revealed that the isolated complex had a molecular weight of approximately 1.5 MDa, consistent with a dimer of heterotrimers (fig. S1A).

Focused cryo–electron microscopy (cryo-EM) classification and refinement yielded density maps that resolved all folded regions of the human UBR4 complex ($HsUBR4_2$/$KCMF1_2$/$CALM1_2$) (Fig. 1 and figs. S2 to S4), except the flexible C-terminal extension bearing the catalytic hemi-RING E3 module (*22*). Two UBR4 molecules dimerize through two remote, structurally different interfaces to form a large ring-shaped assembly (Fig. 1). The first interface is composed by the N-terminal domain, involving extensive contacts between adjacent Armadillo repeats. The second, larger interface is formed by Armadillo repeats in the C-terminal portion of UBR4, with laterally aligned helices sealing this contact. CALM1 and KCMF1 bind at this C-terminal interface, as predicted for the trimeric subcomplex (*16*). The C-terminal helix of KCMF1 inserts like a pin into a hole in the UBR4 Armadillo repeat and is covered by a small lid insertion within this region, which also mediates CALM1 binding. This arrangement implies that the proper folding of the Armadillo repeat scaffold and its associated lid depends on KCMF1. Indeed, coexpressing UBR4 with a KCMF1 variant having a mutated pin helix (L318A/F319A/V320S) yields monomeric UBR4 protein (fig. S1B). CALM1, known to mediate calcium regulation of target proteins, engages UBR4 in a canonical manner, with a long hydrophobic CALM1-interacting helix (CIH) of UBR4 docking into the cleft between its two lobes (Fig. 1) (*23*). Although the C-lobe of CALM1 adopts a calcium-bound conformation (fig. S1C), treatment with the calcium-chelator EGTA did not affect the CALM1 occupancy, suggesting that the cofactor is stably bound rather than acting as a reversible regulator (fig. S2). Still, its precise function in the UBR4 complex needs to be discovered.

The structural motifs that constitute the ringlike scaffold of the E4 complex were well defined by EM density, but the functional domains lining the inner cavity or extending outward into the periphery displayed high flexibility, as reflected by their lower local resolution (figs. S1D, S2, and S4). The largest intrusion into the E4 cavity is a multidomain appendage formed by the β propeller (BP), the UBR box, and the β sandwich 1 (BS1) of UBR4 together with the N-terminal zinc binding domains of KCMF1 (Fig. 1). A second appendix protrudes from the C-terminal dimerization region, where a zinc finger domain positions another β sandwich (BS2) and two associated zinc fingers near the central cavity. Lastly, an extension formed by the C-terminal Armadillo repeats projects outwards from the UBR4 arena. This extension houses a ubiquitin-like (UBL) domain and a flexibly tethered hemi-RING module, the catalytic E4 domain, which was unresolved in the EM map.

[1]Research Institute of Molecular Pathology, Vienna BioCenter (VBC), Vienna, Austria. [2]Vienna BioCenter PhD Program, Doctoral School of the University of Vienna and Medical University of Vienna, Vienna BioCenter (VBC), Vienna, Austria. [3]Department of Life Sciences, Korea University, Seoul, South Korea. [4]Vienna Biocenter Core Facilities, Vienna BioCenter, Vienna, Austria. [5]Max Perutz Labs, Vienna Biocenter (VBC), Vienna, Austria. [6]University of Vienna, Center for Molecular Biology, Department of Microbiology, Immunobiology, and Genetics, Vienna, Austria. [7]MRC Laboratory of Molecular Biology, Cambridge, UK. [8]Medical University of Vienna, Vienna, Austria. ***Corresponding author: daniel.grabarczyk@imp.ac.at (D.B.G.); tim.clausen@imp.ac.at (T.C.)** †Present address: Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA, USA. ‡Present address: Boston Children's Hospital, Boston, MA, USA.

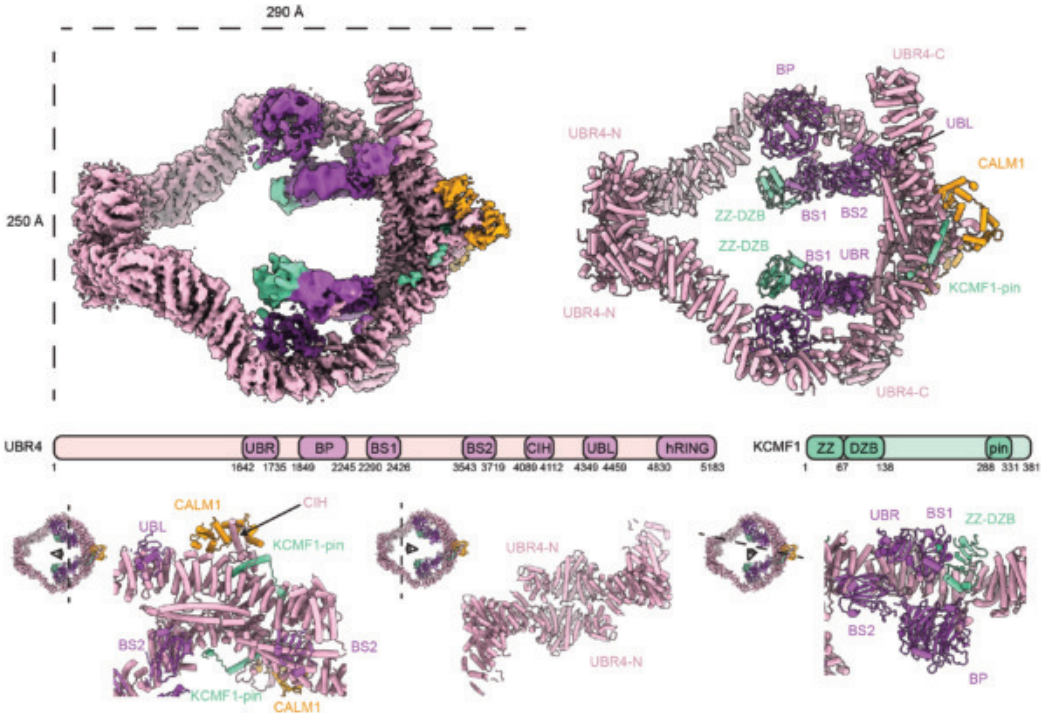


Fig. 1. Architecture of the human UBR4 complex. (Top) Composite cryo-EM density map and model colored by protein component and domain as indicated. (Middle) A schematic domain architecture of UBR4 and KCMF1. (Bottom) Detailed views of structural features.

Ubiquitin K48 chain extension within the UBR4 ring

To investigate how the distinct functional motifs of the UBR4 complex support ubiquitination, we devised an E4 assay monitoring the conjugation of the two ubiquitin variants Ub* and Ub-K0 (Fig. 2A). Ub* (ubiquitin-ΔGG) lacks the C-terminal diglycine motif required for ubiquitin transfer, whereas Ub-K0 has all lysine residues mutated to arginine. This setup enabled us to follow a single ubiquitination event, where Ub-K0 is transferred onto Ub*, and then no further reaction can occur. As expected, a Ub-K0–Ub* band occurred only in the presence of both ubiquitin variants and UBR4 (Fig. 2A), offering a specific readout of E4 activity, ideal for mutational analyses and systematic substrate screens.

We had previously shown that UBR4 only ubiquitinates orphan proteins that have already been ubiquitinated by other E3 ligases (*16*). This E4 activity depends on the UBL domain, which was predicted by AlphaFold3 (fig. S5A) (*24*), to bind ubiquitin and orient K48 toward the E2~Ub conjugate at the hemi-RING, favoring K48-linked chain extension on preubiquitinated substrates (*16*). In line with this, formation of free ubiquitin chains by the UBR4 complex required K48 of ubiquitin (fig. S5B). To probe the UBL/Ub interface, we mutated three residues in the predicted binding site of the UBL domain (UBL-3A). Analytical size-exclusion chromatography (fig. S5C), which was performed with the isolated and stably folded UBL variants (fig. S5D), showed that wild-type UBL interacted with ubiquitin, whereas the UBL-3A variant did not.

We then incubated the UBR4 complex with its cognate E2 enzyme (UBE2A) and analyzed the resulting state by cryo-EM (fig. S6 and table S2). The reconstruction contained extra density for UBE2A, enabling us to model the E2 bound to the C-terminal extension of UBR4, despite the low resolution in this region (Fig. 2B). As predicted by AlphaFold3, the E2 is bound in a backside orientation, stabilized by a specific β hairpin structure inserted within an extended helix of the C-terminal protrusion of UBR4 (*16*). The largely hydrophobic interface covers two sides of UBE2A yet leaves its canonical RING-binding face free for hemi-RING engagement. As a consequence, the E2 becomes precisely positioned near the UBL domain at the narrow constriction site of the substrate binding arena. In our cryo-EM map, the catalytic hemi-RING is not engaged in direct contacts with UBE2A but appears as diffuse density above the E2 (fig. S6), consistent with its reported low affinity for UBE2A (*22*). We therefore modeled the ubiquitin transfer state by fitting the UBE2A-UBR4(hemi-RING) crystal structure (*22*) onto UBE2A of our cryo-EM model (Fig. 2C). Further, we used the yeast UBR1-UBE2A-Ub structure to model the donor ubiquitin (*25*), whereas the acceptor ubiquitin was placed using the AlphaFold3-predicted UBL/Ub complex (fig. S5A). In this model, guided by our cryo-EM structure, UBE2A and the UBL domain align the acceptor ubiquitin so that its K48 residue points toward the E2~Ub thioester at the hemi-RING, thus favoring K48 ubiquitin chain extension. Consistent with this putative E4 mechanism, Ub-K0–Ub* formation was abolished by the UBL-3A mutation, deletion of the hemi-RING (ΔRING), or by disrupting the backside interface of UBE2A (E2B-5A) (Fig. 2D). In native substrates, the UBL-bound ubiquitin would be covalently attached through its C terminus to a lysine residue in the target protein. As the ubiquitin C terminus points into the central cavity of the UBR4 complex, potential substrate recognition sites should reside within this region (Fig. 2E).

Selection of Ub-marked proteins to be degraded

The UBR4 complex has mobile E4 catalytic arms on top of a huge substrate binding arena. This organization is reminiscent of the giant ubiquitin ligases BIRC6, HUWE1 and UBR5, which use an array of receptor domains to target diverse substrates (fig. S7A) (*26*–*32*). While analyzing our cryo-EM data, we noted additional low-resolution density inside the central cavity of the UBR4 complex (Fig. 3A). As we could assign all ordered domains of the UBR4 complex, we suspected that this extra density represents a copurified substrate bound in the inner cavity, visualizing the mechanism of substrate recognition. The density was connected to the E4 by the ZZ-DZB motifs of KCMF1, hinting that these domains act as substrate receptors. To test this idea, we reconstructed a UBR4 complex lacking ZZ-DZB and determined its cryo-EM structure using the same workflow as for the wild-type complex (fig. S8 and table S2). In the resulting map, the ZZ-DZB domains were absent as was the connected extra density (Fig. 3A). To identify captured substrates, we performed a mass spectrometry (MS) analysis searching for proteins that copurified with wild-type E4 but were lost in the ΔZZ-DZB sample (Fig. 3B). Single-stranded DNA binding protein (SSBP1), a mitochondrial protein previously reported to bind to UBR4/KCMF1 in human cells (*13*), was the most enriched protein in the wild-type complex as compared to the ΔZZ-DZB deletion. Many other proteins preferentially retained in wild-type UBR4 were mitochondrial, consistent with reports that UBR4 participates in mitochondrial PQC (*13*, *17*). By contrast, deleting other putative receptors, such as the BS2 or UBR domains, did not enrich obvious targets (fig. S7B).

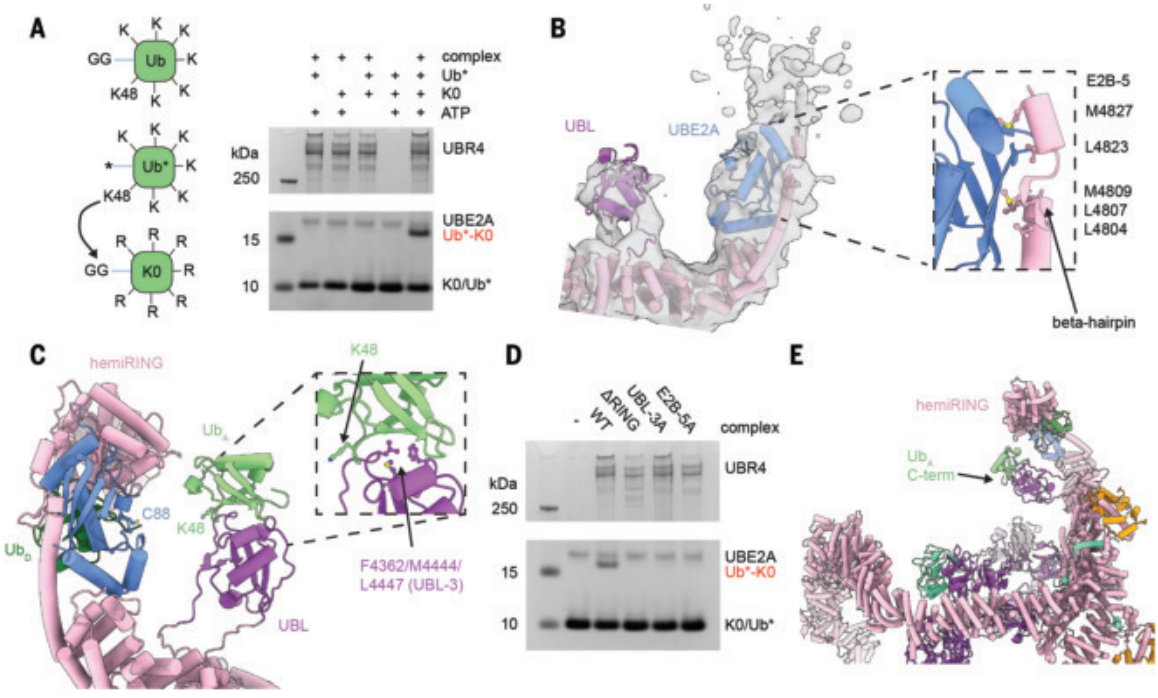


Fig. 2. Mechanism of ubiquitin chain extension by the UBR4 complex. (**A**) (Right) E4 ligase assay showing formation of Ub-K0–Ub* diubiquitin from Ub-K0 and Ub* catalyzed by 200 nM human UBR4 complex when all components of the assay are added for 45 min. (Left) A schematic explaining the assay and the two ubiquitin components. (**B**) Cryo-EM density map and model of the structure of human UBR4 complex obtained in the presence of UBE2A. (Inset) A detailed view of the interaction highlighting hydrophobic residues on the E2-binding helix of UBR4, which are mutated in the E2B-5A UBR4 complex variant. (**C**) Model of the E4 transfer state using our structure overlaid with a crystal structure of the hemiRING-UBE2A complex (PDB ID pdb_00008BTL) and the cryo-EM structure of the yeast UBR1-UBE2A-Ub_D complex (PDB ID pdb_00007MEX), both aligned on UBE2A, and an AlphaFold3 model of the UBL-Ub interaction (fig. S5A) aligned on the UBL domain. The three hydrophobic residues on the UBL domain that interact with ubiquitin and are mutated in the UBL-3A variant are indicated. (**D**) E4 ligase assay as in (A) with 100 nM of the indicated variant complexes for 45 min. WT, wild type. (**E**) Zoomed-out image of the E4 transfer state model showing the position of the C-terminal tail of Ub relative to the UBR4 arena. Single-letter abbreviations for the amino acid residues referenced throughout the text and figures are as follows: L, Leu; A, Ala; F, Phe; V, Val; S, Ser; R, Arg; C, Cys; K, Lys; M, Met; G, Gly; Q, Gln; T, Thr; P, Pro; D, Asp; E, Glu; I, Ile; Y, Tyr.

The UBR4 complex has been implicated in targeting unimported mitochondrial precursors that still carry their unprocessed MTS (*13*, *17*). To ask whether the insect cell (*Trichoplusia ni)* proteins co-purifying with the UBR4 complex retained an MTS, we applied a tailored MS analysis mapping their N-terminal sequences (fig. S7C). Some hits, such as SSBP1 and ACADS, had a processed MTS, whereas others, such as NDUFA5 and MRPS17, contained an intact signal sequence. To test potential substrates in ubiquitination assays, we produced recombinant variants of the identified proteins with the detected N termini. We generated each variant with and without a C-terminally fused Ub* to explore whether the UBR4 complex can extend chains on preubiquitinated substrates (*16*, *33*), or whether it relies on KCMF1 as priming E3 ligase (*34*). To simplify the read-out, we used K0 ubiquitin, which prevents chain formation. Robust ubiquitination was only observed for Ub*-fused TnSSBP1, TnMRPS17, and TnNDUFA5 but not for their unfused counterparts (Fig. 3C), indicating that the UBR4 complex acts as a chain-extending E4 ligase, not a priming ligase. Mutating individual lysines on ubiquitin showed that generated chains were K48 linked (fig. S9A). In a control experiment, we used HUWE1 as priming E3 ligase and monitored the ubiquitination of TnNDUFA5, which was predicted to be a good HUWE1 substrate (*28*). In contrast to KCMF1, HUWE1 was able to initiate ubiquitin chains on TnNDUFA5 (fig. S9B), which were further extended by the UBR4 complex (fig. S9C). These data suggest that KCMF1 does not function as priming E3 ligase in the UBR4 complex. Consistently, KCMF1 lacks a known E3 domain (fig. S9D), did not interact with any E2 in an AlphaFold pulldown screen (fig. S9E), and did not catalyze ubiquitination with any E2 we tested (fig. S10).

To pinpoint substrate features recognized by the UBR4 complex, we adapted our E4 assay. Appending two model MTSs through a linker onto the N terminus of Ub* enhanced K48 chain extension compared with that of the MTS-free control (Fig. 3D). Thus, substrate features beyond ubiquitin itself can enhance recruitment to the UBR4 complex. Because the ZZ-DZB motif of KCMF1 has been implicated in binding N-terminal arginine (^{N}R) N-degrons (*35*), we measured the binding of the isolated ZZ-DZB domain to synthetic peptides by isothermal titration calorimetry (ITC) (fig. S11A). No interaction was observed for the reported ligand R1-C2 sulfonate, ^{N}RC(SO_3), and, consistently, a cocrystal structure of the ZZ-DZB:peptide complex showed both R1 and C(SO_3)2 sidechains protruding out of the binding pocket (Fig. 3E and fig. S11B). Screening additional peptides revealed tight binding to an ^{N}RT motif, and the cocrystal structure captured specific contacts with both the T2 side chain and the N-terminal amino group (Fig. 3E, fig. S11C, and table S3). ITC binding studies showed that the second residue is more important than the first, as an ^{N}LT peptide also interacted with the ZZ-DZB domains, but ^{N}RL, ^{N}RR, or ^{N}RA peptides did not (fig. S11A). A threonine at position 2 is characteristic of many processed mitochondrial proteins, for example, identified TnSSBP1 (^{N}ST) and TnACADS (^{N}FT); previously copurified human proteins, such as ABHD10 (^{N}KT), SARS2 (^{N}TT), ACOT9 (^{N}LT) (*13*); and the Dengue virus protein NS5 (^{N}GT) binding UBR4 (*36*). In agreement with these data, E4 assays showed stronger chain extension on a substrate with an ^{N}RT motif compared with that of other N-termini (Fig. 3F and fig. S12A). Determining whether this specificity was due to the ZZ-DZB motif was complicated, as the deletion produced more free chains for unknown

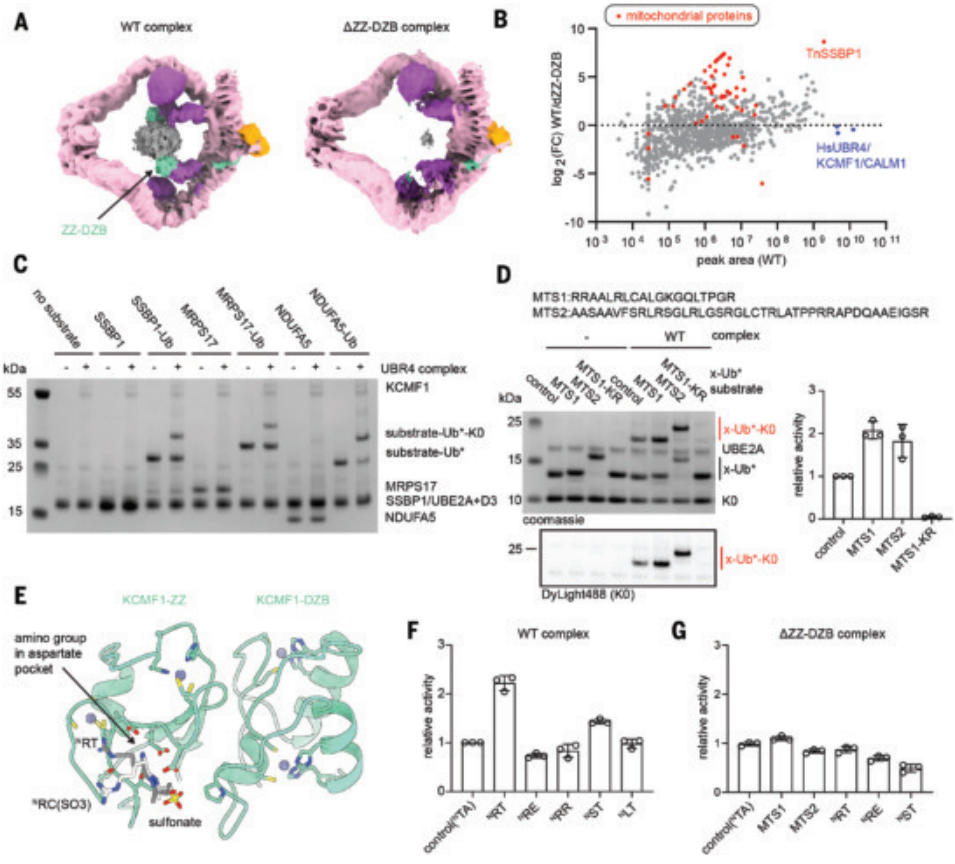


Fig. 3. KCMF1 mediates selection of UBR4 substrates. (**A**) (Left) A cryo-EM map of the human UBR4 complex structure refined with a global mask showing fuzzy central substrate density. (Right) The ΔZZ-DZB UBR4 complex refined with the same approach. (**B**) MS analysis of insect cell proteins copurified with the wild-type and ΔZZ-DZB UBR4 complexes, where peak intensity in the wild-type is plotted against $\log_2$(fold change) between the two datasets to identify proteins that are present in the wild-type but not the ΔZZ-DZB UBR4 complex. The core complex components are colored blue, and mitochondrial proteins are colored red. (**C**) Ubiquitination assay using 200 nM UBR4 complex, UBE2A, UBE2D3, and K0-Ub for 30 min in the presence of 2 μM of the indicated recombinantly purified insect cell proteins with or without Ub* fused by a linker to the C terminus. (**D**) E4 ligase assay with 200 nM of the human UBR4 complexes for 30 min with the indicated MTS or control with no MTS fused through a linker to the N terminus of Ub*. KR indicates a K48R mutation in the fused Ub*. Reactions contained 1:10 DyLight488 Ub-K0, which was used to quantify product bands relative to the control substrate and plot these on the right (± SD). (**E**) Two overlaid crystal structures of the $KCMF1^{ZZ+DZB(\triangle linker)}$ domain with either a $^{N}RC(SO_3)K$ peptide or a $^{N}RTGG$ peptide. Only the two N-terminal residues are shown for clarity, with notable residues on KCMF1 shown in stick representation. (**F** and **G**) Quantified E4 ligase assays (± SD) as in (D) with the indicated substrates for (F) 200 nM WT UBR4 complex for 30 min and (G) 200 nM ΔZZ-DZB UBR4 complex for 15 min. Assay gels are shown in fig. S12, A and B.

reasons (fig. S9F). Nevertheless, its specificity profile differed from that of the wild-type, showing little preference among control, MTS, or N-degron–containing constructs but still showing reduced activity on ^{N}RE and ^{N}ST termini (Fig. 3G and fig. S12B). Together, these data identify the ZZ-DZB as a major substrate receptor and that other domains in the E4 arena might finetune specificity.

Evolutionary conservation of the UBR4 ubiquitination arena

UBR4 plays important roles in various eukaryotes, including *Arabidopsis thaliana* (*37*, *38*) and *Caenorhabditis elegans* (*39*), yet its partnership with KCMF1 and calmodulin has not been demonstrated outside animals. To test whether the human E4 architecture and mechanism represents a universal PQC solution, we first reconstituted the *C. elegans* UBR4 complex by coexpressing CeUBR4, CeKCMF1 (ZK652.6), and CeCALM. The purified complex had a molecular weight of 1.1 MDa, consistent with a dimer of trimers (fig. S13A). Activity assays showed that the CeUBR4 complex preferred MTS-Ub* substrates over Ub* alone but showed only weak selectivity for the ^{N}RT-Ub* degron (fig. S13B).

To explore structural differences, we determined the cryo-EM structure of CeUBR4 (Fig. 4A, figs. S14 to S16, and table S4). Although the nematode E4 complex preserves the ring shape of the human enzyme, it displays notable adaptations, particularly in the dimerization interfaces that result in different topologies of the ring-shaped arena. Whereas the domains aligning the N-terminal ends exhibit different, unrelated scaffolds, the C-terminal interface is sealed by distinct interaction with the two cofactor proteins. Most notably, in the CeUBR4 complex, CALM1 is absent, and its binding site is filled by insertions in CeKCMF1 (Fig. 4B). The necessity of occupying this interface in the absence of CALM1 suggests a role in maintaining the structural integrity of the E4 complex. Proteomic analysis of purified CeUBR4 pointed to a distinct substrate preference, with mitochondrial medium-chain acyl–coenzyme A dehydrogenase (ACADM) showing stronger enrichment than SSBP1, which is preferred by the human UBR4 complex (fig. S13C and table S5). In line with this finding, we observed well-defined extra density in the central arena of CeUBR4, closely resembling the elongated shape of an orphaned subunit from the tetrameric ACADM complex (Fig. 4C). According to the cryo-EM data, the bound substrate not only contacts the ZZ domain but also two long helices in the N-terminal region. These two helices, which are specific to CeUBR4, protrude from the rigid E4 core into the central cavity, exposing a series of hydrophobic residues (fig. S13D) that may help to recognize and bind the hydrophobic surface characteristic of orphaned subunits.

The observed adaptations of the CeUBR4 complex prompted us to investigate the ancestral design of the UBR4 core further. We turned to *A. thaliana* UBR4 (also known as BIG), which diverged from metazoan UBR4 early in eukaryotic evolution and has distinct roles in hormone signaling (*37*, *38*). Because KCMF1 is critical for E4 assembly, we first performed an in-depth bioinformatic analysis to identify the elusive plant ortholog. Drought-induced protein 19 (DI19) emerged as the best candidate: It shows strong homology to HsKCMF1 in the C-terminal pin region (fig. S13E), has related biological functions to AtUBR4 (*37*, *40*), and was one of the top hits alongside AtUBR4 in an N-degron proximity labeling screen (*38*). Coexpression of AtUBR4, AtDI19, and AtCALM1 followed by cryo-EM analysis on the resulting complex (fig. S17 and table S3) yielded a medium-resolution map of the dimerization interface, which we could model using AlphaFold3.

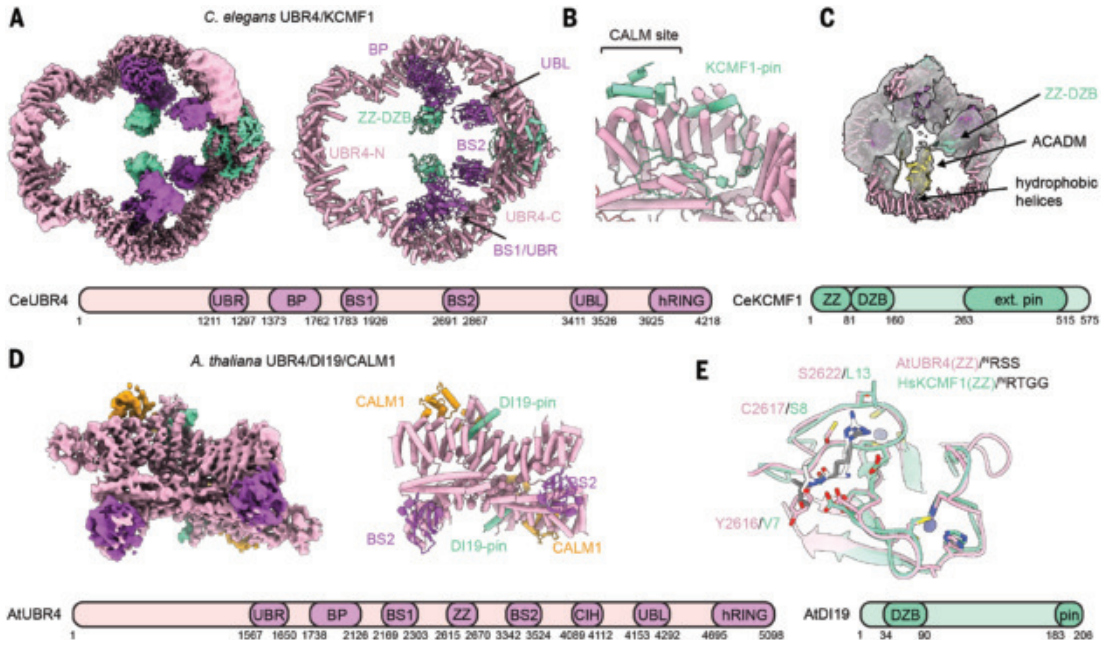


Fig. 4. Evolutionary conservation of the UBR4 complex. (**A**) Cryo-EM composite map and model of the *C. elegans* UBR4 complex colored by subunit and domain, as indicated. The domain architecture is illustrated below. (**B**) Detailed view of the extended interface of CeKCMF1 with CeUBR4. (**C**) Low-resolution cryo-EM map of the CeUBR4 complex from a global three-dimensional classification (fig. S14) showing a bound substrate. An AlphaFold3 model of the copurified protein TnACADM is modeled in the density. (**D**) Cryo-EM map of the dimer interface core of the AtUBR4 complex, colored by domain and subunit, as indicated. The domain architecture derived from bioinformatic comparison with the human UBR4 complex is shown below. (**E**) Crystal structure of the AtUBR4 ZZ domain in complex with an ^{N}RSS peptide superimposed with the HsKCMF1 ZZ domain crystal structure in complex with an N-RTGG peptide. Sequence differences in the N-degron binding pocket are highlighted.

The resulting structure confirmed a highly similar arrangement to the human complex (Fig. 4D and fig. S17). Despite this structural similarity, one notable difference is that the DI19 cofactor lacks the ZZ domain. Instead, an internal ZZ domain is found in AtUBR4 itself (fig. S18). To explore whether the plant ZZ domain displays distinct specificity to HsKCMF1-ZZ, we determined the cocrystal structure of the AtUBR4-ZZ in complex with an ^{N}RS-containing peptide (Fig. 4E, fig. S13F, and table S3). We observed multiple differences in the substrate binding pocket—for example, a specific interaction of the peptide's R1 residue with a serine residue in the ZZ motif—whereas the corresponding residue in HsKCMF1 ZZ is a leucine. In agreement with this, ITC showed that the AtUBR4 ZZ domain binds to ^{N}RA and ^{N}RR peptide sequences (fig. S13G), whereas the HsKCMF1 ZZ domain did not. Thus, the AtUBR4 complex could act as a traditional type 1 N-degron receptor, consistent with in vivo studies (*38*).

Discussion

Our data show that the evolutionarily conserved UBR4 complex forms a megadalton E4 arena in which inward-facing ZZ-DZB domains of KCMF1 act not as E3 ligases but as substrate receptors. UBR4 recognizes preubiquitinated substrates based on their orphan protein character (*16*), the presence of an MTS (*17*), or N-degrons (*41*). The pronounced preference for threonine in position 2 points to the targeting of processed mitochondrial proteins, eventually leaking from mitochondria into the cytosol, as can be now explored in vivo. The breadth of these partially degenerate signals accounts for the large number of reported UBR4 substrates.

Although the ZZ-DZB motif of KCMF1 is central for recognition, efficient capture likely depends on multivalent, low-affinity interactions supplied by accessory UBR4 domains, such as the BP, UBR, and BS2, that also line the inner cavity. Organism-specific variations demonstrate this principle: In plants, specificity depends on an internal AtUBR4 ZZ motif, whereas in nematodes, two elongated helices at the N terminus extend into the central cavity to capture substrates, presumably recognizing the hydrophobic surface of orphaned proteins. Beyond domain composition, the ring itself imposes an additional size filter that favors small, orphaned proteins over multiprotein complexes. The functional importance of the ring architecture is emphasized by its conserved shape and dimension in CeUBR4 despite pronounced differences in its helical scaffold and dimerization interfaces. Moreover, UBR4 disease mutations are spread throughout the ring scaffold, highlighting its crucial role for proper PQC function (fig. S19A).

E4 ligases, like UBR4, play a dual role in amplifying ubiquitination signals and refining specificity in protein degradation (*42*, *43*). Rather than relying solely on the priming E3 ligase for substrate selection, our findings emphasize the equally crucial role of E4 ligases in enhancing specificity. Although both chain-initiating E3 ligases and chain-extending E4 ligases individually may exhibit weak substrate specificity, their sequential action yields an effective and specific proofreading mechanism to target aberrant proteins (fig. S19, B and C). Moreover, the likelihood of sequential recognition increases with a substrate's residence time. For example, mitochondrial precursor proteins imported efficiently into mitochondria escape recognition, whereas precursors stalled in the cytosol under stress become marked for destruction. Such temporal control ensures that truly dysfunctional or defective proteins are removed but transient intermediates are spared. Sequential E3 and E4 ligase action should also facilitate the formation of specialized PQC pathways toward distinct substrate types. The HsUBR4 complex operates alongside other broad PQC enzymes, such as HUWE1, BIRC6, UBR5, HERC1, and HERC2 (*18*, *44*–*48*), each targeting distinct yet overlapping sets of substrates. This combinatorial setting allows for fine-tuned regulation of PQC pathways in specific cell types and physiological conditions. Through its intricate structure, selective targeting of preubiquitinated substrates, and synergy with other PQC ligases, UBR4 is thus able to adopt a crucial role

in cellular homeostasis, enabling the precise and efficient elimination of defective proteins.

REFERENCES AND NOTES

1. Y. T. Kwon, A. Ciechanover, *Trends Biochem. Sci.* **42**, 873–886 (2017).
2. L. Buetow, D. T. Huang, *Nat. Rev. Mol. Cell Biol.* **17**, 626–642 (2016).
3. S. Juszkiewicz, R. S. Hegde, *Mol. Cell* **71**, 443–457 (2018).
4. M. Koegl *et al.*, *Cell* **96**, 635–644 (1999).
5. C. Liu, W. Liu, Y. Ye, W. Li, *Nat. Commun.* **8**, 14274 (2017).
6. H. J. Meyer, M. Rape, *Cell* **157**, 910–921 (2014).
7. S. Lipkowitz, A. M. Weissman, *Nat. Rev. Cancer* **11**, 629–643 (2011).
8. A. M. Pickrell, R. J. Youle, *Neuron* **85**, 257–273 (2015).
9. C. M. Brennan *et al.*, *Genes Dev.* **33**, 1031–1047 (2019).
10. E. M. Torres *et al.*, *Science* **317**, 916–924 (2007).
11. S. T. Kim *et al.*, *PLOS ONE* **13**, e0202260 (2018).
12. C. Belzil *et al.*, *J. Biol. Chem.* **288**, 24452–24464 (2013).
13. J. H. Hong *et al.*, *Mol. Cell. Proteomics* **14**, 674–685 (2015).
14. L. C. Hunt *et al.*, *Nat. Commun.* **12**, 1418 (2021).
15. T. Tasaki *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 3800–3805 (2013).
16. S. Carrillo Roas *et al.*, *Mol. Cell* **85**, 815–828.e10 (2025).
17. D. L. Haakonsen *et al.*, *Nature* **626**, 874–880 (2024).
18. R. G. Yau *et al.*, *Cell* **171**, 918–933.e20 (2017).
19. M. Kathiresan, S. Animesh, R. Morris, J. Kreuzer, K. C. Patra, L. Shi, J. Merritt, X. Yin, C. H. Benes, N. Bardeesy, W. Haas Protein interactome homeostasis through an N-recognin E3 ligase is a vulnerability in aneuploid cancer. bioRxiv 2023.05.04.539299 [Preprint] (2023); https://doi.org/10.1101/2023.05.04.539299.
20. Y. Nakatani *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15093–15098 (2005).
21. D. E. Leto *et al.*, *Mol. Cell* **73**, 377–389.e11 (2019).
22. L. Barnsby-Greer *et al.*, *Nat. Struct. Mol. Biol.* **31**, 351–363 (2024).
23. H. Kurokawa *et al.*, *J. Mol. Biol.* **312**, 59–68 (2001).
24. J. Abramson *et al.*, *Nature* **630**, 493–500 (2024).
25. M. Pan *et al.*, *Nature* **600**, 334–338 (2021).
26. L. Dietz *et al.*, *Science* **379**, 1112–1117 (2023).
27. J. F. Ehrmann *et al.*, *Science* **379**, 1117–1123 (2023).
28. D. B. Grabarczyk *et al.*, *Nat. Chem. Biol.* **17**, 1084–1092 (2021).
29. M. Hunkeler, C. Y. Jin, E. S. Fischer, *Science* **379**, 1105–1111 (2023).
30. M. Hunkeler *et al.*, *Mol. Cell* **81**, 3468–3480.e7 (2021).
31. L. A. Hehl *et al.*, *Nat. Chem. Biol.* **20**, 190–200 (2024).
32. Z. Hodáková *et al.*, *EMBO J.* **42**, e113348 (2023).
33. A. Shulkina *et al.*, *Nat. Commun.* **16**, 3894 (2025).
34. Z. Yang, D. L. Haakonsen, M. Heider, S. R. Witus, A. Zelter, T. Beschauner, M. J. MacCoss, M. Rapé, The molecular basis of integrated stress response silencing. bioRxiv, 2024.2012.2002.626349 (2024); https://doi.org/10.1101/2024.12.02.626349.
35. A. J. Heo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2107993118 (2021).
36. J. Morrison *et al.*, *PLOS Pathog.* **9**, e1003265 (2013).
37. P. Gil *et al.*, *Genes Dev.* **15**, 1985–1997 (2001).
38. H. Zhang *et al.*, *Plant Cell* **36**, 3177–3200 (2024).
39. M. M. Rinschen *et al.*, *Hum. Mol. Genet.* **25**, 1328–1344 (2016).
40. S. Maitra Majee, E. Sharma, B. Singh, J. P. Khurana, *Plant Direct* **4**, e00234 (2020).
41. D. E. Jeong *et al.*, *Commun. Biol.* **6**, 1214 (2023).
42. V. Pant, G. Lozano, *Genes Dev.* **28**, 1739–1751 (2014).
43. V. Anton *et al.*, *Mol. Cell* **83**, 2976–2990.e9 (2023).
44. L. D. Cervia *et al.*, *Cancer Discov.* **13**, 766–795 (2023).
45. A. Shulkina, K. Hacker, J. F. Ehrmann, V. Budroni, A. Mandlbauer, J. Bock, D. B. Grabarczyk, L. Cochella, T. Clausen, G. A. Versteed, TRIM52 is a primate-specific player in the DNA repair process under tight proteolytic control by a triad of giant E3 ligases. bioRxiv 2024.2005.2016.594269 [Preprint] (2024); https://doi.org/10.1101/2024.05.16.594269.
46. I. Schwartz, V. Budroni, M. Meyenberg, H. Hornegger, K. Hacker, S. Schwartz, Z. Hodakova, D. B. Grabarczyk, J. F. Ehrmann, S. Scinicariello, D. Haselbach, J. Menche, T. Clausen, G. E. Karagöz, G. A. Versteed, Multiple ubiquitin ligases protect human genome integrity by targeting cancer-associated APOBEC3 deaminases for degradation. bioRxiv 2024.2004.2023.590688 (2024); https://doi.org/10.1101/2024.04.23.590688.
47. Y. Yagita, E. Zavodszky, S. Y. Peak-Chew, R. S. Hegde, *Cell* **186**, 3443–3459.e24 (2023).
48. E. Zavodszky, S. Y. Peak-Chew, S. Juszkiewicz, A. J. Narvaez, R. S. Hegde, *Science* **373**, 998–1004 (2021).

ACKNOWLEDGMENTS

We thank the EM facility of the Vienna BioCenter Core Facilities (VBCF), particularly H. Kotisch for collecting the cryo-EM datasets. We thank the staff at Beamline 5C of the Pohang Accelerator Laboratory, South Korea. We also thank the VBCF Protein Technologies facility, particularly A. Crousilles for help with CD spectroscopy and E. Roitinger, R. Imre, O. Hudecz, and G. Dürnberger from the VBCF Proteomics facility for performing and analyzing MS experiments. Lastly, we thank all members of the Clausen lab for discussions. **Funding:** This work was supported by the Austrian Research Promotion Agency Headquarter grant 852936 (T.C.), Marie Skłodowska-Curie grant 847548 (D.B.G.), Austrian Science Fund 10.55776/DOC.funds 112-B27 (J.F.E.) and Esprit grant 10.55776/ESP218B (P.M.), WWTF grants 10.47379/LS21029 (L.E.B.) and 10.47379/LS21009 (L.D.), and a BI fonds fellowship (L.E.B.). This work was also supported by grants RS-2020-NR049540, RS-2021-NR056577, and RS-2022-NR067411 from the National Research Foundation of Korea (H.K.S.). The Research Institute of Molecular Pathology is supported by Boehringer Ingelheim. **Author contributions:** D.B.G. and T.C. designed experiments. D.B.G., P.M., J.F.E., R.K., J.N., and L.D. performed cloning, expression, and purification. D.B.G., P.M., and A.S. performed biochemical experiments. D.B.G. generated and analyzed cryo-EM data. W.S.Y., J.L., and J.S.S. generated and analyzed crystallography data. A.S. performed the bioinformatic analysis. A.M., G.A.V., E.Z., H.K.S., R.S.H., and T.C. supervised research and provided scientific input. D.B.G. and T.C. coordinated the research project and prepared the manuscript with input from all authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Cryo-EM maps and atomic coordinates have been deposited in the Protein Data Bank (PDB) with accession codes pdb_00009QWS, pdb_00009QWX, pdb_00009QWZ, pdb_00009QWU, pdb_00009QX0, pdb_00009QX1, pdb_00009QX2, pdb_00009QX5, and pdb_00009QT9 and in the Electron Microscopy Data Bank (EMDB) with accession codes EMD-52491, EMD-53426, EMD-53430, EMD-53431, EMD-52494, EMD-53428, EMD-53432, EMD-52488, EMD-53425, EMD-52504, EMD-53433, EMD-53434, EMD-53435, EMD-52513, EMD-52516, and EMD-53348. Crystal structures have been deposited in the PDB with accession codes pdb_00009LGS, pdb_00009JNI, and pdb_00009UPZ. Uncropped gels are provided in fig. S21. The MS proteomics data have been deposited to the ProteomeXchange Consortium through the PRIDE partner repository with the dataset identifier PXD063485.

SUPPLEMENTARY MATERIAL

science.org/doi/10.1126/science.adv9309
Materials and Methods; Figs. S1 to S21; Tables S1 to S5; References (*49*–*55*)

Submitted 13 January 2025; accepted 24 June 2025

10.1126/science.adv9309

RIVER DYNAMICS

Vegetation changes the trajectory of river bends

Michael Hasson[1]*, Alvise Finotello[2], Alessandro Ielpi[3], Mathieu G. A. Lapôtre[1]

A primary axiom in geoscience is that the evolution of plants drove global changes in river dynamics. Notably, the apparent sinuosity of rivers, derived from the variance of sediment accretion direction measured in rocks, substantially increased when land plants evolved, around 425 million years ago. This led to the hypothesis that the rise of vegetation triggered river meandering. Recent studies of barren, meandering rivers challenge this notion, but the Paleozoic shift in the geometry of river deposits remains unexplained. Here, we suggest that it occurred because vegetation changes how river bends move through space. Using satellite images to monitor river migration, we found that bank vegetation alters the orientation of point bar accretion, resulting in a 62% increase in the inferred variance of flow direction. These results explain why meandering rivers have been underrecognized in prevegetation stratigraphy.

The long-held hypothesis that vegetation is required for rivers to develop a single-thread (one channel) meandering planform, in contrast to a multithread, braided planform (*1*–*4*), has been challenged in recent years (*5*–*14*). Although vegetation influences river dynamics in numerous ways—including slowing migration rates (*10*), creating muddier floodplains (*15*), reducing channel-floodplain connectivity (*12*), and strengthening the relationship between curvature and migration rates (*16*)—it is not, per se, a prerequisite for river meandering (*8*).

Despite increasing recognition of single-thread river deposits in the prevegetation sedimentary record, they appear to be rarer than would be expected based on the abundance of such rivers in modern, unvegetated basins [e.g., (*10*, *16*)]. This disparity begs the question of whether they are not recognized because the accepted criteria for their recognition need revision. Unvegetated meandering rivers produce different deposits than vegetated ones (*11*, *12*, *14*). Notably, their deposits show a greater degree of amalgamation owing to faster migration rates and contain downstream-accreting outer bank bars. Furthermore, they lack both levees and inclined heterolithic stratification (IHS), which forms as sediment accumulates in point bars on the inner banks of meander bends. The absence of IHS is notable because it has been used as a marker for meandering rivers in the rock record (*12*, *17*).

Paleocurrent variance—a measurement of the directional variability in river flow—is a criterion that has been commonly used to identify highly sinuous deposits in the rock record, often in concert with analysis of accretionary-bar growth (*18*–*31*). Free bars in braided rivers dominantly migrate downstream, such that the measured circular variance in accretion vectors (i.e., the angular variation) is relatively low (*18*, *29*, *32*–*39*). By contrast, sinuous, meandering rivers are expected to produce accretion measurements with greater variance, as point bars primarily accrete away from the channel belt axis rather than downstream (*18*–*21*, *29*, *40*–*44*).

Here, we show that vegetation measurably influences the migration trajectories of meanders: Point bars in unvegetated meandering rivers more frequently migrate downstream, parallel to river centerlines (translation), in contrast to vegetated bends that dominantly migrate perpendicular to centerlines (expansion). This difference in channel trajectory, linked to riparian vegetation, means that vegetated and unvegetated rivers produce accretion sets oriented perpendicular and parallel to the channel belt axis, respectively. As a result, the typical circular variance for bar accretion in vegetated rivers is 62% higher than that in unvegetated rivers such that, using established thresholds of variance, unvegetated meandering rivers would be identified as low-sinuosity multithreaded rivers (*29*).

These results indicate that the use of variance in bar accretion orientation is insufficient to identify most unvegetated meandering rivers in the rock record. Instead, a comparison of barform accretion and bedform migration is necessary to infer the record of the flow patterns distinctive of river meandering, because such flow patterns cause bedforms to migrate obliquely to the dominant, downstream flow direction [e.g., (*23*, *25*–*27*, *30*, *31*, *40*–*46*)]. Low-variance prevegetation meandering river deposits may have been misinterpreted as braided, and meandering rivers may therefore have been more widespread before the advent of land plants (and, more broadly, in arid or unvegetated environments) than previously recognized.

Migration angles

We measured the migration trajectories of 4482 individual meander bends from 49 unconfined meandering rivers of varying riparian vegetation densities (24 unvegetated, 7 semivegetated, and 18 vegetated) through time [Fig. 1, table S1, and materials and methods; (*47*)]. Rivers span a wide range of bank vegetation densities [quantified using the normalized difference vegetation index (NDVI); materials and methods], climates [quantified using the aridity index (AI); materials and methods], and flow variabilities [assessed using the monthly standard deviation in aridity index (SDAI) as a crude proxy; fig. S1 and materials and methods]. Although climate and vegetation density are linked, there is notable variation in vegetation density within climate zones (fig. S2).

Using a semiautomated workflow, we segmented bends as the reaches between consecutive inflection points (where curvature is zero; materials and methods). We measured the migration angle, θ, of each bend as the angle between a vector linking the centroids of the bend between two consecutive time steps and a tangent line to the closest point on the centerline of the channel belt (Fig. 1, B and D). The migration angle, θ, was always calculated relative to the downstream direction, such that positive values indicate bends migrating to the right of the channel belt trace and negative values indicate bends migrating to the left. We then calculated migration angle distributions for each analyzed river (fig. S3) and aggregated angles within vegetation classes for analysis (Fig. 2 and Table 1).

As bars migrate, they produce accretion surfaces with orientations that can be measured, either from subsurface stratigraphy or from their surface expressions on the alluvial plain (such as scroll bars; figs. S4 and S5). Tracking point bar trajectories in modern systems thus provides a direct proxy for bar accretion direction as preserved in alluvial stratigraphy. Absolute values of θ <45° (or >135°) represent the predominance of translation, whereas absolute values between 45° and 135° represent the predominance of expansion. Few bends are expected to be purely expansional or translational, such that distributions of θ represent the overarching relative predominance of expansion and translation across rivers.

Because the data are directional, we used circular variance (rather than linear) to compare distributions (*29*, *48*, *49*). The circular variance of aggregated θ values in unvegetated rivers (0.26) is less than that of vegetated (0.41) or semivegetated rivers (0.46), indicating the relative predominance of translation in unvegetated rivers (Fig. 2).

Vegetation and flow variability

Kolmogorov-Smirnov tests demonstrated that migration angles, θ, are statistically different between unvegetated and vegetated rivers, as well

[1]Department of Earth and Planetary Sciences, Stanford University, Stanford, CA, USA. [2]Department of Geosciences, University of Padova, Padova, Italy. [3]Department of Earth, Environmental and Geographic Sciences, University of British Columbia, Okanagan Campus, Kelowna, BC, Canada. *Corresponding author. Email: mhasson@stanford.edu

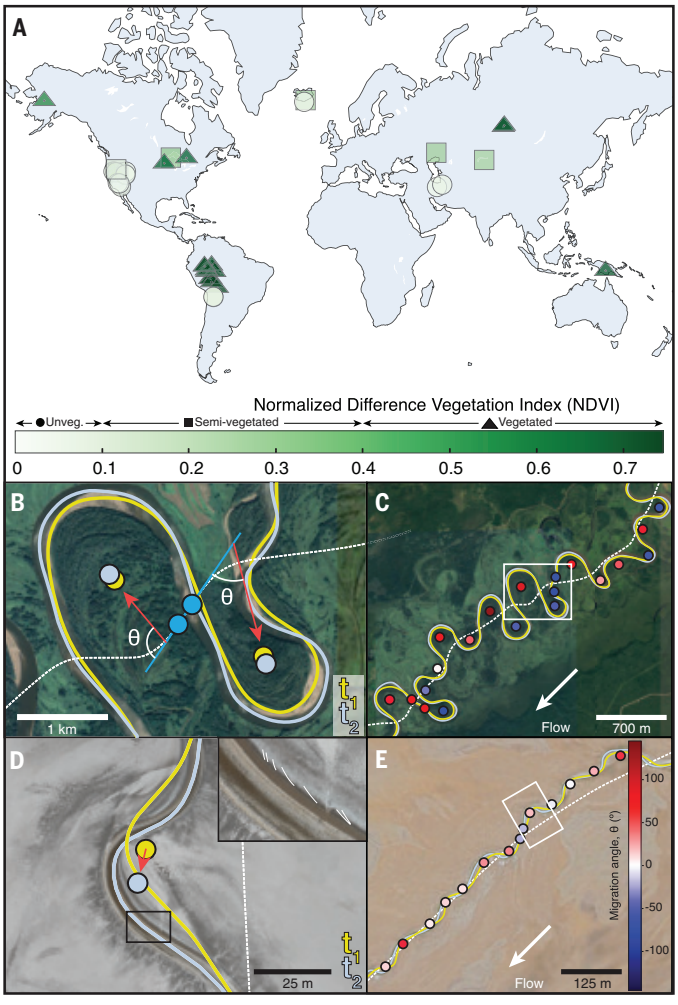


Fig. 1. River meanders in vegetated and unvegetated landscapes. (**A**) Map of analyzed rivers, colored by vegetation density (NDVI). (**B** and **C**) Example reach of meander bends that dominantly expand (Chulym River, 1984–2016). (**D** and **E**) Example reach of meander bends that dominantly translate (Dixie Valley Wash, 2000–2023). Yellow (time step 1; t_1) and light blue (time step 2; t_2) lines and filled yellow (t_1) and light blue (t_2) circles represent channel centerlines and bend centroids at each time step. In (B) and (D), white dashed lines indicate the approximate channel belt trace, and red arrows indicate point bar accretion direction. Blue circles show the closest point to the t_1 bend centroids on the channel belt trace, and the tangent lines to these points, shown in blue, were used to calculate migration angle, θ. In (C) and (E), centroids of meander bends are colored by θ; darker colors indicate higher migration angles (more expansion), and lighter colors indicate lower migration angles (more translation). White boxes in (C) and (E) indicate the regions shown in (B) and (D), respectively. The inset in (D) is a close-up of the point bar apron showing bedforms migrating up-bar that demonstrate the presence of helicoidal flow (crests outlined in white). [(B), (C), and (E) basemap credit: CNES/Airbus; (D) M.H. and M.G.A.L]

as between unvegetated and semivegetated rivers (see supplementary text). Migration angles are statistically similar between vegetated and semivegetated rivers (Table 2). Bends in rivers with any amount of vegetation tend to have higher, more expansional migration angles as opposed to unvegetated rivers, which have lower, more translational migration angles (Fig. 3). Piecewise linear regressions of median migration angle as a function of NDVI and the monthly SDAI reveal breakpoints in both relationships [at NDVI = 0.164 ± 0.058 and SDAI = 0.169 ± 0.053; standard error (SE)] based on the Davies test for the existence of at least one breakpoint [Fig. 3, B and C; (*50*)].

Bend morphology

Theoretical and practical insights into river morphodynamics suggest that translation should be more prominent in meandering rivers with poorly sinuous bends, which should therefore have lower migration angles than bends with higher sinuosity [(*51*–*53*); fig. S6]. It thus seems plausible that the observed differences in migration trajectory between vegetated and unvegetated rivers could arise from differences in bend sinuosity (Fig. 4). To test this hypothesis, we analyzed distributions of migration angles across quintiles of bend sinuosity. In vegetated rivers, migration trajectory is not significantly different between the lowest and highest sinuosity quintiles ($p = 0.36$). By contrast, higher sinuosity is associated with higher migration angles and expansion in unvegetated rivers. A comparison of trajectory distributions between the lowest and highest sinuosity quintiles reveals a significant difference for unvegetated bends ($p < 0.05$).

We further note that autogenic translation can occur where short bends with locally high curvatures are found (*54*). Although these short bends typically display relatively low sinuosity, there is no consistent relationship between curvature and sinuosity (fig. S7). This effect is due to curvature being a local parameter that describes how abruptly a channel changes direction, whereas sinuosity is a bend-related metric that describes the path length of a bend relative to a straight line between its end points. For calculations of curvature and sinuosity, see materials and methods.

Why do bends translate?

Bend translation can be caused by both allogenic and autogenic factors. The breakpoint in the relationship between migration angle and vegetation density indicates that vegetation roughly functions as a binary control on channel trajectory (Fig. 3). For vegetation densities below the breakpoint (NDVI <0.164), slight increases in NDVI drive marked increases in migration angle and bend expansion. Furthermore, principal components analysis (PCA) reveals that NDVI and sinuosity [which increases with vegetation density; (*16*)] are the most important variables (fig. S8). Although this result strongly supports our hypothesis, we investigated several other possible controls on migration behavior: flow variability, sediment flux, floodplain confinement and heterogeneity, and bend morphology.

Flow variability

Flow variability is known to affect migration patterns by shifting the locus of maximum erosion in meander bends downstream during major floods (*55*). It is thus somewhat surprising that greater SDAI values are associated with more expansional bend migration (Fig. 3C). Furthermore, as with NDVI (NDVI = 0.164 ± 0.035), a breakpoint exists in the relationship between migration angle and SDAI (0.169 ± 0.053, SE). However, whereas an NDVI value of about 0.1 corresponds to a real, physical threshold (i.e., the presence or absence of vegetation), an SDAI value of 0.169 does not align with any established regime change in discharge variability. Furthermore, migration angle is uncorrelated with the standard deviation of discharge (as measured from vegetated rivers with reliable data; fig. S9).

Disentangling the effects of discharge variability and vegetation is challenging because perennial flow facilitates the establishment of vegetative cover, and meandering rivers with scarce vegetation are rarely gauged. SDAI is an imperfect proxy for flow variability because it does not capture the punctuated flow events associated with, for example, monsoon precipitation in desert basins, given that it is calculated as the standard deviation of monthly averages. However, it is a reliably measurable quantity in small, ungauged, unvegetated rivers (unlike discharge). Nonetheless, the much clearer, step-like increase in migration angle with NDVI, and the physical implication of the NDVI value at which it occurs, supports the hypothesis that flow variability is not a primary driver of the observed trend but merely correlates with the primary driver—vegetation (fig. S9).

Sediment flux

Sediment flux has been linked to meander migration rates [e.g., (*56*)]. To understand whether it affects spatial changes in migration patterns,

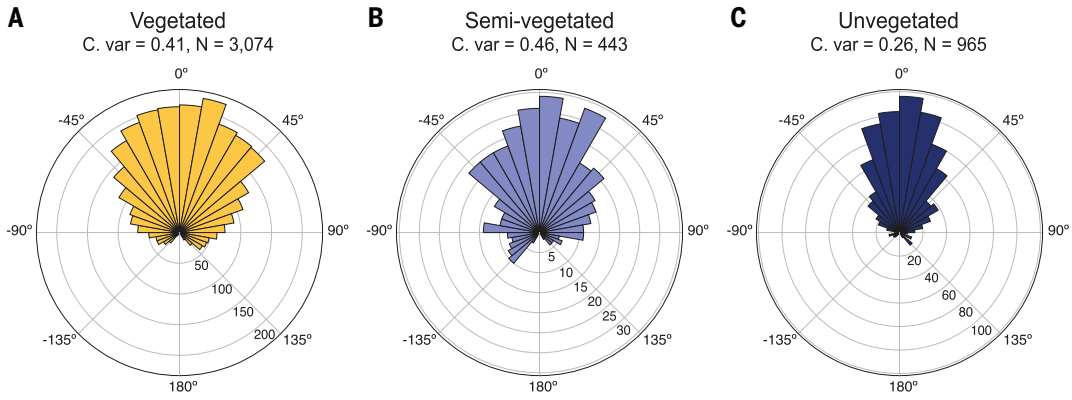


Fig. 2. Vegetation increases the probability of meander expansion. Rose diagrams of migration angle, θ, distributions for rivers of varying bank vegetation densities (quantified by NDVI): (**A**) vegetated, (**B**) semivegetated, and (**C**) unvegetated. We aggregated measurements from all bends of each vegetation class. Vegetated and semivegetated rivers are composed of more bends that expand away from the river centerline, whereas unvegetated rivers are composed of bends that mainly translate downstream. C. var., circular variance.

Table 1. Variance metrics.

	Vegetated	Semivegetated	Unvegetated
Linear variance (obsolete)*	3281.18	3377.10	1461.26
Linear standard deviation (obsolete)*	57.28°	58.11°	38.23°
Circular variance	0.41	0.46	0.26
Circular standard deviation	51.77°	54.75°	41.69°

*These are reported for the sake of comparison with classic literature, even though the use of these metrics has been superseded by circular variance.

Table 2. Kolmogorov-Smirnov test. The Kolmogorov-Smirnov test between migration angles in vegetated and semivegetated rivers indicates that the distributions are not significantly different, in contrast to the comparison of distributions between unvegetated and vegetated and unvegetated and semivegetated rivers. Circular mean, circular variance, and the number of samples are also reported. Circular mean is expected to be near zero because our measurement method defines zero as downstream, with negative and positive values reflecting migration to the left and right, respectively.

	Vegetated versus semivegetated	Unvegetated versus vegetated	Unvegetated versus semivegetated
Kolmogorov-Smirnov statistic	0.051	0.11	0.15
p value	0.26	$<10^{-3}$	$<10^{-3}$
	Vegetated	Semivegetated	Unvegetated
Circular mean	4.24°	−0.41°	5.84°
Circular variance	0.41	0.46	0.26
Circular standard deviation	51.77°	54.75°	41.69°
Number of samples	2994	443	965

we compared modeled values of bed and suspended sediment fluxes (*57*) with our calculations for migration angle. Only nonpermafrost vegetated rivers greater than 30 m wide were analyzed to ensure reliable modeled values (supplementary text). No relationship was observed between median migration angle and sediment flux, suggesting that the latter is not a primary driver of migration trajectory for these rivers (fig. S10).

Floodplain confinement and lithological heterogeneities

Migration dynamics not only depend on river curvature (*16*, *58*) but are also influenced by the erodibility of the substrate through which the river cuts. Erodibility, in turn, can vary owing to both allogenic factors [e.g., the presence of bedrock valley flanks; (*59*–*62*)] and autogenic processes [e.g., the formation of erosion-resistant plug bars left behind by cutoff channels; (*63*, *64*)]. Erosion-resistant bank materials are often related to floodplain heterogeneity, and specifically to the abundance of cohesive sediment [as a channel migrates against, e.g., abandoned channel fills or mud-rich units such as older crevasse splays; (*45*, *65*–*71*)]. Although theoretical frameworks have demonstrated that downstream migration is intrinsic to meander development [e.g., (*51*–*53*, *72*–*76*)], the translational component of migration is generally dwarfed by its expansional component in vegetated rivers, such that translation is often attributed to the aforementioned forcings when observed in the rock record.

Rivers in this study are not systematically confined by topography. Whereas low-erodibility regions can influence migration patterns, they are not a prerequisite for translation because autogenic translation can occur purely as a function of morphodynamics in homogeneous floodplains (*51*–*54*). Before the appearance of terrestrial vegetation, sand and mud were better mixed in floodplains and channel deposits, resulting in lower degrees of floodplain heterogeneity [(*10*, *12*); fig. S4]. Therefore, areas of low bank erodibility likely became more common after the advent of vegetation owing to changes in the distribution of mud in fluvial systems. If translation were primarily attributed to isolated floodplain regions of low erodibility in the investigated rivers, the greater floodplain heterogeneity of vegetated rivers would cause a higher proportion of vegetated river bends to translate downstream relative to unvegetated ones, contrary to observations (Fig. 2). Isolated regions of low erodibility due to floodplain heterogeneity are thus unlikely to be the primary driver of bend translation in unvegetated rivers.

Vegetation increases avulsion timescale and sinuosity

We hypothesize that the differences in migration angle distributions between vegetated and unvegetated bends arise from the effects of vegetation on meander development. Classic meander development theory states that, after initial development from a nearly straight channel, low-sinuosity bends primarily translate until the bend becomes sufficiently sinuous that expansion becomes predominant (*51*–*53*). In the terminal phase of development, expansion and translation slow until the bend undergoes cutoff and the cycle restarts.

Without bank vegetation, sediment is transported further from the channel and levees are subdued or absent (*12*, *14*, *77*). Avulsions occur when bed aggradation reaches around one channel depth (*78*), or when levee slopes are greater than the in-channel slope [termed gradient advantage; (*79*, *80*)].

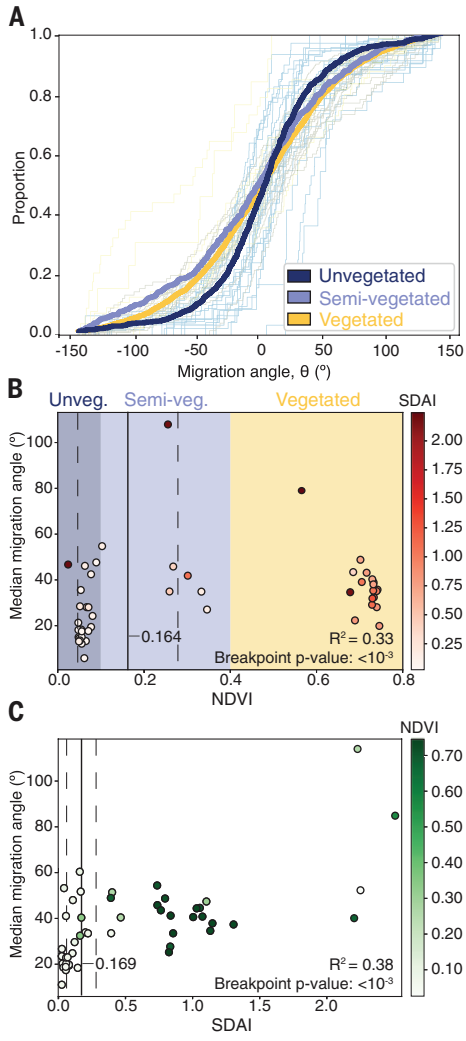


Fig. 3. Influence of vegetation and climate variability on migration trajectory. (**A**) Cumulative distribution functions (CDFs) of migration angle, θ. Thin lines show CDFs of each river in each class, and thick lines show aggregated data for each vegetation class. (**B**) Median migration angle as a function of NDVI, a measure of vegetation density. Points represent each studied river and are colored by SDAI to show variability within vegetation classes. Dark colors indicate seasonal climates, and light colors indicate consistent climates. For low-NDVI rivers, the median θ value increases sharply with slight increases in NDVI. The breakpoint (identified through piecewise linear regression) in the relationship at NDVI = 0.164 reflects the presence or absence of vegetation, which governs migration trajectory. (**C**) Median migration angle as a function of the standard SDAI. The breakpoint in the relationships suggests similar threshold behavior, but low-SDAI rivers tend to have low NDVIs, such that vegetation is still interpreted as the primary driver. In (B) and (C), solid vertical lines show the breakpoint identified through piecewise linear regression. Dashed vertical lines indicate the standard error ($N = 49$). Red lines indicate linear regressions for data on both sides of the breakpoint. R^2, coefficient of determination.

Subdued levees have very low slopes such that gradient advantage is unlikely to be a trigger. However, given the very low levee heights of unvegetated rivers, the bed aggradation required to trigger an avulsion (for a given width and sediment flux) is lower, such that we hypothesize that unvegetated rivers have shorter avulsion timescales than vegetated rivers of similar width.

Although modern unvegetated meandering rivers are typically mud rich, the vast majority of silt and clay-sized sediment is transported as sand-sized aggregates that are mixed into the same architectural elements with quartz and lithic sand grains. This implies transport equivalence: Large, low-density aggregates are transported alongside smaller, denser, individual grains [(*12*, *14*); supplementary text]. Suspended sediment flux affects levee development and migration rates, but the composition of the suspended sediment itself is unlikely to.

Sediment flux is challenging to constrain in ungauged unvegetated rivers, although we found no significant relationship between sediment flux and migration angle in larger (>30 m wide) nonpermafrost vegetated rivers (fig. S10 and supplementary text). Furthermore, unvegetated rivers are thought to have elevated sediment fluxes relative to vegetated rivers of similar width and migration rate (*56*), which would further increase avulsion frequency due to enhanced bed aggradation (*81*).

In addition to undergoing avulsions more frequently, chute cutoffs are more frequent in unvegetated rivers, further limiting sinuosity (*82*). Expansional, unvegetated river bends are not common, but observed examples of neck cutoff suggest that very high sinuosity can develop given sufficiently long avulsion timescales (fig. S11). These instances, and our results as a whole, imply that highly sinuous, expansional bends can occasionally form in unvegetated rivers, but the probability of them forming and being preserved in the rock record is substantially lower than for vegetated meandering rivers.

Because planform morphology and meander development are scale independent (*16*, *83*, *84*), the reduced avulsion timescales and sinuosity are interpreted to be due to bank vegetation density rather than river or catchment size. Modern unvegetated meandering rivers are typically less than 30 m wide (*10*); however, they do not represent an upper size bound [e.g., (*8*)]. Rather, larger rivers form in larger catchments that integrate hydrological processes over greater areas such that they are more likely to be perennial and therefore support vegetation. Any controls on avulsion and cutoff timescales, related to lack of bank vegetation, are expected to remain similar for larger rivers that may have formed before the evolution of vegetation.

The statistically significant shift in migration angle distributions, associated with sinuosity quintiles, implies that bends in unvegetated rivers align well with theoretical relationships between meander development and sinuosity (Fig. 4). By contrast, vegetated river bends do not, as the statistical similarity of the lowest and highest sinuosity quintile distributions implies less dependence of migration angle on local bend sinuosity. Rather, plants stabilize riverbanks, anchor meander inflection points, and promote the formation of highly sinuous, high-amplitude bends that frequently undergo neck cutoffs [(*16*); see supplementary text]. These cutoffs, in turn, facilitate the formation of a few sharp, translation-prone bends and alter river migration both locally and nonlocally (*10*, *54*, *85*–*87*), such that the morphodynamics of one bend depend on its neighbors. Hence, the tendency of vegetated rivers to experience more frequent cutoffs explains why their migration angles are not simply a function of local bend sinuosity, and thus why the relationship between migration trajectories and bend sinuosity deviates from theoretical expectations for vegetated rivers.

Implications for the sedimentary record

Accurately reconstructing sedimentary environments—and particularly fluvial ones—has profound implications for our understanding of how the entire Earth system has evolved. Specifically, characterizing the planform geometry of ancient rivers has implications for critical shifts in global biogeochemical cycling [e.g., (*88*–*90*)] and our fundamental understanding of how terrestrial life developed (*4*). Until recently, the paucity of recognized meandering river deposits led to the hypothesis that vegetation and meandering rivers coevolved and, by extension, that single-threaded rivers would have been exceedingly rare before the advent of vegetation (*1*–*3*, *90*–*94*). Whereas single-threaded rivers can and have formed in the absence of vegetation for at least 2 billion years (*2*, *25*, *26*, *28*, *30*, *31*, *46*, *77*, *95*–*100*), the data presented in this work suggest they likely have been underrecognized and encourage modification of the criteria to identify them in the rock record.

Multiple lines of evidence are typically necessary to reconstruct river planform geometry. Field observations of the sedimentology and

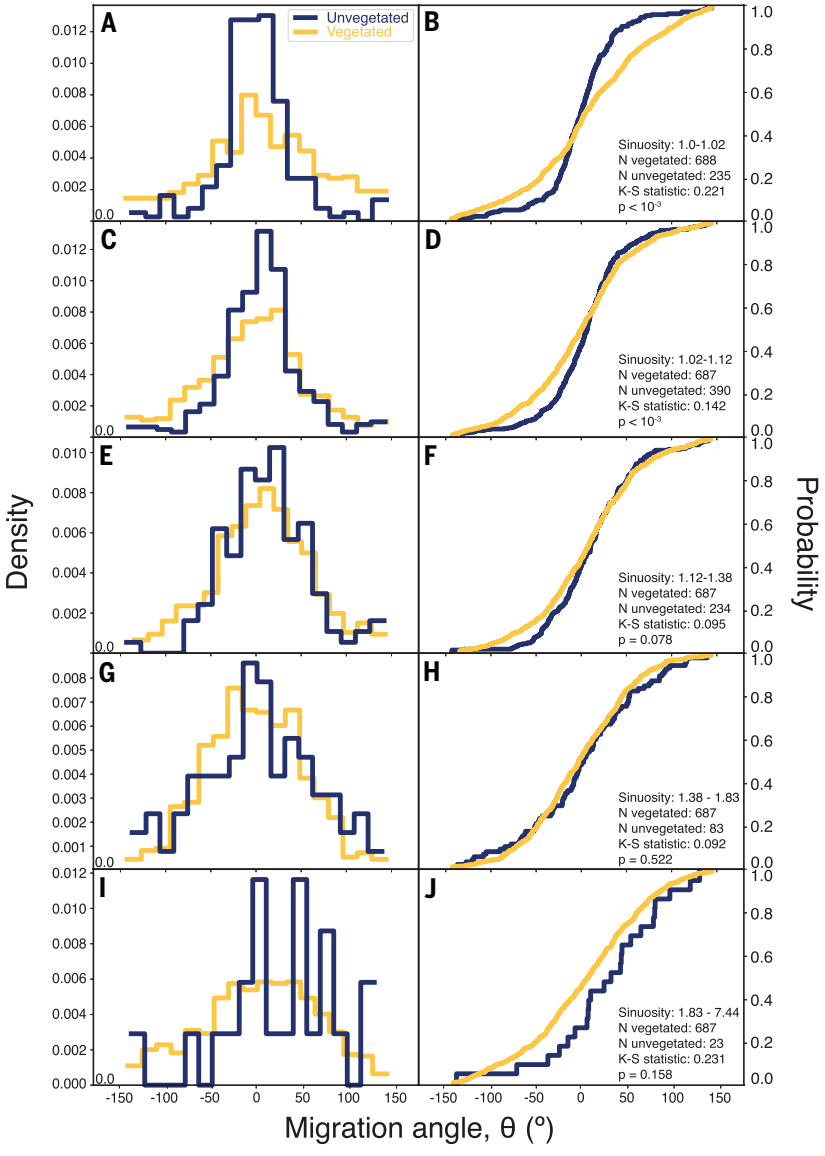


Fig. 4. Influence of bend sinuosity on migration trajectory. (**A** to **J**) Normalized histograms and cumulative distribution functions of migration angle, θ, subset by quintiles of sinuosity (as measured for all vegetated rivers). Sinuosity quintiles range from 1 to 1.02 (B), 1.02 to 1.12 (D), 1.12 to 1.38 (F), 1.38 to 1.83 (H), and 1.83 to 7.44 (J). We combined vegetated and semivegetated rivers because distributions of θ are not significantly different. For unvegetated meandering rivers, low sinuosity bends are more likely to translate than higher sinuosity bends. The θ distributions between the lowest and highest sinuosity quintiles are significantly different in unvegetated rivers ($p = 0.004$), whereas the distributions are not different in vegetated rivers ($p = 0.344$). The p values are associated with two-sample Kolmogorov-Smirnov (K-S) statistics.

stratigraphy of rocks—including the degree of amalgamation, grain size and variability, and general organization of deposits—often provide the foundation for analysis. However, few stratigraphic features have universal, definitive origins: for example, a previous study (*33*) demonstrated that lateral accretion (typically associated with meandering planforms) dominates the deposits of the braided Brahmaputra River. Conversely, sandy meandering rivers can produce deposits with extreme amalgamation, no preserved mud, and almost no evidence of lateral accretion, such that fluvial style is indeterminable without planform rock exposure (*24*). Furthermore, because the difference between barforms and bedforms is one of relative scale, multiple scales of architectural elements must be clearly observable for architectural analysis. Classifying elements is often challenging in poorly exposed prevegetation rocks that are already expected to display profoundly different stratigraphic character than postvegetation deposits.

Paleocurrent analysis is often used to supplement stratigraphic observations because it can elucidate past flow structure and, therefore, paleo-planform. Historically, paleocurrent analysis has focused on reconstructing sinuosity using the aggregated orientations of all measurable current and accretion structures in outcrop [e.g., (*18*–*21*); table S2]. More recently, the need to account for architectural hierarchy (i.e., the relative scales associated with bar accretion and bedform migration) has been recognized and used to understand river planform geometry (*23*, *25*–*28*, *30*, *31*, *46*, *101*). The value of this more rigorous methodology is well demonstrated by the reinterpretation of parts of the Proterozoic record, for example, the ~1-Ga fluvial deposits of the Torridonian succession of Scotland: Initial interpretations relied heavily on aggregated paleoflow measurements to infer deposition in braided rivers (*102*–*104*). More recent investigations, accounting for stratigraphic architecture (and thus the expectation of bedforms migrating oblique to barform accretion), have shown that numerous intervals contain among the most conclusive examples of prevegetation meandering rivers recorded to date (*26*, *31*, *99*). Accounting for depositional architecture when assessing paleoflow is therefore a robust approach to account for the processes associated with each scale of deposition that may substantially alter apparent paleocurrent variance.

As bar-scale features are generally thought to have the highest preservation potential (*39*, *67*), their orientations have been heavily studied in the floodplains of modern rivers (*29*, *32*, *33*, *35*–*38*, *42*–*44*). Whereas these analyses have shown the depositional variability within rivers of similar planform geometries, they universally highlight the focused, downstream migration of bars in braided and multithread rivers and the tendency of point bars in meandering rivers to migrate much more variably. Measurements presented in this work are solely from one channel generation, such that they are expected to be lower, for example, than those reported in a previous study (*29*) that documents floodplain records of multiple channel generations. However, the relative decrease in circular variance between vegetated and unvegetated rivers in this work is similar to that between single-threaded (meandering) and multithreaded (wandering and braided) rivers in that same work.

We have shown that many modern unvegetated meandering rivers are predominantly composed of downstream-migrating bars. Therefore, using aggregated current and accretion directions of their deposits to reconstruct channel sinuosity (as has been done extensively for ancient river systems) would likely lead to the misinterpretation that those rivers were braided rather than meandering. These results highlight the need for detailed analyses of depositional features observed in modern analogs (*11*, *12*, *14*). Additionally, paleocurrent analyses should only be relied upon when bedforms and barforms can clearly be differentiated to demonstrate the past existence of distinctive flow patterns (e.g., helicoidal flow cells in meander bends). If bedforms and barforms cannot be readily distinguished, unimodal, downstream migration of amalgamated architectural elements cannot be taken as evidence for a braided planform (Fig. 3). Together, the propensity of unvegetated meanders to translate downstream and use of aggregated paleocurrent and accretion measurements may have driven a fundamental misunderstanding of the prevegetation record, with cascading impacts on the accepted views of the coevolution of the hydrosphere, biosphere, and geosphere on Earth.

REFERENCES AND NOTES

1. S. A. Schumm, *Geol. Soc. Am. Bull.* **79**, 1573–1588 (1968).
2. D. G. Long, *Fluvial Sedimentology Memoir* **5**, 313–341 (1977).
3. N. S. Davies, M. R. Gibling, *Geology* **38**, 51–54 (2010).
4. M. R. Gibling, N. S. Davies, *Nat. Geosci.* **5**, 99–105 (2012).
5. M. E. Donselaar, M. C. Cuevas Gozalo, S. Moyano, *Sediment. Geol.* **283**, 1–14 (2013).
6. Y. Matsubara, A. D. Howard, *Earth Surf. Process. Landf.* **39**, 1365–1377 (2014).
7. A. Ielpi, R. H. Rainbird, D. Ventra, M. Ghinassi, *Nat. Commun.* **8**, 15250 (2017).
8. M. G. A. Lapôtre, A. Ielpi, M. P. Lamb, R. M. E. Williams, A. H. Knoll, *J. Geophys. Res. Earth Surf.* **124**, 2757–2777 (2019).
9. M. G. M. Santos *et al.*, *Sediment. Geol.* **385**, 1–14 (2019).
10. A. Ielpi, M. G. A. Lapôtre, *Nat. Geosci.* **13**, 82–86 (2020).
11. A. Ielpi, M. G. A. Lapôtre, A. Finotello, M. Ghinassi, A. D'Alpaos, *Earth Surf. Process. Landf.* **45**, 1717–1731 (2020).
12. M. Hasson, M. C. Marvin, A. Gunn, A. Ielpi, M. G. A. Lapôtre, *Sedimentology* **70**, 2272–2301 (2023).
13. J. S. Levy, B. Cvijanovich, *Geomorphology* **432**, 108705 (2023).
14. M. M. Douglas, K. L. Miller, M. P. Lamb, *Geol. Soc. Am. Bull.* **137**, 522–540 (2024).
15. S. S. Zeichner *et al.*, *Science* **371**, 526–529 (2021).
16. A. Finotello *et al.*, *Nat. Commun.* **15**, 1968 (2024).
17. A. Ielpi, M. G. A. Lapôtre, *J. Sediment. Res.* **89**, 399–415 (2019).
18. A. D. Miall, *J. Sediment. Res.* **44**, 1174–1185 (1974).
19. R. I. Ferguson, *Geol. Soc. Am. Bull.* **88**, 212–214 (1977).
20. J. P. Le Roux, *J. Sediment. Res.* **62**, 283–291 (1992).
21. P. Ghosh, *J. Sediment. Res.* **70**, 449–455 (2000).
22. D. G. F. Long, E. C. Turner, *Precambrian Res.* **214–215**, 192–209 (2012).
23. M. Ghinassi, P. Billi, Y. Libsekal, M. Papini, L. Rook, *J. Sediment. Res.* **83**, 1066–1084 (2013).
24. A. J. Hartley *et al.*, *Geology* **43**, 679–682 (2015).
25. A. Ielpi, R. H. Rainbird, *Sedimentology* **62**, 1950–1977 (2015).
26. M. G. M. Santos, G. Owen, *Precambrian Res.* **272**, 226–243 (2016).
27. L. E. Lebeau, A. Ielpi, *Sediment. Geol.* **357**, 53–71 (2017).
28. R. Barkat, P. P. Chakraborty, S. Saha, K. Das, *Precambrian Res.* **350**, 105930 (2020).
29. C. P. Galeazzi, R. P. Almeida, A. H. do Prado, *Geology* **49**, 1402–1407 (2021).
30. A. Marconato *et al.*, *J. S. Am. Earth Sci.* **114**, 103707 (2022).
31. W. J. McMahon, S. T. Herron, J. A. Craig, N. S. Davies, *Geol. Soc. Lond. Spec. Publ.* **540**, 75–101 (2023).
32. D. J. Cant, R. G. Walker, *Sedimentology* **25**, 625–648 (1978).
33. C. S. Bristow, in *Recent Developments in Fluvial Sedimentology*, vol. 39, F. G. Ethridge, R. M. Flores, M. D. Harvey, Eds. (SEPM Society for Sedimentary Geology, 1987).
34. A. D. Miall, *Sediment. Geol.* **55**, 233–262 (1988).
35. R. L. Skelly, C. S. Bristow, F. G. Ethridge, *Sediment. Geol.* **158**, 249–270 (2003).
36. I. A. Lunt, J. S. Bridge, *Sedimentology* **51**, 415–432 (2004).
37. G. H. Sambrook Smith, P. J. Ashworth, J. L. Best, J. Woodward, C. J. Simpson, *Sedimentology* **53**, 413–434 (2006).
38. G. H. Sambrook Smith *et al.*, *J. Sediment. Res.* **79**, 629–642 (2009).
39. N. N. Cyples, A. Ielpi, R. W. Dirszowsky, *J. Sediment. Res.* **90**, 131–149 (2020).
40. R. C. Selley, *J. Geol.* **76**, 99–110 (1968).
41. J. S. Bridge, *J. Sediment. Res.* **55**, 579–589 (1985).
42. M. Słowik, *J. Soils Sediments* **11**, 1262–1278 (2011).
43. D. Yue *et al.*, *Interpretation* **7**, T437–T454 (2019).
44. R. Dara, N. Kettridge, M. O. Rivett, S. Krause, D. Gomez-Ortiz, *J. Appl. Geophys.* **171**, 103863 (2019).
45. A. Ielpi, M. Ghinassi, *Sedimentology* **61**, 1923–1960 (2014).
46. A. Ielpi, R. H. Rainbird, *Sedimentology* **63**, 1550–1581 (2016).
47. M. Hasson, A. Finotello, A. Ielpi, M. Lapotre, Data for: Vegetation changes the trajectory of river bends, Version 1. Zenodo (2025); https://doi.org/10.5281/zenodo.15116956.
48. R. G. F. Krause, T. A. M. Geijer, *J. Sediment. Res.* **57**, 779–780 (1987).
49. P. Berens, *J. Stat. Softw.* **31**, 1–21 (2009).
50. R. B. Davies, *Biometrika* **74**, 33–43 (1987).
51. G. Seminara, G. Zolezzi, M. Tubino, D. Zardi, *J. Fluid Mech.* **438**, 213–230 (2001).
52. G. Zolezzi, G. Seminara, *J. Fluid Mech.* **438**, 183–211 (2001).
53. X. Guo, D. Chen, G. Parker, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 23448–23454 (2019).
54. Z. Sylvester, P. R. Durkin, S. M. Hubbard, D. Mohrig, *Geol. Soc. Am. Bull.* **133**, 2439–2456 (2021).
55. M. Ghinassi, J. Moody, D. Martin, *Geol. Soc. Am. Bull.* **131**, 71–83 (2018).
56. E. Greenberg, V. Ganti, *Earth Planet. Sci. Lett.* **634**, 118674 (2024).
57. S. Cohen, A. J. Kettner, J. P. M. Syvitski, B. M. Fekete, *Comput. Geosci.* **53**, 80–93 (2013).
58. Z. Sylvester, P. Durkin, J. A. Covault, *Geology* **47**, 263–266 (2019).
59. T. J. Nicoll, E. J. Hickin, *Geomorphology* **116**, 37–47 (2010).
60. D. G. Smith, S. M. Hubbard, J. R. Lavigne, D. A. Leckie, M. Fustic, in *From River to Rock Record: The Preservation of Fluvial Sediments and Their Subsequent Interpretation*, vol. 97, S. K. Davidson, S. Leleu, C. P. North, Eds. (SEPM Society for Sedimentary Geology, 2011).
61. H. Amini, S. Lanzoni, F. Monegaglia, M. Tubino, G. Zolezzi, *Geomorphology* **466**, 109427 (2024).
62. K. A. Fryirs, J. M. Wheaton, G. J. Brierley, *Earth Surf. Process. Landf.* **41**, 701–710 (2016).
63. İ. Güneralp, B. L. Rhoads, *Geophys. Res. Lett.* **38**, L14401 (2011).
64. M. Bogoni, thesis, University of Padova (2017).
65. W. J. Turnbull, E. L. Krinitzky, F. J. Weaver, *J. Soil Mech. Found. Div.* **92**, 121–136 (1966).
66. B. Makaske, H. J. T. Weerts, *Sedimentology* **52**, 651–668 (2005).
67. D. G. Smith, S. M. Hubbard, D. A. Leckie, M. Fustic, *Sedimentology* **56**, 1655–1669 (2009).
68. B. J. Willis, H. Tang, *J. Sediment. Res.* **80**, 440–454 (2010).
69. M. Ghinassi, A. Ielpi, M. Aldinucci, M. Fustic, *Sediment. Geol.* **334**, 66–96 (2016).
70. M. Ghinassi *et al.*, *Sedimentology* **65**, 1354–1377 (2018).
71. A. Ielpi, M. G. A. Lapôtre, M. R. Gibling, C. K. Boyce, *Nat. Rev. Earth Environ.* **3**, 165–178 (2022).
72. S. Ikeda, G. Parker, K. Sawai, *J. Fluid Mech.* **112**, 363–377 (1981).
73. G. Parker, K. Sawai, S. Ikeda, *J. Fluid Mech.* **115**, 303–314 (1982).
74. A. D. Howard, T. R. Knutson, *Water Resour. Res.* **20**, 1659–1667 (1984).
75. T. Sun, P. Meakin, T. Jøssang, K. Schwarz, *Water Resour. Res.* **32**, 2937–2954 (1996).
76. D. Chen, J. G. Duan, *J. Hydrol. (Amst.)* **321**, 59–76 (2006).
77. D. G. F. Long, in *From River to Rock Record: The Preservation of Fluvial Sediments and Their Subsequent Interpretation*, vol. 97, S. K. Davidson, S. Leleu, C. P. North, Eds. (SEPM Society for Sedimentary Geology, 2011).
78. D. Mohrig, P. L. Heller, C. Paola, W. J. Lyons, *Geol. Soc. Am. Bull.* **112**, 1787–1803 (2000).
79. R. Slingerland, N. D. Smith, *Geology* **26**, 435–438 (1998).
80. J. H. Gearon *et al.*, *Nature* **634**, 91–95 (2024).
81. A. Ielpi, M. G. A. Lapôtre, *J. Sediment. Res.* **92**, 695–703 (2022).
82. R. Maitan *et al.*, *Geology* **52**, 336–340 (2024).
83. J. Hooke, in *Proceedings of Sediment Dynamics and the Hydromorphology of Fluvial Systems Conference 2006*, J. S. Rowan, A. Werritty, R. W. Duck, Eds. (IAHS, 2006), pp. 127–135.
84. J. Hooke, *Geomorphology* **439**, 108825 (2023).
85. J. Schwenk, E. Foufoula-Georgiou, *Geophys. Res. Lett.* **43**, 12,437–12,445 (2016).
86. A. Ielpi, M. G. A. Lapôtre, A. Finotello, M. Ghinassi, *Earth Surf. Process. Landf.* **46**, 357–370 (2021).
87. D. J. Furbish, *Geol. Soc. Am. Bull.* **103**, 1576–1589 (1991).
88. S. J. Lyster, A. C. Whittaker, E. A. Hajek, *Geology* **50**, 822–826 (2022).
89. M. A. Torres *et al.*, *Earth Surf. Dyn.* **5**, 711–730 (2017).
90. E. Cotter, *Fluvial Sedimentology Memoir* **5**, 361–383 (1977).
91. P. G. Eriksson *et al.*, *Sediment. Geol.* **120**, 5–53 (1998).
92. P. G. Eriksson, A. J. Bumby, J. J. Brümer, M. van der Neut, *Sediment. Geol.* **190**, 25–46 (2006).
93. N. S. Davies, M. R. Gibling, M. C. Rygel, *Sedimentology* **58**, 220–258 (2011).
94. M. R. Gibling *et al.*, *Proc. Geol. Assoc.* **125**, 524–533 (2014).
95. S. M. Casshyap, *J. Sediment. Res.* **38**, 920–942 (1968).
96. D. A. Pretorius, *Gold in the Proterozoic Sediments of South Africa: Systems, Paradigms, and Models* (University of the Witwatersrand, 1974).
97. A. Button, N. Tyler, in *Economic Geology Seventy-Fifth Anniversary Volume*, B. J. Skinner, Ed. (Economic Geology Publishing Company, 1981), pp. 687–709.
98. A. Ielpi, M. Ghinassi, *Sediment. Geol.* **325**, 1–16 (2015).
99. V. Ganti, A. C. Whittaker, M. P. Lamb, W. W. Fischer, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 11652–11657 (2019).
100. P. P. Chakraborty, R. Bailwal, P. P. Paul, A. Sharma, *Front. Earth Sci.* **10**, 1 (2022).
101. A. Ielpi, *Geol. Mag.* **154**, 609–624 (2016).
102. G. E. Williams, *Nature* **209**, 1303–1306 (1966).
103. G. E. Williams, in *North Atlantic—Geology and Continental Drift*, vol. 12, M. Kay, Ed. (American Association of Petroleum Geologists, 1969).
104. P. G. Nicholson, in *Tectonic Controls and Signatures in Sedimentary Successions*, L. E. Frostick, R. J. Steel, Eds. (Wiley, 1994), pp. 183–202.
105. E. Greenberg, MeanderMigration, Version 1.0.1. Zenodo (2023); https://doi.org/10.5281/zenodo.8341861.

ACKNOWLEDGMENTS

Funding: There are no funding sources to be acknowledged. **Author contributions:** Conceptualization: M.H., M.G.A.L.; Methodology: M.H., M.G.A.L.; Investigation: M.H., M.G.A.L.; Formal analysis: M.H., M.G.A.L., A.I., A.F.; Visualization: M.H.; Resources: A.I., A.F., M.H., M.G.A.L.; Supervision: M.L.; Writing – original draft: M.H., M.G.A.L.; Writing – review & editing: M.H., M.G.A.L., A.I., A.F. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The datasets generated during and/or analyzed during this study are available in Zenodo (*47*). Code used to analyze sediment flux and water discharge data was published by (*56*) (version 1.0.1) and is available in Zenodo (*105*). Minor code blocks for processing river centerlines were adapted from (*58*), the original version of which is available at: https://github.com/zsylvester/curvaturepy. Adapted code published in the Zenodo repository for this work is specifically attributed to (*58*) where appropriate.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adv4939
Materials and Methods; Supplementary Text; Figs. S1 to S10; Tables S1 and S2; References (*106–135*); Data S1 and S2

Submitted 20 December 2024; accepted 23 June 2025

10.1126/science.adv4939

NANOMATERIALS

Deciphering icosahedra structural evolution with atomically precise silver nanoclusters

Feng Hu[1], Gaoyuan Yang[2], Lu-Ming Zheng[1], Gui-Jie Liang[2], Quan-Ming Wang[1]*

Determining the atomic structure of nanoparticles (NPs) is critical for understanding their structural evolution and properties. However, controlling the growth of multiply-twinned metal NPs remains challenging because of numerous competing pathways. In this work, we report the synthesis of two giant silver icosahedral nanoclusters, $[Ag_{213}(C{\equiv}CR1)_{96}]^{5-}$ and $[Ag_{429}Cl_{24}(C{\equiv}CR2)_{150}]^{5-}$ (Ag_{213} and Ag_{429}, R1 =3,4,5-$F_3C_6H_2$ and R2 = 4-$CF_3C_6H_4$), achieved through ligand engineering and kinetic control. Single-crystal x-ray diffraction reveals that Ag_{213} and Ag_{429} have multilayered icosahedral Ag_{141} | (Ag_{13}@Ag_{42}@Ag_{86}) and Ag_{297} (Ag_{13}@Ag_{42}@Ag_{92}@Ag_{150}) cores, respectively. Notably, Ag_{429} with 260 valence electrons is the largest Ag^0-containing nanocluster reported to date. These two giant silver nanoclusters are metallic in nature, as confirmed by their plasmonic absorption and pump-power–dependent excited-state dynamics. Their atomically precise structures support the layer-by-layer evolution from nuclei to seeds of silver icosahedra.

The shape control of metal nanoparticles (NPs) is very important owing to their morphology-dependent properties (*1*–*4*). Understanding the evolution pathways of NPs is essential for controlling shapes and tailoring properties (*5*–*8*). NP synthesis typically involves three stages: the formation of nuclei, seeds, and nanocrystals. The mechanism of nucleation has been extensively studied (*9*–*11*), and the evolution from seeds to nanocrystals has been revealed by electron microscopy (*5*, *12*). However, it remains challenging to visualize the growth from nuclei to seeds owing to the small size of the evolving species. We propose that determining the structure of giant metal nanoclusters (2 to 4 nm) may offer invaluable insights into the growth process of NPs. In this context, we synthesized large silver nanoclusters (2.1 and 2.6 nm) with an icosahedral morphology, and their well-defined atomic structures serve as bridges between nuclei and seeds.

Icosahedral morphology is commonly found in many noble metal NPs. An icosahedron is a regular polyhedron composed of 20 coplanar tetrahedra with 12 vertices, 30 edges, and 20 faces (*13*, *14*). Its {111} faces, corners, and edges are responsible for many interesting electronic, optical, and catalytic effects that have been reported. Silver icosahedral species are energetically favored at small size (<2 nm, ~300 atoms) (*15*). Controlling the synthetic parameters, including surface ligands and reduction conditions, is essential to producing large-sized silver icosahedra (*3*, *16*, *17*).

Because slow reduction is generally beneficial to the formation of icosahedra (*15*), we used weak reductant diphenylsilane (Ph_2SiH_2) to reduce alkynyl-silver precursors in an attempt to synthesize large icosahedral silver NPs. We managed to grow two x-ray–quality single crystals of 20-fold twinned silver nanoclusters, $[Ph_4P]_5[Ag_{213}(C{\equiv}CR1)_{96}]$ (Ag_{213}) and $[Ph_4P]_5[Ag_{429}Cl_{24}(C{\equiv}CR2)_{150}]$ (Ag_{429}) (R1 = 3,4,5-$F_3C_6H_2$; R2 = 4-$CF_3C_6H_4$) (Fig. 1), and resolved their atomic structures. Ag_{213} and Ag_{429} have three- and four-layered icosahedral core structures, respectively, which strongly suggests a layer-by-layer evolution from icosahedral nuclei to seeds. Additionally, Ag_{213} and Ag_{429} are metallic in nature, as confirmed by their plasmonic absorption and pump-power–dependent excited-state dynamics.

Synthesis

Instead of using $NaBH_4$ (*16*), a strong reducing agent extensively used in metal nanocluster synthesis, we used Ph_2SiH_2 as a weak reductant to effectively slow down the reduction rate and limit the generation of intractable precipitates (*18*, *19*). Organic ligands on the surface of nanoclusters play an important role in regulating the size and arrangement of the kernel. Ligands with small steric hindrance may adopt a compact arrangement in the ligand shell (*20*), which favors the formation of a large kernel. Previously, 3,5-ditrifluoromethyl phenyl acetylene has been used during the synthesis of $[Ag_{112}Cl_6(C{\equiv}CR)_{51}]^{3-}$ (Ag_{112}) (*18*). In this work, we chose phenyl acetylenes (3,4,5-trifluoromethyl phenyl acetylene, 3,4,5-$F_3C_6H_2C{\equiv}CH$ and *p*-trifluoromethyl phenyl acetylene, 4-$CF_3C_6H_4C{\equiv}CH$), which have lower steric hindrance, and successfully prepared two well-defined silver icosahedral nanoclusters, Ag_{213} and Ag_{429}. The introduction of fluoro groups in these alkynyl ligands is based on a ligand engineering strategy that intercluster F···F interactions may provide extra stabilization force favoring the efficient packing of the clusters, thereby enabling the formation of stable crystals (*21*). The chlorides in Ag_{429} come from Ph_4PCl rather than dichloromethane because Ag_{429} could not be obtained in the absence of Ph_4PCl. When Ph_4PCl was replaced with Ph_4PBr in the synthesis, black crystals with an ultraviolet-visible (UV-vis) absorption spectrum nearly identical to that of Ag_{429} were obtained, which suggests that Ph_4PBr can also be used for the synthesis of Ag_{429} containing bromides (fig. S1).

Structural analysis of Ag_{213} and Ag_{429}

Single-crystal x-ray diffraction revealed that Ag_{213} is composed of 213 silver atoms and 96 alkynyl ligands, whereas Ag_{429} consists of 429 silver atoms, 24 Cl ligands, and 150 alkynyl ligands. The metal kernel sizes of Ag_{213} and Ag_{429} are 2.1 and 2.6 nm, respectively (figs. S2 and S3). The molecular formulae of Ag_{213} and Ag_{429} are $[Ph_4P]_5[Ag_{213}(C{\equiv}CR1)_{96}]$ and $[Ph_4P]_5[Ag_{429}Cl_{24}(C{\equiv}CR2)_{150}]$, respectively. Although the cations in Ag_{213} and Ag_{429} were not resolved, the existence of Ph_4P^+ was confirmed by electrospray ionization time-of-flight mass spectrometry in the positive ion mode (fig. S4). In the negative ion mode, Ag_{213} shows two main signals at mass/charge ratio (m/z) = 7583 and 9532, corresponding to $[Ag_{213}(C{\equiv}CR_1)_{96}]^{5-}$ (7573) and $[Ag_{213}(C{\equiv}CR1)_{96}]^{4-}$ (9466), respectively (fig. S5A). Similarly, Ag_{429} exhibits two peaks at m/z = 12195 and 14634, corresponding to $[Ag_{429}Cl_{24}(C{\equiv}CR2)_{150}]^{6-}$ (12082) and $[Ag_{429}Cl_{24}(C{\equiv}CR2)_{150}]^{5-}$ (14498), respectively (fig. S5B). The mass difference between measured and simulated is a result of their very large molecular weight and the dynamic dissociation and/or association of ligands and solvent molecules.

As shown in Fig. 2, the four-layered core-shell structure of Ag_{213} is composed of the icosahedral core Ag_{141} (Ag_{13}@Ag_{42}@Ag_{86}) and the outermost shell $Ag_{72}(C{\equiv}CR1)_{96}$. The first layer Ag_{12} encloses a center silver atom to form a centered icosahedral Ag_{13} with I_h symmetry (Fig. 2A), and the second layer is Ag_{42} with a Mackay icosahedron structure (Fig. 2B). The two layers form a complete Ag_{55} icosahedral structure. The third layer Ag_{86} (Fig. 2C and fig. S6) is derived by symmetrically truncating the six vertices on the waist of the complete icosahedron Ag_{92}. The absence of these six silver atoms causes the Ag_{141} core to have C_{2h} rather than I_h symmetry (Fig. 2D and fig. S7). Surrounding the Ag_{141} core is an Ag_{72} shell built up from two triple-spiral fan-shaped Ag_{18} and 12 Ag_3 triangles (Fig. 2, E and F). The average Ag-Ag bond lengths of the first, second, third, and fourth layers are 2.887 Å, 2.929 Å, 2.955 Å, and 3.014 Å, respectively (table S1).

[1]Department of Chemistry, Engineering Research Center of Advanced Rare Earth Materials (Ministry of Education), Tsinghua University, Beijing, P. R. China. [2]Hubei Key Laboratory of Low Dimensional Optoelectronic Materials and Devices, Hubei University of Arts and Science, Xiangyang, P. R. China. *Corresponding author. Email: qmwang@tsinghua.edu.cn

Ag_{429} has a five-layered concentric structure. It is composed of an icosahedral core Ag_{297} (Ag_{13}@Ag_{42}@Ag_{92}@Ag_{150}) and a peripheral protective shell $Ag_{132}(C{\equiv}CR2)_{150}$. The first three layers of silver atoms are the complete Mackay icosahedra Ag_{12}, Ag_{42}, and Ag_{92} (Fig. 2, G to I), which form a centered icosahedron Ag_{147} with 20 triangle {111} faces. The fourth layer of Ag_{150} is derived from a larger icosahedral Ag_{162} by symmetrically removing 12 vertices (Fig. 2J). Therefore, the Ag_{297} core consisting of four layers still has I_h symmetry. The projection of Ag_{297} along the [100], [111], and [023] axes shows the C_2, C_3, and C_5 symmetry (Fig. 2, K to M). Surrounding the Ag_{297} core is an Ag_{132} shell consisting of six dumbbell-shaped Ag_{14} and eight triangular Ag_6 (Fig. 2, N and O). The average bond lengths of the first, second, third, fourth, and fifth layers of silver are 2.864 Å, 2.900 Å, 2.932 Å, 2.940 Å, and 2.997 Å, respectively (table S2).

An icosahedral nanocrystal is composed of 20 vertex- and face-sharing fcc tetrahedra. In the Ag_{141} core of Ag_{213}, there are three kinds of tetrahedral units (Ag_{18}, Ag_{19}, and Ag_{20}; see Fig. 3, A to C). Ag_{20} is a tetrahedral unit with a complete four-layer fcc arrangement (Ag:Ag_3:Ag_6:Ag_{10}). Ag_{19} and Ag_{18} units are derived from Ag_{20} by removing one and two vertex silver atoms, respectively. The Ag_{141} core consists of 12 Ag_{18} units, 6 Ag_{19} units, and 2 Ag_{20} units. The Ag_{297} core in Ag_{429} has 20 identical incomplete Ag_{32} tetrahedra (Fig. 3D), which is derived from removing three vertex silver atoms from a complete Ag_{35} tetrahedron. The Ag_{32} unit has five close-packed layers of silver atoms stacking in the pattern of Ag:Ag_3:Ag_6:Ag_{10}:Ag_{12}. The icosahedral configuration, characterized by {111} facets and near-spherical geometry, is well known to minimize surface free energy. However, its stability will be compromised with increasing size owing to rapidly accumulating strain energy. In Ag_{213} and Ag_{429}, the apex truncation of tetrahedral units and formation vertex-deficient icosahedral shells are identified as an effective way to release strain (*22*–*24*). For an icosahedron, the number of atoms $S_{(n)}$ in each layer can be calculated by $S_{(n)} = 10n^2 + 2$ (where S is the number of surface atoms in respective layers and n is the number of concentric layers) (*13*). Unlike Ag_{112} with a small two-layered icosahedron core (Ag_{55}, 1+S_1+S_2) (*18*), the cores of Ag_{213} and Ag_{429} are three-layered icosahedron Ag_{141} (1+S_1+S_2+S_3-6; Fig. 3E) and four-layered icosahedron Ag_{297} (1+S_1+S_2+S_3+S_4-12; Fig. 3F).

Alkynyl ligands exhibit diverse coordination patterns on the surface of Ag_{213} and Ag_{429}. They can be divided into two categories: (i) bridging on the surface of the nanocluster (fig. S8, A to D) and (ii) connecting the sublayer and the outermost layer (Fig. 4, A and B, and fig. S8, E and F). There are five kinds of alkynyl ligands in Ag_{213}. Ten alkynyl ligands (fig. S8A) and 20 alkynyl ligands (fig. S8B) are directly attached to the fan-shaped Ag_{18} and triangular Ag_3 surfaces (fig. S9, A and B). The centers of six pentagonal Ag_5 on Ag_{86} are occupied by the alkynyl ligands in the μ_6-η^1, η^1, η^1, η^1, η^1, η^1 mode (fig. S9, C and D). The other alkynyls filled the cracks on the surface of Ag_{213} kernel (fig. S9, E to H). Each pentagonal vertex of Ag_{86} is occupied by one μ_3-η^1, η^1, η^1 coordinated alkynyl ligand. Thirty alkynyl ligands link the Ag_{18} and Ag_3 units together. In Ag_{429}, 36 and 24 alkynyl ligands (fig. S10, A and B) are distributed on the surface of hexagonal Ag_7 and triangular Ag_6, respectively. There are six μ_5-η^1, η^1, η^1, η^1, η^1 coordinated alkynyl ligands (fig. S10, C and D) in the quadrilateral Ag_4. The vertices of the pentagon on Ag_{150} are connected to five μ_3-η^1, η^1, η^1 coordinated alkynyl ligands (fig. S10, E and F). Twenty-four alkynyl ligands (fig. S10, G and H) act as bridges between Ag_6 and Ag_7 units. Each chloride takes μ_2-η^1, η^1 mode as a bridge on an edge of pentagonal Ag_5 in Ag_{429} (Fig. 4C and fig. S8G). As shown in fig. S11, it is noteworthy that the surface ligands of Ag_{213} display intercluster F···F interactions with distances ranging from 2.778 to 2.969 Å. These F···F interactions are also observed in Ag_{429}, with distances ranging from 2.723 to 2.983 Å. These F···F distances are comparable to those of conventional F···F interactions (*21*).

Ligand-protected giant silver nanoclusters are rare because of the difficulties in synthesis and structure determination. Ag_{429} is the only well-defined NP with more than 400 silver atoms. Thiolate-protected giant silver nanoclusters (~2.5 nm) have been reported (*25*), exhibiting decahedral morphology based on a pentagonal bipyramid. By sharp contrast, the alkynyl-protected Ag_{213} and Ag_{429} synthesized in this work are icosahedral. This difference in shape is likely caused by the different binding modes of alkynyl and thiolate ligands: Whereas thiolate

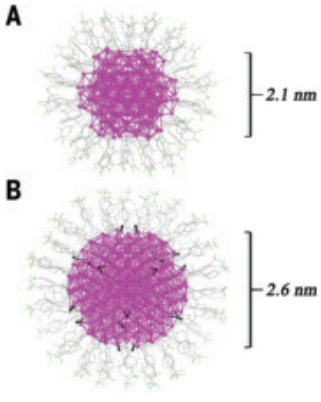


Fig. 1. Silver icosahedra. (**A** and **B**) Molecular structures of $[Ag_{213}(C{\equiv}CR1)_{96}]^{5-}$ (A) and $[Ag_{429}Cl_{24}(C{\equiv}CR2)_{150}]^{5-}$ (B) (R1 = 3,4,5-$F_3C_6H_2$; R2 = 4-$CF_3C_6H_4$) (Ag, pink; Cl, black; F, green; C, gray) determined by single-crystal x-ray diffraction. All hydrogen atoms are omitted for clarity.

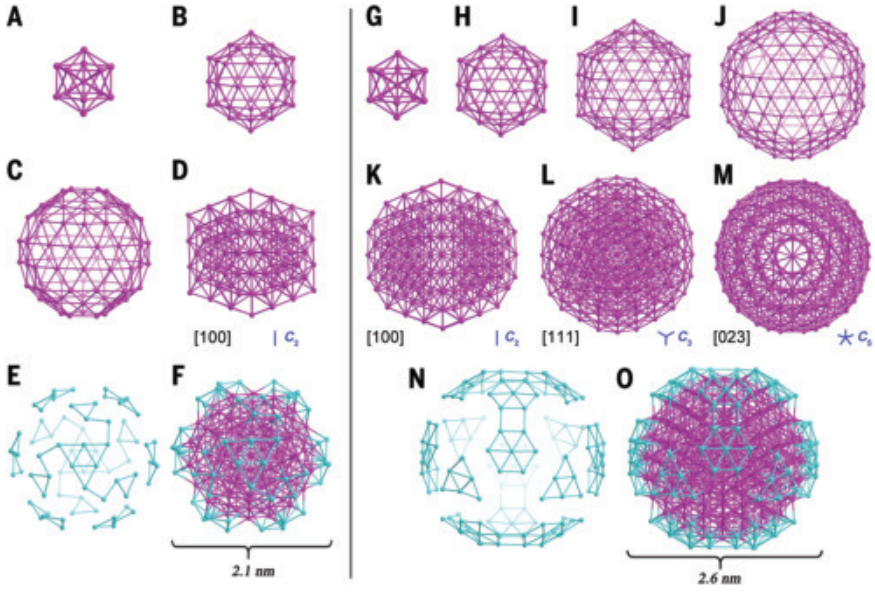


Fig. 2. Structural anatomy of Ag_{213} and Ag_{429}. (**A** to **F**) The four-layered core-shell structure of Ag_{213}. (A) Inner Ag_{13} centered icosahedron. (B) Second layer of Ag_{42} icosahedron. (C) Third layer of Ag_{86} icosahedron. (D) Icosahedral Ag_{141} core viewed along C_2 axis. (E) Outermost shell of Ag_{72}. (F) Overall Ag_{213} kernel structure with the outermost shell of Ag_{72} highlighted in blue. (**G** to **O**) The five-layered core-shell structure of Ag_{429}. (G) Inner Ag_{13} centered icosahedron. (H) Second layer of Ag_{42} icosahedron. (I) Third layer of Ag_{92} icosahedron. (J) Fourth layer of Ag_{150} icosahedron. [(K) to (M)] Icosahedron Ag_{297} viewed along C_2, C_3, and C_5 axes, respectively. (N) Outermost shell of Ag_{132}. (O) Overall Ag_{429} kernel structure with the outermost shell of Ag_{132} highlighted in blue.

ligands primarily coordinate with surface silver atoms (*26*, *27*), alkynyl ligands can both coordinate to these atoms and bridge the outermost layer with the underlying sublayer. Alkynyl ligands may adopt both σ and π bonding to coordinate silver atoms (*28*, *29*) (fig. S8, E and F). As shown in fig. S12, because of the binding of alkynyl ligands to the sublayer silver atoms, the Ag-Ag bond lengths between the sublayer and the outermost layer are substantially longer (2.942 and 2.937 Å for Ag_{213} and Ag_{429}, respectively) compared with those in between inner layers (2.834 and 2.823 Å). In addition, the formation of icosahedral nanocrystals tends to be more likely when the reduction rate is kept sufficiently slow because low reduction rate favors kinetically controlled nucleation (*15*). The weak reducing agent Ph_2SiH_2 instead of the strong $NaBH_4$ reduces the silver precursor needed to obtain Ag_{213} and Ag_{429}.

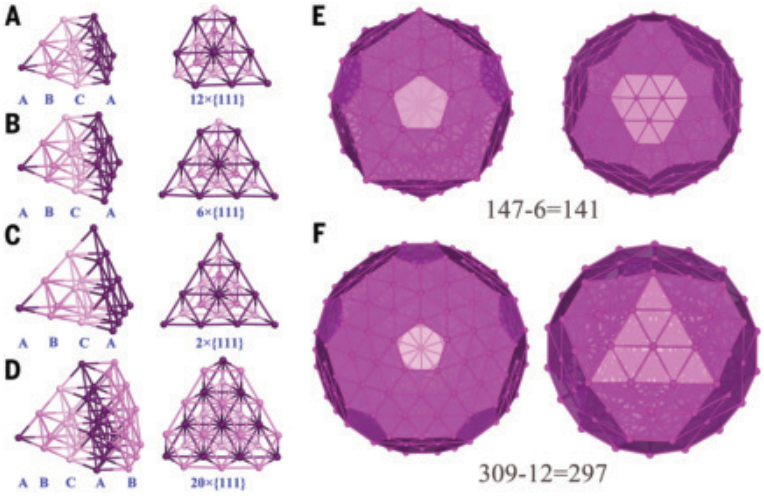


Fig. 3. The tetrahedral units and representative facets. (**A** to **D**) Various tetrahedral units. (**E**) Top and side view of Ag_{141} core in polyhedral drawing. (**F**) Top and side view of Ag_{297} core in polyhedral drawing.

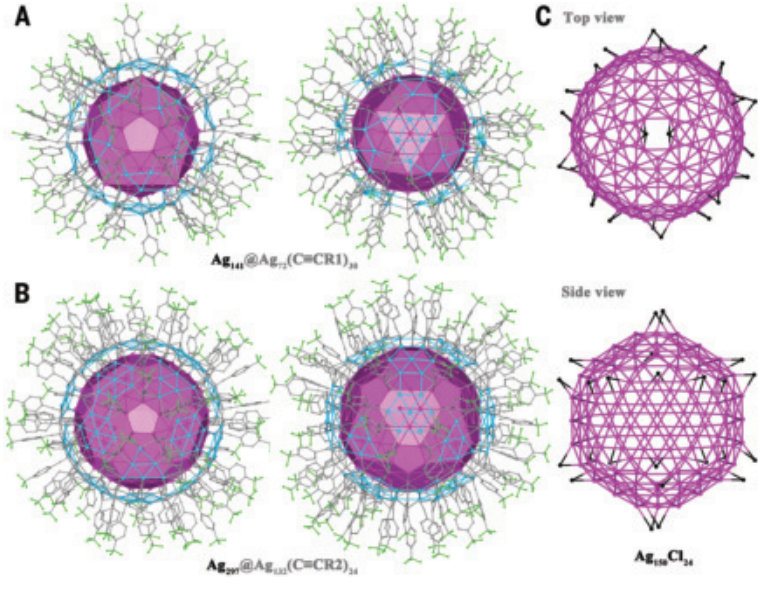


Fig. 4. Peripheral metal-ligand shell. (**A**) Binding motifs on the surface of Ag_{213} (Ag, pink and blue; Cl, black; F, green; C, gray). (**B**) Binding motifs on the surface of Ag_{429}. (**C**) The binding mode of Cl on the Ag_{150} layer of Ag_{429}.

Steady-state absorption and transient absorption

UV-vis absorption spectra of Ag_{213} and Ag_{429} in dichloromethane at room temperature (Fig. 5, A and B) reveal single, prominent absorption peaks at 478 and 475 nm, respectively, attributed to the localized surface plasmon resonance of large-sized silver nanoclusters. The absence of multiple excitonic absorption peaks, typical of molecule species, supports their metallic nature. The stability of Ag_{213} and Ag_{429} was monitored with absorption spectroscopy (fig. S13). In the absence of light, Ag_{213} and Ag_{429} in dichloromethane could be stable for 1 and 3 days, respectively. Under ambient light, their stability durations were reduced to 4 hours and 12 hours, respectively. Their stability also decreased at elevated temperature—both Ag_{429} and Ag_{213} can stay intact for only 1 hour at 40°C.

Metal nanoclusters of 1- to 3-nm size are ideal candidates for studying the boundary between molecular and metallic states (*30*, *31*). Time-resolved transient absorption (TA) spectroscopy has recently been applied to study the electronic dynamics of gold nanoclusters (*19*); however, similar studies on silver nanoclusters have not been reported. We measured the TA spectra of Ag_{213} and Ag_{429}. The UV-vis absorption spectra before and after the TA experiments are identical, which indicates that no decomposition occurred (fig. S14). The linear relationship between the maximum intensity of TA signals and pump fluence rules out the possibility of multiphoton effects and optical damage to the sample (fig. S15). Under 400-nm excitation, both TA data maps of Ag_{213} and Ag_{429} are similar to Ag NPs (*32*), which shows a strong ground-state bleaching (GSB) at 485 (Ag_{213}) or 483 nm (Ag_{429}) and a wide excited-state absorption band at 520 to 700 nm (Fig. 5, C and D).

The GSB at 485 and 483 nm corresponds to the absorption at 478 and 475 nm in steady-state absorption of Ag_{213} and Ag_{429}, respectively. The TA data maps show that the TA signals of Ag_{213} and Ag_{429} mainly attenuated at 5 ps (Fig. 5, E and F). Decay kinetics under different pump fluences were tested, and the spectral features of Ag_{213} and Ag_{429} are not dependent on pump fluences (figs. S16 and S17). The kinetic traces of Ag_{213} at selected wavelengths (485 nm) indicate that it takes 50 ps for the TA signal to completely decay to zero (Fig. 5G). However, it takes 100 ps for the TA signal of Ag_{429} at 483 nm to decay to zero (Fig. 5H). The GSB signals for Ag_{429} show an unusual increase at 20-ps time delay with substantially stronger pump fluences (fig. S18), similar to what was observed for Ag NPs by Hodak *et al.* (*32*).

The GSB at 485 nm was selected to study the power-dependent dynamics of Ag_{213}. The decay process of Ag_{213} shows pump-power–dependent excited-state dynamics (fig. S19A). Through global fitting analysis of Ag_{213} pumped with an excitation pulse energy of 64 nJ per pulse, the lifetime of electron-phonon coupling (τ_{e-ph}) is found to be 0.97 ps (fig. S20). Similarly, Ag_{429} shows a positive dependence on the pump fluence (fig. S19B). However, in addition to τ_{e-ph}, an additional long-time component corresponding to phonon-phonon coupling (τ_{ph-ph}) appears in the TA maps. Global fitting analysis of Ag_{429} pumped with an excitation pulse energy of 44 nJ per pulse determines τ_{e-ph} and τ_{ph-ph} to be 0.69 and 3.79 ps, respectively (fig.S21). The intrinsic τ_{e-ph} (τ_0) values for Ag_{213} and Ag_{429} are determined to be 0.64 and 0.42 ps, respectively, by extrapolating the linear trend in relaxation rates to zero laser pulse energy (Fig. 5, I and J, and tables S3 and S4). Using the same method, the intrinsic τ_{ph-ph} of Ag_{429} is determined to be 3.4 ps (fig. S22) The room-temperature electron-phonon coupling constant G is typically used to represent the rate of energy flow from the hot electron gas to the lattice ($G = \gamma^* T_0/\tau_0$, where $\gamma = 66\ J\ m^{-3}\ K^{-2}$ and T_0 is the laboratory temperature). G values for Ag_{213} and Ag_{429} are calculated to be 3.07×10^{16} and

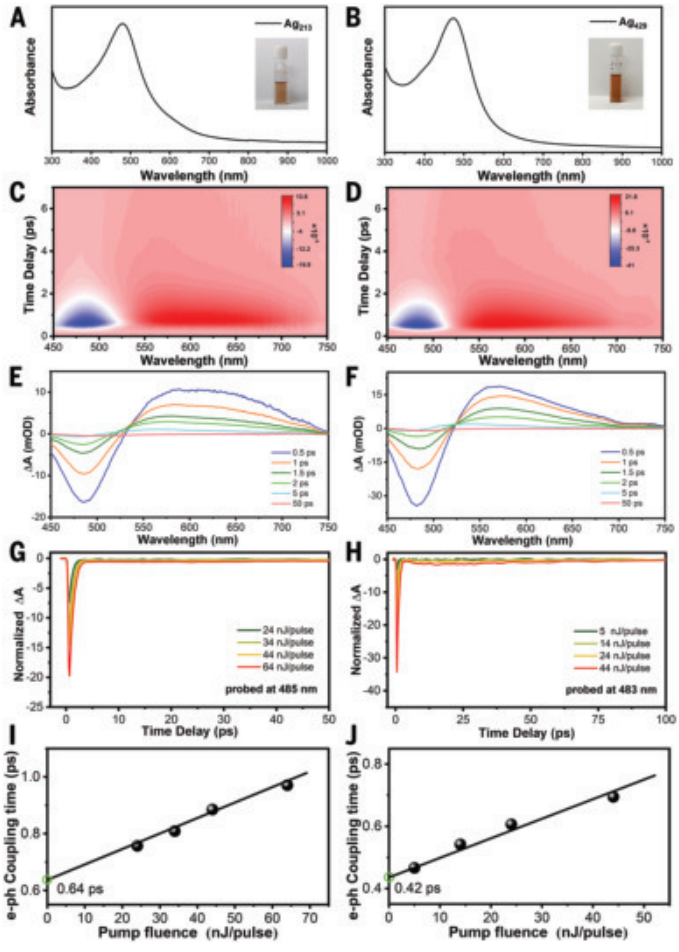


Fig. 5. Steady-state absorption spectra and excited-state dynamics of Ag_{213} and Ag_{429}. (**A** and **B**) UV-vis absorption spectra of Ag_{213} (A) and Ag_{429} (B) in CH_2Cl_2 at room temperature. (**C** and **D**) TA data maps of Ag_{213} (C) and Ag_{429} (D) pumped at 400 nm in CH_2Cl_2. (**E** and **F**) TA spectra of Ag_{213} (E) and Ag_{429} (F) as a function of time delay between 0.5 and 50 ps. (**G** and **H**) Kinetics traces of Ag_{213} (G) and Ag_{429} (H) probed at selected wavelengths with different laser excitation pulse energies. (**I** and **J**) Electron-phonon coupling time of Ag_{213} (I) and Ag_{429} (J) as a function of 400-nm pump fluence.

4.68×10^{16} W m^{-3} K^{-1}, which are comparable to those of Ag NPs (*32*) and are much larger than those of gold nanoclusters (1×10^{16} to 2×10^{16} W m^{-3} K^{-1}) of similar size (*31*). Based on the analysis of TA data under 400-nm excitation, the excited-state lifetimes of Ag_{213} and Ag_{429} are dependent on the pump fluence, displaying metallic electron dynamics.

The first thiolate-protected silver nanocluster $Ag_{44}(SR)_{30}^{4-}$ (Ag_{44}) exhibits a one-layered icosahedral structure (*33*, *34*). Ag_{44}, Ag_{112} (*18*), Ag_{213}, and Ag_{429} are silver nanoclusters with one-, two-, three-, and four-layered icosahedral cores, respectively. The UV-vis absorption spectra of Ag_{44} and Ag_{112} display multiple excitonic transitions, indicative of molecular characteristics. By contrast, Ag_{213} and Ag_{429} exhibit strong surface plasmon resonance absorption and pump fluence–dependent excited-state dynamics, confirming their metallic nature.

Discussion

We have isolated two cases of 20-fold twinned silver icosahedra, Ag_{213} and Ag_{429}, with precise atomic arrangement. Ag_{213} and Ag_{429} have 120 and 260 valence electrons, respectively, and they are metallic in nature. Ag_{213} and Ag_{429} have multilayered icosahedral Ag_{141} ($Ag_{13}@Ag_{42}@Ag_{86}$) and Ag_{297} ($Ag_{13}@Ag_{42}@Ag_{92}@Ag_{150}$) cores, respectively. Their atomically precise structures provide invaluable information for supporting the layer-by-layer evolution of silver icosahedra. Ag_{429} is the largest well-defined silver nanocluster containing Ag^0. This work proves that single-crystal x-ray diffraction is still the most powerful method to determine well-defined structures. Because examples of giant metal nanoclusters are rare, ligand engineering deserves extensive investigation in term of generating structural diversity of large-sized nanoclusters. In addition, heterometal doping may be a promising approach to improve the stability and enhance the properties of metal nanoclusters. It is expected that interesting functional changes will emerge as a result of the structural evolution of metal nanoclusters and NPs.

REFERENCES AND NOTES

1. Y. Sun, Y. Xia, *Science* **298**, 2176–2179 (2002).
2. L. Liu, A. Corma, *Chem. Rev.* **118**, 4981–5079 (2018).
3. Y. Xia, Y. Xiong, B. Lim, S. E. Skrabalak, *Angew. Chem. Int. Ed.* **48**, 60–103 (2008).
4. B. Wiley, Y. Sun, Y. Xia, *Acc. Chem. Res.* **40**, 1067–1076 (2007).
5. M. R. Langille, J. Zhang, M. L. Personick, S. Li, C. A. Mirkin, *Science* **337**, 954–957 (2012).
6. J. A. Scholl, A. L. Koh, J. A. Dionne, *Nature* **483**, 421–427 (2012).
7. S. Jeon *et al.*, *Science* **371**, 498–503 (2021).
8. M. Song *et al.*, *Science* **367**, 40–45 (2020).
9. V. K. LaMer, R. H. Dinegar, *J. Am. Chem. Soc.* **72**, 4847–4854 (1950).
10. U. Gasser, E. R. Weeks, A. Schofield, P. N. Pusey, D. A. Weitz, *Science* **292**, 258–262 (2001).
11. S. Auer, D. Frenkel, *Nature* **409**, 1020–1023 (2001).
12. Q. Zhang, J. Xie, J. Yang, J. Y. Lee, *ACS Nano* **3**, 139–148 (2009).
13. A. Mackay, *Acta Crystallogr.* **15**, 916–918 (1962).
14. A. Haji-Akbari *et al.*, *Nature* **462**, 773–777 (2009).
15. H. Wang, S. Zhou, K. D. Gilroy, Z. Cai, Y. Xia, *Nano Today* **15**, 121–144 (2017).
16. R. Jin, C. Zeng, M. Zhou, Y. Chen, *Chem. Rev.* **116**, 10346–10413 (2016).
17. Y. Li, R. Jin, *J. Am. Chem. Soc.* **142**, 13627–13644 (2020).
18. F. Hu, J.-J. Li, Z.-J. Guan, S.-F. Yuan, Q.-M. Wang, *Angew. Chem. Int. Ed.* **59**, 5312–5315 (2020).
19. F. Hu *et al.*, *J. Am. Chem. Soc.* **143**, 17059–17067 (2021).
20. C. Zeng, Y. Chen, A. Das, R. Jin, *J. Phys. Chem. Lett.* **6**, 2976–2986 (2015).
21. K. Reichenbächer, H. I. Süss, J. Hulliger, *Chem. Soc. Rev.* **34**, 22–30 (2005).
22. D. Wales, *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*, *Cambridge Molecular Science* (Cambridge Univ. Press, 2004).
23. A. I. Kirkland, P. P. Edwards, D. A. Jefferson, D. G. Duff, *Annu. Rep. Prog. Chem. Sect. C Phys. Chem.* **87**, 247–304 (1990).
24. Z. Sun *et al.*, *Nat. Commun.* **16**, 1595 (2025).
25. H. Yang *et al.*, *Nat. Commun.* **7**, 12809 (2016).
26. B. Bhattarai *et al.*, *Acc. Chem. Res.* **51**, 3104–3113 (2018).
27. J. Yan, B. K. Teo, N. Zheng, *Acc. Chem. Res.* **51**, 3084–3093 (2018).
28. Z. Lei, X.-K. Wan, S.-F. Yuan, Z.-J. Guan, Q.-M. Wang, *Acc. Chem. Res.* **51**, 2465–2474 (2018).
29. W.-Q. Shi *et al.*, *Science* **383**, 326–330 (2024).
30. M. Zhou, R. Jin, *Annu. Rev. Phys. Chem.* **72**, 121–142 (2021).
31. C. Yi, H. Zheng, L. M. Howard, C. J. Ackerson, K. L. Knappenberger, *J. Phys. Chem. C* **119**, 6307–6313 (2015).
32. J. H. Hodak, I. Martini, G. V. Hartland, *J. Phys. Chem. B* **102**, 6958–6967 (1998).
33. A. Desireddy *et al.*, *Nature* **501**, 399–402 (2013).
34. H. Yang *et al.*, *Nat. Commun.* **4**, 2422 (2013).

ACKNOWLEDGMENTS

F.H. acknowledges support from the Shuimu Tsinghua Scholar Program. We thank the National Center for Protein Sciences at the Shanghai Synchrotron Radiation Facility for the crystallographic data collection. **Funding:** This study was supported by the National Key R&D Program of China (2022YFA1503900), the National Natural Science Foundation of China (92361301 and 22401171), and the China Postdoctoral Science Foundation (2023T160357). **Author contributions:** Measurements and data analysis: F.H., G.Y., L.-M.Z., G.-J.L., Q.-M.W.; Project concept: Q.-M.W., F.H.; Synthesis: F.H., L.-M.Z.; Writing – original draft: F.H., G.-J.L.; Writing – review & editing: Q.-M.W., F.H., G.-J.L. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. The x-ray crystallographic structures reported in this article have been deposited at the Cambridge Crystallographic Data Centre (CCDC) (deposition nos. 2433339 for Ag_{213} and 2433340 for Ag_{429}).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adx6639
Materials and Methods; Supplementary Text; Figs. S1 to S22; Tables S1 to S4
Submitted 23 March 2025; accepted 26 June 2025
10.1126/science.adx6639

HORSE EVOLUTION

Selection at the *GSDMC* locus in horses and its implications for human mobility

Xuexue Liu[1,2,3]†, Yaozhen Jia[2]†, Jianfei Pan[1,2], Yanli Zhang[2], Ying Gong[2], Xintong Wang[2], Yuehui Ma[2], Nadir Alvarez[3,4,5], Lin Jiang[2]*, Ludovic Orlando[1]*

Horsepower revolutionized human history through enhanced mobility, transport, and warfare. However, the suite of biological traits that reshaped horses during domestication remains unclear. We scanned an extensive horse genome time series for selection signatures at 266 markers associated with key traits. We detected a signature of positive selection at *ZFPM1*—known to be a modulator of behavior in mice—occurring ~5000 years ago (ya), suggesting that taming was one of the earliest steps toward domestication of horses. Intensive selection at *GSDMC* began ~4750 ya with the domestication bottleneck, leading regulatory variants to high frequency by ~4150 ya. *GSDMC* genotypes are linked to body conformation in horses and to spinal anatomy, motor coordination, and muscular strength in mice. Our results suggest that selection on standing variation at *GSDMC* was crucial for the emergence of horses that could facilitate fast mobility in human societies ~4200 ya.

Horses have provided humans with fast mobility, traction, and new modes of warfare for more than 4000 years (*1*). Over the past decade, next-generation sequencing of large archaeological bone collections has transformed our understanding of horse domestication (*2*–*7*). Spatiotemporal shifts in genetic diversity identified the lower Don-Volga steppes as the origin of modern domesticated horses (DOM2) and dated the emergence of horse-based mobility to ~4200 years ago (ya) (*6*, *7*). Two genetic modifications—common in DOM2 but rare in other lineages—likely adapted locomotion and behavior to growing human demand and exposure (*6*). Genomic evidence also showed that Przewalski's horses (PrZ), long considered the last truly wild horses (*8*), descend from a separate domestication ~5500 ya on a distinct Central Asian lineage (*3*, *7*).

Concurrently, genomic tools and advanced phenotyping have mapped the genetic basis of key disorders and traits such as speed, ambling, and coat coloration. Targeted polymerase chain reaction (PCR) genotyping of ancient horses further informed the selection history of specific traits. This research, typically focused on single traits, has revealed early selection for coat-coloration diversification (*9*), fluctuating preferences for leopard spotted (*10*) and chestnut (*11*) phenotypes, and the medieval emergence of natural amblers (*12*). However, no study has traced the full spectrum of known trait-associated markers, despite rich genome time series (*2*–*7*) and improved methods for inferring selection timing and strength (*13*–*15*).

[1]Centre d'Anthropobiologie et de Génomique de Toulouse (CNRS/Université de Toulouse), Toulouse, France. [2]State Key Laboratory of Animal Biotech Breeding, Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS), Beijing, P. R. China. [3]Department of Genetics and Evolution, University of Geneva, Geneva, Switzerland. [4]Naturéum—State Museum of Natural Sciences of the Canton de Vaud, Lausanne, Switzerland. [5]Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland. ***Corresponding author. Email: ludovic.orlando@utoulouse.fr (L.O.); jianglin@caas.cn (L.J.)** †These authors contributed equally to this study.

We analyzed allelic trajectories at 266 trait-associated markers to track selection on behavior, coloration, conformation, locomotion, performance, and disease susceptibility. Our results suggest early selection on behavior, limited coat color diversification during initial domestication, and selection for increased body size from the Iron Age onward. Intensive selection for mutations regulating *GSDMC* transcription coincided with the rise of DOM2 breeding and large-scale horse-based mobility. These regulatory variants affect body conformation in horses and spinal anatomy, motor coordination, and limb muscular strength in mice.

Results

Horse demographic history

Ancient DNA time series directly document temporal allele frequency shifts, enhancing power to detect selection (*16*). Current methods infer selection when observed frequency shifts deviate from neutral expectations. To model demography, we analyzed the site-frequency spectrum of biallelic autosomal single nucleotide polymorphism (SNPs) in a global panel of 71 individuals from 65 DOM2 breeds and 15 PrZ horses (table S1 and fig. S1) using a two-step approach. First, we used SMC++ (*17*) to obtain prior information on divergence time and demographic change, accounting for structure within DOM2 (figs. S2 to S4). We then refined the model with fastsimcoal2 (*18*) (table S2), incorporating prior evidence of post-divergence contact (*8*) (Fig. 1A and figs. S5 and S6).

The best-fit model dates the DOM2-PrZ split to ~4367 generations ago [(ga); ~32,316 ya, based on the most recent generation time estimate] (*7*), coinciding with increased climatic instability and intensifying glaciation (*19*). While maintaining post-divergence contact, both lineages remained largely isolated starting from the onset of the Last Glacial Maximum [~26.5 thousand years ago (kya)] (*20*) until the early 20th century, when captive breeding rescued PrZ horses from extinction (*8*). This discontinuous contact history paralleled dynamic demographic trajectories (Fig. 1A) as PrZ populations declined steadily after divergence and more sharply over the last ~94 generations (~696 ya), whereas DOM2 experienced a sharp, albeit brief, expansion post-divergence before declining and expanding again ~4200 ya with domestication. Our results largely align with previous models (*4*, *8*, *21*) and support the widespread dispersal of DOM2 horses by the late third millennium BCE (before common era), as evidenced in archaeological assemblages and material culture (*6*, *7*).

Identification of selection signatures

We used the DOM2 demographic model and a comprehensive horse genome time series to test for selection at 266 trait-associated markers (table S3). After excluding 69 loci with (nearly) flat allelic trajectories, 197 were left for analysis (tables S4 and S5). We applied CLUES2 (*13*) to test for selection over neutrality (Fig. 2, D to F, fig. S7, and table S6), under models assuming constant selection or one to two shifts in selection regimes (Supplementary Information section 4, fig. S7, and table S6). Timing, dominance, and selection coefficients were inferred through grid search (Fig. 2, G to I). SLiM (*22*) simulations confirmed sufficient power to detect regime shifts (figs. S8 to S10), although selection onset could be slightly underestimated.

Out of the given candidates, 44 passed multiple testing correction (Fig. 1B and table S6). However, 17 of 197 neutral-matched controls also appeared under selection, indicating a false discovery rate (FDR) of 8.63% (table S7). Raising the significance threshold reduced false discovery and revealed more selection candidates among trait-associated than a matched set of genome-wide random markers ($-\log_{10}$ *P*-value = 7.11; FDR = 5.07%, enrichment-fold relative to neutral set = 3.3 versus 2.4, respectively; tables S7 and S8).

We further refined candidates using a Bayesian framework based on particle marginal Metropolis-Hastings (PMMH) (*14*), which models

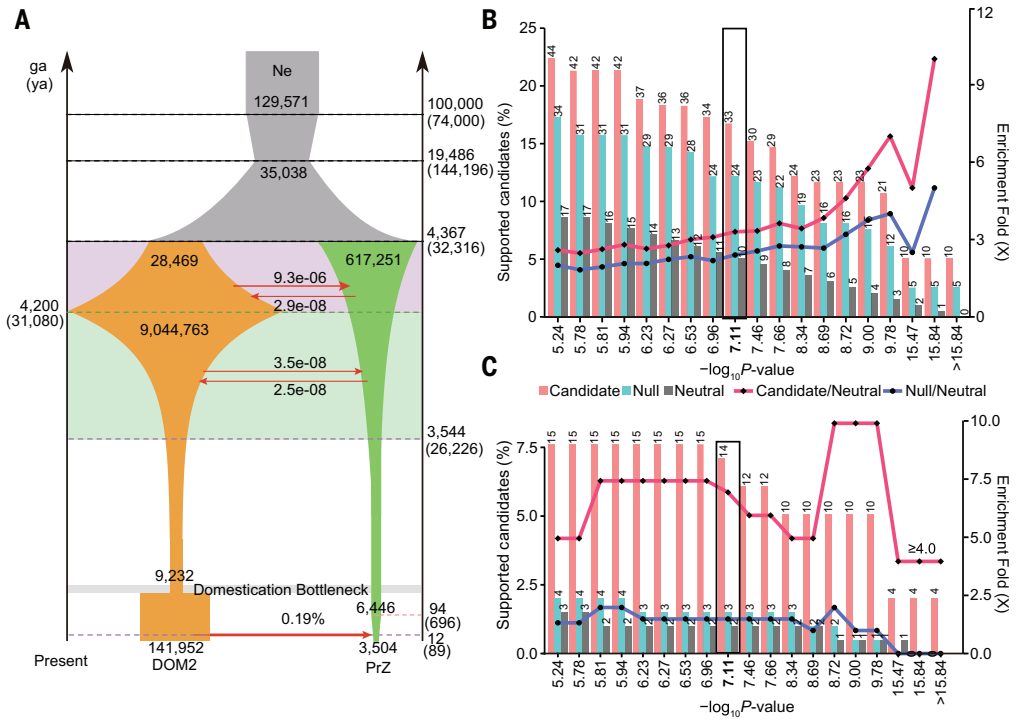


Fig. 1. Horse demography and selection detection power. (**A**) Best fastsimcoal2 (*18*) demographic model for the DOM2 (orange) and PrZ (green) lineages, with periods of contact post-divergence. Migration rates are shown near the red arrows. The numbers reported on each lineage indicate the effective size (Ne) at the time marked by the dashed lines. Time is shown in generations ago (ga) and years ago (ya), assuming the average generation time of 7.4 years from Librado and colleagues (*7*). (**B**) CLUES2 (*13*) selection candidates for various $-\log_{10}P$-value thresholds. In addition to trait-associated markers (red), matched genome-wide sets of neutral loci (gray) and random variants (blue) are considered. The left vertical axis indicates the fraction of candidates passing the $-\log_{10}P$-value threshold (bars), while the right vertical axis represents the ratio of the number of selection candidates identified amongst trait-associated markers and random variants, relative to null loci (continuous lines). (**C**) Same as (B), but restricted to those candidate markers also supported by PMMH (*14*) analyses. The black boxes highlight the $-\log_{10}P$-value threshold retained. Numbers above each bar indicate the number of loci that passed the respective threshold.

two epochs and jointly estimates dominance and selection coefficients (Fig. 2, J to L, figs. S11 to S23, and table S9). PMMH (*14*) validated 14 CLUES2 (*13*) hits and rejected all but two neutral controls, reducing FDR to 1.02% (Fig. 1C and table S10). Trait-associated markers yielded 4.67 times more selection candidates than a genome-wide matched set of random markers (14 versus 3). The latter included more markers than the neutral set, suggesting that additional true selection targets remain undetected in the genome time series. We focused on the 14 trait-associated loci identified (Fig. 3, with detailed examples in Fig. 2 and figs. S11 to S23), leveraging their known genotype-phenotype associations for clearer interpretation.

Selection history of trait-associated markers

Our conservative set of trait-associated markers reveals selection regime shifts over the last ~10 to ~680 generations (~74 to ~5032 ya; table S6). It includes causative variants with Mendelian inheritance as well as heritable variants of unknown, small, mild, or severe effects (table S3). Below, we outline nine markers with the most notable phenotypic effects (Figs. 2 and 3; see Supplementary section 4 for the remaining five).

Three markers influencing body size showed evidence of selection: The first is a *TBX3* enhancer, validated in knockout (KO) mice, affecting hindlimb development, and accounting for 20.27% of withers height variation in Asian breeds (~95 to 155 cm) (*23*). The "G" allele was selected against in early domestication, then underwent strong positive selection ~370 ga (~2738 ya), followed by milder selection since ~220 ga (~1628 ya) (Fig. 2, C, F, I, and L, and Fig. 3). The second, *LCORL/NCAPG_1*, ~59.5 kb upstream of the *LCORL/NCAPG* gene, is one of the most influential body size loci, explaining 68.50% of global withers height variation (*24*). The "C" allele is associated with increased stature, leading to an average additive size increment of ~1.0 and ~3.7 cm in Franches-Montagnes and Thoroughbreds (*25*), respectively, and of ~3.6% of the withers height in Spanish purebreds (*26*) (table S3). Selection shifted from negative to mild positive ~350 ga (~2590 ya) (fig. S12). The third marker, ~4.2 Mb upstream of *LCORL/NCAPG_1*, is part of a quantitative tract locus affecting height (*27*). Positive selection for the "A" allele began ~170 ga (~1258 ya) (fig. S13). The timing of these shifts follows the rise of large-scale mounted cavalry (~900 BCE) (*28*), suggesting coordinated, moderate-to-strong selection (coefficients = 0.83 to 22.75%) for increased body size from the Iron Age onward.

Four selected markers relate to coat coloration (Fig. 3). The first affects *MC1R*, with horses homozygous for the recessive "T" allele developing chestnut coats (*29*). This allele was positively selected for until ~260 ga (~1924 ya, coefficients = 1.36 to 1.69%), but has since faced negative selection (−1.05 to −1.53%), suggesting shifting coat coloration preferences during antiquity (fig. S14 and table S9). The three other markers are associated with silver coats [*PMEL_5* (*30*), fig. S15], and white patterning [*KIT_3* (*31*), fig. S16, and *KIT_28*, fig. S17 (*32*)]. *KIT_28* underwent positive selection (0.64 to 1.39%) through most of domestication until 20 ga (148 ya), while *PMEL_5* and *KIT_3* experienced negative selection until 10 to 30 ga (74 to 222 ya) (−0.13 to −0.50%; Fig. 3 and table S9). No positive selection was detected for 18 other coat coloration markers, including those for leopard spotting, white, gray, cream, or modifiers for white markings (table S6). These results challenge the hypothesis of early widespread selection for full coat color diversity (*9*, *33*), pointing instead to fluctuating preferences shaped by sociocultural and geographic contexts.

Two loci show selection during the earliest stages of DOM2 domestication (*6*). One affects *ZFPM1*, a modulator of anxiety and contextual fear learning in mice (*34*). DNA methylation at this locus is negatively correlated with well-being in humans (*35*). Selection for the "C" allele, ~13.4 kb downstream of the longest transcript, began ~680 ga (~5032 ya), ~40 generations (~296 years) before the domestication bottleneck started [~640 ga, ~4736 ya; (*6*)] (Fig. 2, A, D, G, and J, Fig. 3, and tables S6 and S9), possibly reflecting early taming efforts prior to DOM2 formation. The second marker, ~6.7 kb upstream of *GSDMC*, involves a "C" allele previously hypothesized to affect horseback anatomy (*6*). Human variants near *GSDMC* are strongly associated with chronic back pain and lumbar spinal stenosis—a lower back condition that causes pain and impairs walking (*36*, *37*). The "C" allele had one of the

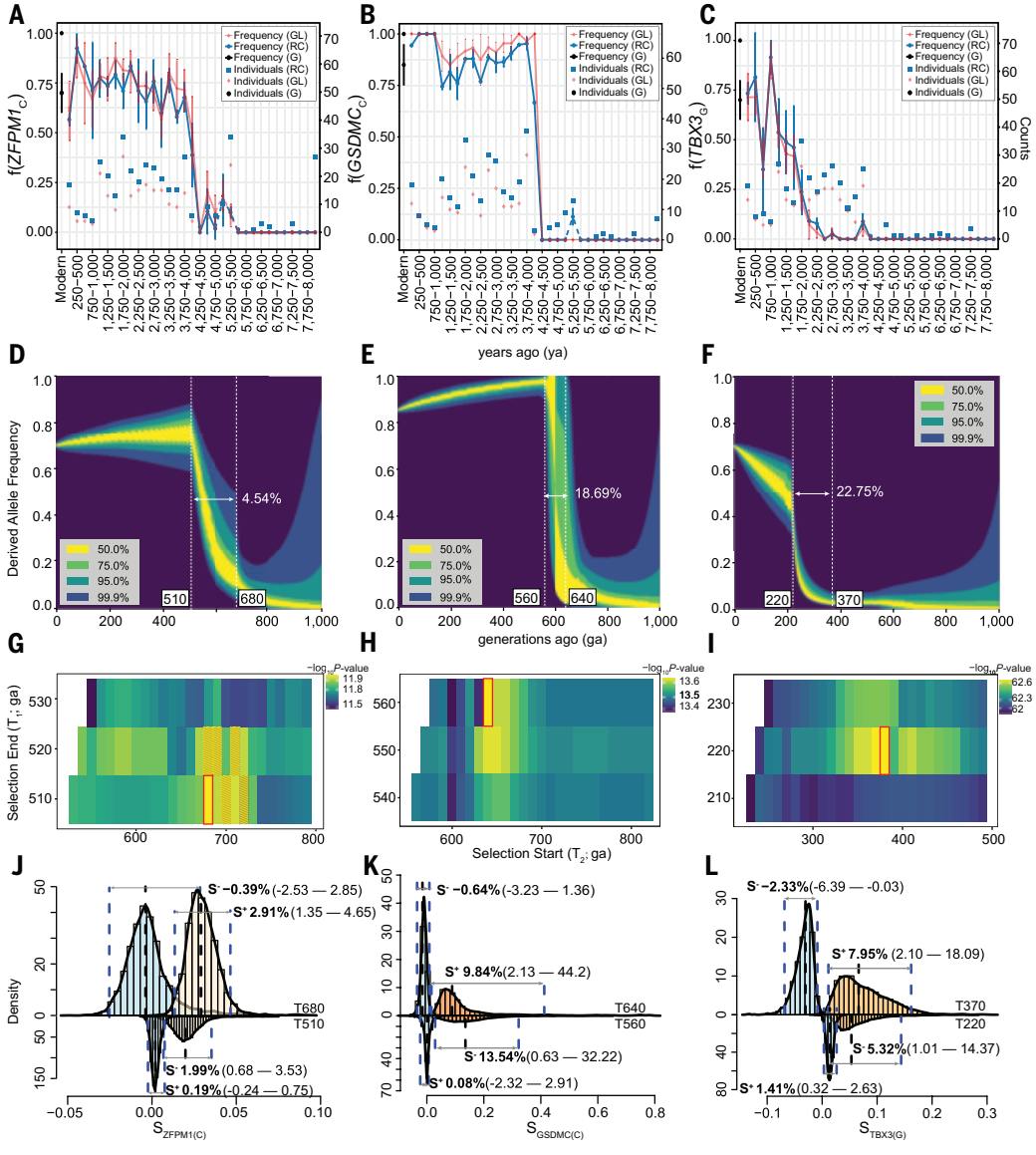


Fig. 2. Allele trajectories and selection analyses at three loci. (**A**) mapDATAge allele trajectory for the "C" allele at position chr3:35,763,853 of *ZFPM1* (time windows = 250 years, step size = 100 years). Allele frequencies are estimated using read counts (RC), or genotype likelihoods (GL), from the corresponding numbers of individuals showing sufficient sequence coverage. The present-day frequency is estimated from the genotypes of modern DOM2 horses (G) and is shown with black dots and error bars. Solid lines indicate DOM2 and their closest pre-DOM2 relatives, while dashed lines extend estimates to other population groups. The left vertical axis represents allele frequencies (continuous lines) while the right vertical axis indicates the number of individuals with non-null sequence coverage (squares and diamonds). (**B**) Same as (A) but for the "C" *GSDMC* allele (chr9:73,114,851). (**C**) Same as (A), but for the "G" *TBX3* allele (chr8:20,644,555). (**D**) CLUES2 (*13*) posterior distribution of *ZFPM1* allelic trajectory. The percentage indicates the point-estimate returned for the selection coefficient at the time of selection, between T_1 (= 510) and T_2 (= 680) generations ago (ga). (**E**) Same as (D), but for the "C" *GSDMC* allele, with T_1 = 560 and T_2 = 640 ga. (**F**) Same as (D), but for the "G" *TBX3* allele, with T_1 = 220 and T_2 = 370 ga. (**G**) $-\log_{10}$*P*-value surface of CLUES2 (*13*) best-fitting models for a combination of T_1 and T_2 variables (*ZFPM1*). The best-fitting combination is highlighted in red. Hatched areas indicate analyses flagged as "unreliable" by CLUES2 (*13*). (**H**) Same as (G), but for the "C" *GSDMC* allele. (**I**) Same as (G), but for the "G" *TBX3* allele. (**J**) Posterior distributions of selection coefficients at *ZFPM1* before (gray) and after 680 ga (top), or before (colored) and after 510 ga (bottom). (**K**) Same as (J), but for the "C" *GSDMC* allele and considering 640 ga and 560 ga. (**L**) Same as (J), but for the "G" *TBX3* allele and considering 370 ga and 220 ga. Colors reflect the gradient shown in Fig. 3.

top two selection coefficients (9.84 to 18.69%; Fig. 2, B, E, H, and K, and tables S6 and S9), with selection spanning ~640 to 560 ga (~4736 to 4144 ya; Fig. 3), aligning closely with the domestication bottleneck (~640 to 540 ga, ~4736 to 3996 ya). It thus emerges as a key early target in DOM2 domestication.

Functional consequences of *GSDMC* genetic variation

The selection history of the "C" allele at *GSDMC* prompted investigation into associated genetic and functional variation. The EquCab3 reference genome (*38*) includes two L1 transposable elements (L1M3c and L1-5_EC), ~4.0 kb downstream and 24.1 kb upstream of the start codon, respectively (Fig. 4A). Initial PCR genotyping using EquCab3-based primers failed, suggesting structural variation. We refined the locus using PacBio de novo assemblies (table S11), enabling successful primer design (table S12).

PCR genotyping 247 horses from nine Asian breeds revealed near-complete linkage between the "C" allele and both L1 insertions (Fig. 4B, figs. S24 and S25, and table S13), a pattern confirmed in resequencing data from 981 globally distributed horses (fig. S26 and table S14). Linkage disequilibrium and "C" allele frequencies were lower in nonspecialized Asian breeds, especially in northern China and Mongolia (figs. S27 and S28), consistent with allele frequencies declining after ~2200 to 2450 ya in Asia (fig. S29). Restricting selection analyses to pre-2300 ya samples confirmed strong positive selection (10.46 to 20.82%; Fig. 3 and fig. S30), overlapping the domestication bottleneck.

These findings suggest selection on an extended haplotype, although the causal variant remains unknown. Over the past ~2300 years, this haplotype declined in frequency among several Asian breeds, primarily used for pastoralism and agriculture in China (*39*) (fig. S27). In Chinese breeding, the body length-to-height ratio discriminates horses built for racing from others (*40*). This trait was significantly associated with *GSDMC* genotypes

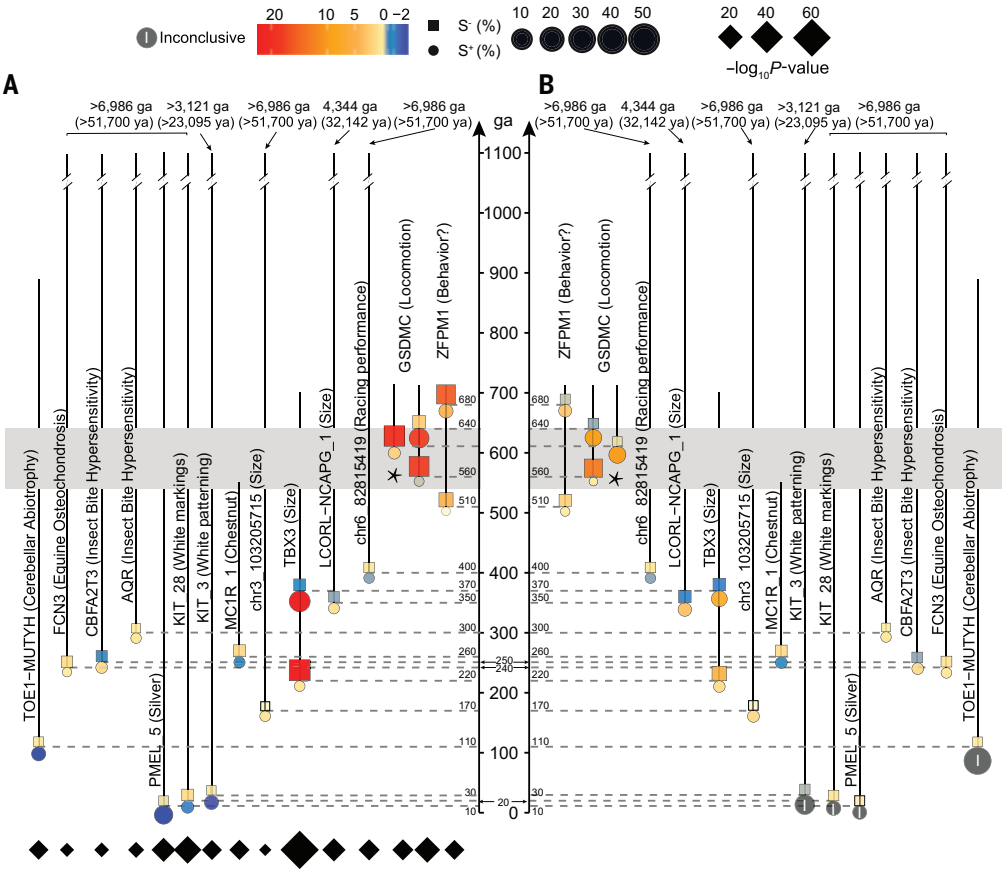


Fig. 3. Overview of the selection history of 14 trait-associated markers. The figure shows those selection candidates confirmed in both CLUES2 (*13*) (**A**) and PMMH (*14*) (**B**) analyses. Squares and circles represent selection coefficients before and after the time indicated by dashed horizontal lines, respectively, with colors and sizes proportional to the selection intensity. Vertical segments indicate the time of first occurrence of the derived allele in the genome time series. The DOM2 domestication bottleneck is highlighted in gray. Black diamonds indicate the $-\log_{10}P$-value of CLUES2 (*13*). PMMH (*14*) posterior distributions of selection coefficients are shown as "inconclusive" (I) if the 95% HPD overlaps with zero and spans a range greater than 0.50.

in Ningqiang horses—both the biallelic SNP and L1 insertions (Fig. 4, C to E)—and with the SNP alone in a global breed panel (Fig. 4F and tables S15 and S16). These results point to a role in body conformation and/or locomotion.

The SNP position within the *GSDMC* promoter, together with nearby structural variants, suggested regulatory potential. Transient transfection in HEK293T cells confirmed that both L1 insertions down-regulate transcription (Fig. 4G and figs. S31 and S32), a pattern mirrored in muscle biopsies from L1M3c homozygous Kazakh horses (Fig. 4H). Combined, these results indicate that the most prevalent DOM2 haplotype down-regulates *GSDMC*, reshaping body conformation and potentially locomotion.

To test functional effects in vivo, we generated a mouse model carrying a deletion spanning exon 12 and the 3′-UTR (Fig. 4, I and J, Δ12; table S17), and conducted expression, anatomical, and behavioral assays (fig. S35). In 12-week-old wild-type (WT) mice, *GSDMC* showed ubiquitous transcription, with large variation across 12 tissues (fig. S33), and lowest expression in muscle. Δ12 heterozygous (HET) and KO mice showed significantly reduced transcription in kidney, colon, liver, and claws (Fig. 4K), echoing the results from our in vitro transfection assays.

CT scans revealed anatomical differences across *GSDMC* genotypes. The kyphosis index (KI), measuring spinal curvature (*41*) (Fig. 4L), was significantly lower in KO mice (Fig. 4M), consistent with the *GSDMC* genetic association with lumbar stenosis in humans (*36*, *37*). Histological analysis of the first lumbar spine (LS1; Fig. 4, L and N to Q, and fig. S34) revealed major differences in KO mice, involving reduced intervertebral disc area (Fig. 4O), disc height (Fig. 4P), and its most central part [i.e., the (total) nuclear pulposus; Fig. 4Q], alongside expansion of its most external part (annulus fibrosus; Fig. 4Q). We hypothesized that these spinal anatomical changes could alter locomotion.

To test this, we assessed muscular strength and motor coordination (Fig. 4, R to U, and figs. S36 to S39). KO mice exhibited greater forelimb strength, hanging longer on limb hanging wire than the WT (Fig. 4R), although no differences were found in four-limb suspension experiments (figs. S36 to S39). On a rotating rod, KO mice had significantly increased latency-to-fall than the WT and HET under standard (Fig. 4S) and accelerating conditions (Fig. 4T), and when weighted with ~10% body mass attached to the tail (Fig. 4U).

Our mouse models demonstrate that the *GSDMC* locus shapes spinal anatomy and locomotor traits, affecting both motor coordination and forelimb strength. In horses, the same locus is associated with body conformation and shows strong selection, beginning with the domestication bottleneck and ending with the DOM2 expansion (*6*, *7*). These findings suggest that standing variation at *GSDMC* was instrumental to the emergence of horses with enhanced locomotor capacity, facilitating increased human mobility and speed of travel from ~4200 ya onward.

Discussion

Despite sufficient statistical power, we detected no evidence of shifting selection regimes at most loci, suggesting that many variants with known phenotypic effects in modern horses were not subject to strong selection in the deeper past. Alternatively, this paucity of detected candidates may reflect spatially restricted selection nearly fixing or purging variants in local populations, leaving minimal impact on global allele frequency trajectories. Another limitation stems from trait architecture. Our framework is optimized to detect selective sweeps under low polygenicity, in which mutations exert large, consistent fitness effects. However, many traits, including temperament, performance, and endurance (*42*), are highly polygenic, shaped by numerous alleles with small, often antagonistic effects. In such traits, directional selection drives rapid mean phenotypic change early on, followed by a slower plateau phase as populations approach the new phenotypic optimum (*43*). In the early phase, opposing-effect alleles exhibit minimal frequency shifts, reducing power to detect selection from time-stamped trajectories. Consequently, many targets of selection on complex traits likely remain undetected in this study. This also holds, to a

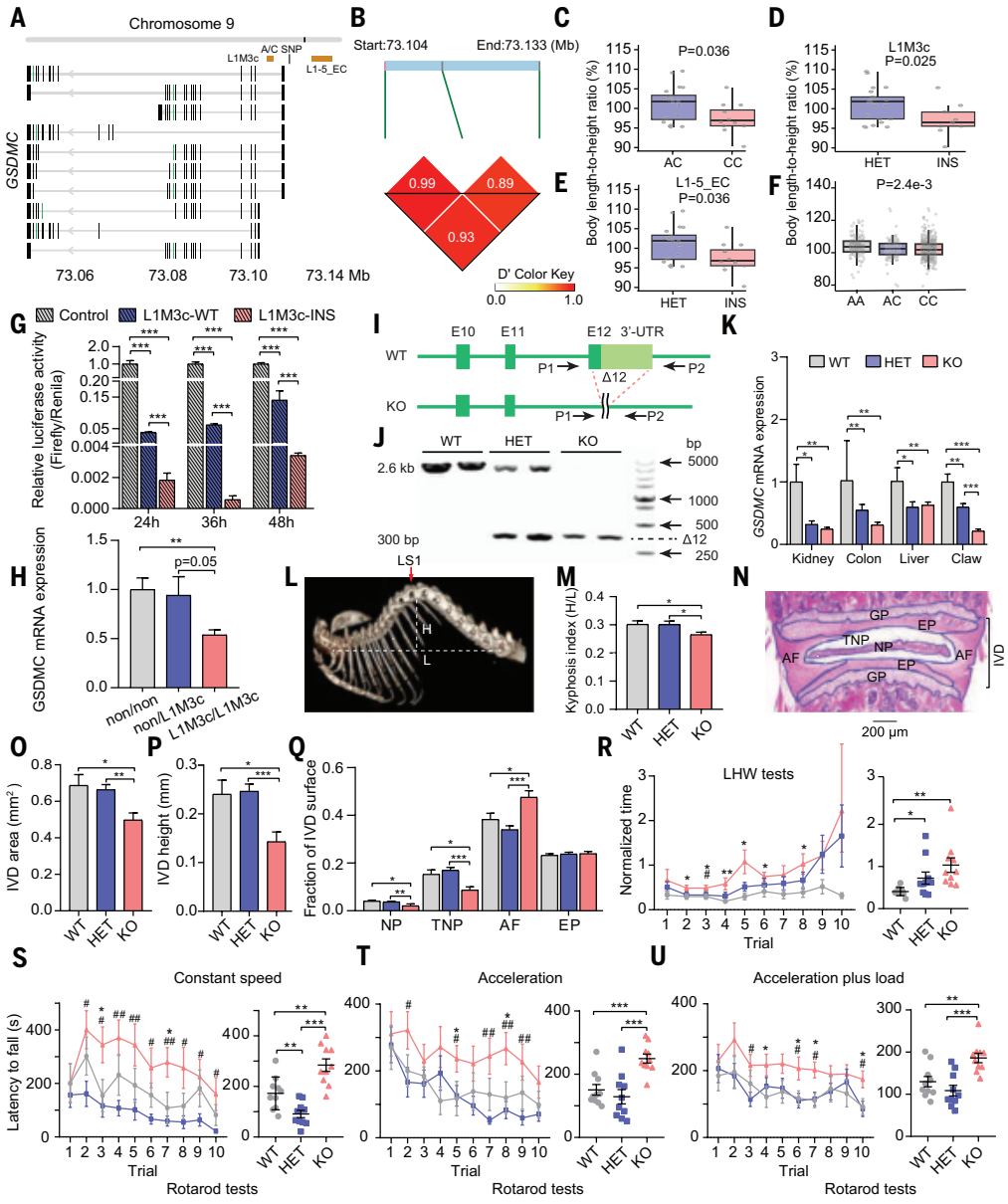


Fig. 4. Functional consequences of genetic variation at *GSDMC*. (**A**) UCSC annotations for the EquCab3 reference (*38*). (**B**) Linkage disequilibrium between the SNP and the two flanking transposable elements in 247 Chinese horses. (**C**) Genetic association analysis between the selected *GSDMC* biallelic SNP marker and body length-to-height ratio in 32 Ningqiang horses. (**D**) Same as (C) for the L1M3c genotypes. (**E**) Same as (D) for the L1-5_EC genotypes. (**F**) Same as (C) in 836 horses. (**G**) Luciferase-based reporter assay for constructions carrying the L1M3c insertion (L1M3c-INS) or not (L1M3cWT). (**H**) Real-time PCR quantification of *GSDMC* muscular transcription for the three genotypes underlying L1M3c insertion ("non" indicates the absence of the insertion). (**I**) Genomic background of WT and KO mice. (**J**) PCR-based mouse genotyping. (**K**) Real-time PCR quantification of *GSDMC* transcription in mouse tissues. (**L**) CT scan of the murine thoracic ridge and spine, highlighting the first lumbar spine (LS1). H and L indicate the two measurements for Kyphosis Index calculation (KI = H/L). (**M**) KI. (**N**) Hematoxylin & Eosin staining of LS1 intervertebral disc (IVD), With annulus fibrosus (AF), end plate (EP), growth plate (GP), nuclear pulposus (NP) and total nuclear pulposus (TNP). (**O**) IVD area. (**P**) IVD height. (**Q**) Fraction of IVD area represented by NP, TNP, AF, and EP. (**R**) Performance on 10 Limb Hanging Wire (LHW) tests. (**S**) Latency-to-fall on rotarod rotating at constant speed ($n = 10$ tests). (**T**) Same as (S) but with accelerating rotation. (**U**) Same as (T), with a load corresponding to 10% of their body mass attached to the mouse tails. *P*-values of 0.05, 0.01 and inferior to 0.01 are shown with one, two or three asterisks (*; WT versus KO) or hashtags (#; KO versus HET), respectively.

lesser extent, for monogenic or weakly polygenic traits, given the modest enrichment of selection candidates in genome-wide matched random sets relative to neutral loci.

Nonetheless, our analyses uncovered significant selection shifts at loci associated with increased withers height, broadly coinciding with the rise of mounted cavalry in warfare (*28*). Unlike in humans, horse height is influenced by a small number of loci, with four explaining up to 83% of global size variation (*24*). Still, the selected variants themselves confer only modest size increases. Although methods for estimating withers height from fragmentary skeletal remains have limitations (*44*), osteometric data indicate average size increases in western Europe from ~125 cm during the Iron Age to ~135 cm in the Roman period (*45*), and ~139 cm in the post-medieval era (*46*). Individual, regional, and contextual variation has also been reported (*45*, *47*, *48*), highlighting the role of nongenetic factors, including nutrition, parasitic burden, and management practices in shaping body size.

Our findings also challenge prevailing models positing selection for diverse coat coloration phenotypes as a hallmark of early horse domestication (*9*). Instead, we propose that an early behavioral shift, driven by selection at the *ZFPM1* locus, paved the way for DOM2 domestication. This locus lies ~230.6 kb upstream of another locus showing signatures of regional adaptation between Asian and European horses (*49*) and was the earliest to undergo positive selection before the domestication bottleneck. In conditional KO mice, *ZFPM1* modulates anxiety and contextual fear memory through neuronal positioning in the dorsal raphe, a core serotonergic brain nucleus (*34*). These findings support the hypothesis that enhanced tameness and stress tolerance were key domestication prerequisites, facilitating closer human-animal interactions and adaptation to life in captivity. However, *ZFPM1* is pleiotropic, as it is also involved in red blood cell production and terminal differentiation (*50*). Future work is needed to clarify the phenotypic consequences of selection at this locus in horses.

Our results indicate that selection at the *GSDMC* locus was instrumental to the emergence and success of the DOM2 bloodline. This locus showed one of the strongest signals of positive selection, precisely concurring with the DOM2 bottleneck. *GSDMC* genetic variation influences the body length-to-height ratio in horses and gene transcription. KO mice exhibit altered spinal anatomy (specifically, decreased curvature through intervertebral disc remodeling), and show changes in motor coordination and forelimb strength. These phenotypes parallel human GWAS linking *GSDMC* to lumbar spinal stenosis and chronic back pain (*36*, *37*), suggesting a conserved role in shaping axial skeleton structure and locomotion. We propose that selection acted on standing regulatory variation at *GSDMC*, enhancing locomotor capacity and contributing to the rise of horse-based mobility ~4200 ya (*6*, *7*). This transformation marked a turning point in human history, as equestrianism became foundational to warfare, transportation, and societal organization until the advent of motorized vehicles.

REFERENCES AND NOTES

1. P. Kelekna, *The Horse in Human History* (Cambridge Univ. Press, 2009).
2. P. Librado *et al.*, *Science* **356**, 442–445 (2017).
3. C. Gaunitz *et al.*, *Science* **360**, 111–114 (2018).
4. A. Fages *et al*, *Cell* **177**, 1419–1435.e31 (2019).
5. L. Orlando, *Annu. Rev. Genet.* **54**, 563–581 (2020).
6. P. Librado *et al.*, *Nature* **598**, 634–640 (2021).
7. P. Librado *et al.*, *Nature* **631**, 819–825 (2024).
8. C. Der Sarkissian *et al.*, *Curr. Biol.* **25**, 2577–2583 (2015).
9. A. Ludwig *et al.*, *Science* **324**, 485 (2009).
10. A. Ludwig *et al.*, *Phil. Trans. Roy. Soc. B* **370**, 20130386 (2015).
11. S. Wutke *et al.*, *Sci. Rep.* **6**, 38548 (2016).
12. S. Wutke *et al.*, *Curr. Biol.* **26**, R697–R699 (2016).
13. A. H. Vaughn, R. Nielsen, *Mol. Biol. Evol.* **41**, msae156 (2024).
14. Z. He, X. Dai, W. Lyu, M. Beaumont, F. Yu, *Mol. Biol. Evol.* **40**, msad008 (2023).
15. W. T. T. Taylor, *et al*, *Science* **379**, 1316–1323 (2023).
16. A.-S. Malaspinas, *Mol. Ecol.* **25**, 24–41 (2016).
17. J. Terhorst, J. A. Kamm, Y. S. Song, *Nat. Genet.* **49**, 303–309 (2017).
18. L. Excoffier *et al.*, *Bioinformatics* **37**, 4882–4885 (2021).
19. I. Malmierca-Vallet, L. C. Sime, DO Community Members, *Clim. Past* **19**, 915–942 (2023).
20. P. U. Clark *et al.*, *Science* **325**, 710–714 (2009).
21. M. Schubert *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, E5661–E5669 (2014).
22. B. C. Haller, P. W. Messer, *Am. Nat.* **201**, E127–E139 (2023).
23. X. Liu *et al.*, *Curr. Biol.* **32**, 480–487.e6 (2022).
24. S. Makvandi-Nejad *et al.*, *PLOS ONE* **7**, e39929 (2012).
25. H. Signer-Hasler *et al.*, *PLOS ONE* **7**, e37282 (2012).
26. N. Sevane, S. Dunner, A. Boado, J. Cañon, *Anim. Genet.* **47**, 570–578 (2016).
27. J. Tetens, P. Widmann, C. Kühn, G. Thaller, *Anim. Genet.* **44**, 467–471 (2013).
28. R. Drews, *Early Riders: The Beginnings of Mounted Warfare in Asia and Europe*. (Routledge, 2004).
29. L. Marklund, M. J. Moller, K. Sandberg, L. Andersson, *Mamm. Genome* **7**, 895–899 (1996).
30. M. Reissmann, J. Bierwolf, G. A. Brockmann, *Anim. Genet.* **38**, 1–6 (2007).
31. L. Patterson Rosa *et al.*, *J. Hered.* **112**, 447–451 (2021).
32. B. Haase *et al.*, *PLOS ONE* **8**, e75071 (2013).
33. A. Linderholm, G. Larson, *Semin. Cell Dev. Biol.* **24**, 587–593 (2013).
34. L. Tikker *et al.*, *J. Neurosci.* **40**, 8669–8682 (2020).
35. B. M. L. Baselmans *et al.*, *Twin Res. Hum. Genet.* **18**, 710–719 (2015).
36. P. Suri *et al.*, *PLOS Genet.* **14**, e1007601 (2018).
37. H. Jiang *et al.*, *Sci. Rep.* **10**, 21069 (2020).
38. T. S. Kalbfleisch *et al.*, *Commun. Biol.* **1**, 197 (2018).
39. Z. Li, Thesis (Murray State University, 2019).
40. C. Chao, C. Fang, L. Liu, W. Liu, *Herbivorous Livestock*, **146**, 17–25 (2017).
41. W. B. Katzman, L. Wanek, J. A. Shepherd, D. E. Sellmeyer, *J. Orthop. Sports Phys. Ther.* **40**, 352–360 (2010).
42. A. Ricard *et al.*, *Front. Genet.* **8**, 89 (2017).
43. L. K. Hayward, G. Sella, *eLife* **11**, e66697 (2022).
44. V. Onar, O. Kahvecioğlu, D. Erdikmen, H. Alpak, C. Parkan, *Revue de médecine vétérinaire* **169**, 157–165 (2018).
45. C. J. Johnstone, Thesis (University of York, 2004).
46. C. Ameen *et al.*, *The International Journal of Equine and Equestrian History* **1**, 100–119 (2021).
47. S. Lepetz, K. Debue, D. Batsukh, in *Masters of the steppe: The Impact of the Scythians and Later Nomad Societies of Eurasia*, S. Pankova, S. J. Simpson, Eds. (Archaeopress, 2020), pp. 227–247.
48. G. Piličiauskienė *et al.*, *Animals* **12**, 1549 (2022).
49. H. Han *et al.*, *BMC Genomics* **24**, 35 (2023).
50. I.-M. Roussis *et al.*, *PLOS Genet.* **21**, e1011617 (2025).

ACKNOWLEDGMENTS

We are grateful to P. Librado and colleagues from CAGT for support and discussions. **Funding:** This work was funded by the following: The National Natural Science Foundation of China (32002144 and 32222079) and The Agricultural Science and Technology Innovation Program of China (ASTIP-IAS01) (to L.J.); The European Union's Horizon 2020 research and Innovation programme under the Marie Sklodowska-Curie grant agreement 101027750 (to X.L.); The Swiss National Science Foundation (310030_188680) (to N.A.); The International Research Programme AnimalFarm (CNRS/Université Toulouse); The CNRS International Emerging Action (HorseBackRiding); The European Research Council (ERC, CoG PEGASUS 681605, and SyG Horsepower 101071707) (to L.O.). **Author contributions:** Conceptualized the study: L.O. Coordinated the study: L.O. Funding: X.L., N.A., L.J., and L.O. Infrastructure and reagents: L.J. and L.O. Genome analyses: X.L. and L.O. Functional validation: X.L., Y.J., J.P., Y.Z., X.W., Y.G., Y.M., L.J., and L.O. Wrote the manuscript: L.O. Wrote the Supplementary Information: L.O., with input from X.L. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adp4581
Materials and Methods; Supplementary Text; Figs. S1 to S39; Tables S1 to S17; References (*51*–*125*); MDAR Reproducibility Checklist; Movies S1 to S3

10.1126/science.adp4581

CLIMATE MITIGATION

Land availability and policy commitments limit global climate mitigation from forestation

Yijie Wang[1], Yakun Zhu[1]†, Susan C. Cook-Patton[2,3], Wenjuan Sun[4], Wen Zhang[5], Philippe Ciais[6], Tingting Li[5], Pete Smith[7], Wenping Yuan[8], Xudong Zhu[9], Josep G. Canadell[10], Xiaopeng Deng[1], Yifan Xu[1]‡, Hao Xu[8], Chao Yue[11], Zhangcai Qin[1]*

Forestation (afforestation and reforestation) could mitigate climate change by sequestering carbon within biomass and soils. However, global mitigation from forestation remains uncertain owing to varying estimates of carbon sequestration rates (notably in soil) and land availability. In this study, we developed global maps of soil carbon change that reveal carbon gains and losses with forestation, primarily in the topsoil. Constraining land availability to avoid unintended albedo-induced warming and safeguard water and biodiversity (389 million hectares available for forestation globally) would sequester 39.9 petagrams of carbon by 2050, substantially below previous estimates. This estimate drops to 12.5 petagrams of carbon with land further limited to existing policy commitments (120 million hectares). Achieving greater mitigation requires expanding dedicated forestation areas and strengthening commitments from nations with considerable but untapped potential.

Forests deliver multiple economic and environmental benefits to society (*1*, *2*). Yet global forest loss has persisted in recent decades, primarily because of disturbances associated with climate change and human activities (*3*, *4*). Forestation aims to increase forest cover by establishing trees on unforested land through reforestation and afforestation. Reforestation involves land that has recently lost its forest, whereas afforestation focuses on land not covered with forest in recent history (e.g., cropland, bare land). Both practices could capture additional carbon within biomass and soils (*5*, *6*). If strategically implemented with careful consideration of location, timing, tree species, and management, forestation could become a cornerstone natural climate solution (*7*, *8*), contributing substantially to rebuilding land carbon stocks and supporting the Paris Agreement's ambitious climate targets (*9*). Current estimates of climate mitigation from global forestation vary widely, ranging from <1 petagrams of carbon per year (Pg C $year^{-1}$) to >10 Pg C $year^{-1}$ (*7*, *8*, *10*, *11*). In comparison, global fossil CO_2 emissions in 2023 were 10.1 Pg C $year^{-1}$ (*12*).

These disparities arise largely from varying estimates of the annual carbon sequestration rate (per area) and the extent of land deemed suitable for forestation (*7*, *8*). Carbon sequestration potential is often estimated by changes in biomass carbon (aboveground and belowground biomass) and soil stocks over a certain period (e.g., 30 years). Although carbon sequestration in biomass has been relatively well assessed through a synthesis of worldwide field measurements (*13*), soil carbon change as a result of forestation is relatively less represented (*14*, *15*). Additionally, estimates of global area available for forestation show a 10-fold difference, ranging from about 200 Mha to >2000 Mha (*6*, *10*), owing to differences in study objectives and methodologies. Many assessments of land availability inadequately account for critical constraints such as albedo effects, water scarcity, biodiversity trade-offs, and policy limitations (*6*, *10*, *11*). Notably, the spatial mismatch between biophysically suitable land ("supply") and policy-driven forestation commitments ("demand") adds further complexity (*16*–*18*). To unlock forestation's potential as a scalable natural climate solution, two priorities are paramount: advancing integrated assessments of biomass and soil carbon dynamics, and conducting holistic assessments of land availability that reconcile biophysical, ecological, and socioeconomic factors. Together, these steps will provide the foundation for strategically framing future forestation efforts and evaluating climate mitigation impacts (*7*, *18*).

In this study, we developed a machine learning model to quantify soil carbon change after forestation. By integrating these results with biomass sequestration estimates (*13*), we mapped global ecosystem carbon sequestration rates (soil and biomass) resulting from potential forestation. Furthermore, we assessed land availability through the lens of national policy commitments, explicitly addressing the "supply-demand" disconnect, and reevaluated forestation's climate mitigation potential. Our findings provide actionable insights for optimizing land-use policies and forestation strategies to maximize climate benefits.

Mapping global carbon sequestration rates

We mapped global carbon sequestration rates after forestation at 0.01° resolution (Fig. 1). Ecosystem sequestration rates (Fig. 1A) are expressed as annual net carbon changes in both biomass and soil over the first 30 years of forestation beginning in 2021 (2021–2050) to align with a 2050 carbon neutrality target. Soil organic carbon (SOC) changes (Fig. 1B) are explicitly estimated to complement a prior study that focused only on biomass (*13*). Using an approach similar to that of Cook-Patton *et al.* (*13*), we developed a data-driven machine learning model (mSOC) to map SOC change after forestation (see materials and methods sections M1 and M2 in the supplementary materials). The model was trained using 1595 paired soil layer–specific observations from sites under afforestation (87%) and reforestation (13%) across various forest types (fig. S1 and table S1). The afforestation and reforestation submodels differ in influencing factors (fig. S2), but both demonstrate reasonable performance with cross-validation coefficients of determination (R^2) of 0.88 for afforestation and 0.85 for reforestation (fig. S3).

Across the areas where new forests could potentially be established, forestation leads to SOC changes (Fig. 1B), with >80% of the changes occurring in the top 20 cm of soil (fig. S4A). SOC sequestration rates are generally lower in tropical and some boreal regions. Local SOC losses are predicted for reforestation on carbon-rich soils in tropical areas (-0.46 ± 0.53 Mg C ha^{-1} $year^{-1}$ in Southeast Asia and -0.23 ± 0.31 Mg C ha^{-1} $year^{-1}$ in the Amazon) and in boreal areas (-0.49 ± 0.63 Mg C ha^{-1} $year^{-1}$ in northern Europe). It takes time for SOC to recover to predisturbance levels after reforestation (*15*, *19*). Factors conducive to SOC loss include coarser soil textures (*19*, *20*) and warmer and wetter climates that accelerate SOC decomposition (*21*) (fig. S5). Afforestation is more likely to increase SOC in places where initial SOC is low (e.g., cropland, bare land), including temperate Asia (0.27 ± 0.35 Mg C ha^{-1} $year^{-1}$) and savannas in subtropical Africa (0.18 ± 0.37 Mg C ha^{-1} $year^{-1}$) (Fig. 1B and fig. S4C). This aligns with field observations that recognize the potential for increasing SOC sequestration in carbon-poor soils (*22*). Even though SOC gain is predicted in areas originally covered with savanna, afforestation remains ecologically unsuitable owing to biodiversity risks (*16*, *23*). Note that the SOC prediction errors are generally small and are minimized in data-rich low latitudes (Fig. 1C).

[1]School of Atmospheric Sciences, Key Laboratory of Tropical Atmosphere-Ocean System (Ministry of Education), Guangdong Province Key Laboratory for Climate Change and Natural Disaster Studies, Sun Yat-sen University, Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai, China. [2]The Nature Conservancy, Arlington, VA, USA. [3]Smithsonian Environmental Research Center, Edgewater, MD, USA. [4]Institute of Botany, Chinese Academy of Sciences, Beijing, China. [5]Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China. [6]Laboratoire des Sciences du Climat et de l'Environnement, CEA CNRS UVSQ, Gif-sur-Yvette, France. [7]Institute of Biological and Environmental Sciences, University of Aberdeen, Aberdeen, UK. [8]Institute of Carbon Neutrality, College of Urban and Environmental Sciences, Peking University, Beijing, China. [9]College of the Environment and Ecology, Xiamen University, Xiamen, China. [10]Global Carbon Project, CSIRO Environment, Canberra, ACT, Australia. [11]Institute of Soil and Water Conservation, Northwest A&F University, Yangling, China. ***Corresponding author. Email: qinzhangcai@mail.sysu.edu.cn** †Present address: School of the Environment and Sustainable Engineering, Eastern Institute of Technology, Ningbo, China. ‡Present address: School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, UK.

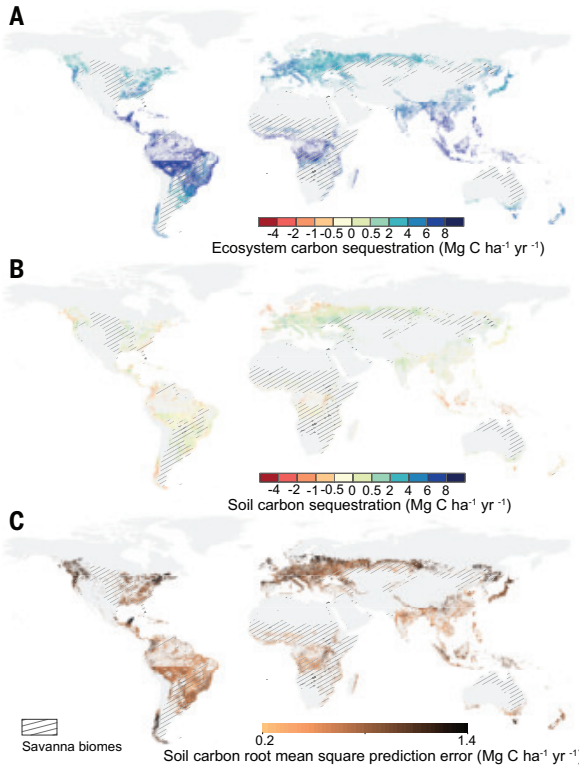


Fig. 1. Estimated carbon sequestration rates with forestation across potential lands. (**A**) Ecosystem carbon sequestration rate (soil and biomass combined). (**B**) Soil carbon sequestration rate (0 to 60 cm). (**C**) Uncertainty of soil carbon sequestration rate (0 to 60 cm). The positive values [(A) and (B)] show carbon gain, and negative values show loss. The color-shaded areas show potential lands estimated in this study (regardless of savanna biome boundaries); the areas under savanna are further constrained. Data are available in data S1 (*24*).

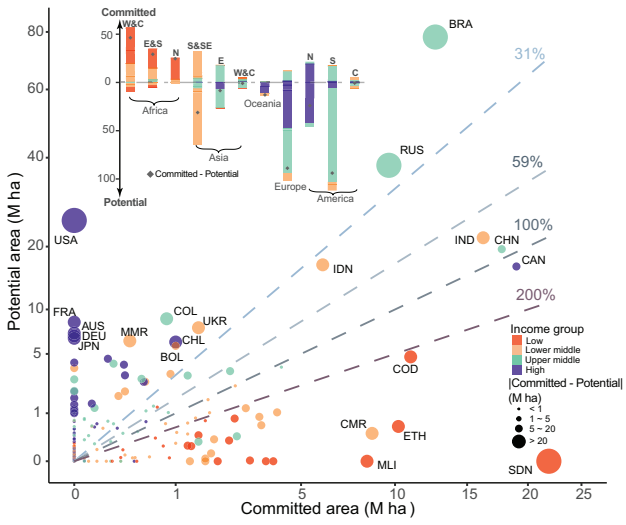


Fig. 2. Potential and committed land for forestation by country and region. The dashed lines indicate the percentage of *x* axis in *y* axis. The 31% and 59% lines refer to the global average proportion of currently achievable (with both potential and commitment) and committed area over total potential area estimated in Qin24, respectively. The inset figure illustrates land availability by region. Country codes are shown for those with forestation potential or commitment >5 Mha. Region names, country codes, and income groups can be found in table S6. Specific country-level data are available in data S2 (*24*).

Ecosystem carbon sequestration varies spatially (Fig. 1A), generally with a larger contribution from biomass than soil (fig. S4B). Globally, the 30-year average potential carbon sequestration rate from both biomass and soil (including savanna) is 3.15 (±1.76) Mg C ha^{-1} $year^{-1}$, with 2.86 (±1.49) and 3.35 (±1.89) Mg C ha^{-1} $year^{-1}$ for afforestation and reforestation, respectively. Within the tropics (between 23°N and 23°S), the total potential sequestration ranges from 0.14 to 8.53 Mg C ha^{-1} $year^{-1}$, with an average of 5.28 (±1.20) Mg C ha^{-1} $year^{-1}$. The highest rate is associated with wet tropical forest ecosystems, with 5.66 (±1.11) Mg C ha^{-1} $year^{-1}$, 5.80 (±1.32) Mg C ha^{-1} $year^{-1}$, and 4.66 (±1.13) Mg C ha^{-1} $year^{-1}$ in South America, Africa, and Southeast Asia, respectively, reflecting high biomass growth rates (fig. S4, C and D). Boreal forests show minimal potential in either biomass or soil carbon sequestration. The spatial variability of ecosystem sequestration is primarily driven by the variability in biomass (fig. S4B). The actual ecosystem carbon sequestration should be further evaluated with dedicated extent of land.

Suitable area for potential forestation

Estimates of overall climate mitigation from forestation heavily rely on the considered available land area. From a land "supply" point of view, Bastin *et al.* (2019) (BA19) found that about 900 Mha were biophysically suitable for increasing tree cover, even beyond current forest biomes (*10*) (fig. S6A). Excluding areas with risks of biodiversity loss in natural grasslands and albedo-induced warming in the boreal region, Griscom *et al.* (2017) (GR17) estimated the potential to be 678 Mha (*8*) (fig. S6B). In this study, we developed a new dataset, referred to as "Qin24" (*24*), and identified that only ~390 Mha of "potential area" could be supplied for forestation (fig. S6C) after excluding areas with potential biodiversity loss (*16*, *23*, *25*), biophysical warming (*26*), and exacerbated water stress (*27*) (see materials and methods section M3). Most of the Qin24 potential area is concentrated in the Americas (42%) and Europe (26%), with Brazil, Russia, and the USA contributing to 36% of the total (Fig. 2). Brazil alone represents one-fifth of the potentially suitable area (78 Mha), primarily consisting of recently deforested regions of the Amazon (*28*). Low- and middle-income countries (LMICs) hold about 71% of this potential (table S2).

From the land "demand" perspective, global commitments across all countries have pledged 229.7 Mha of forestation during 2021–2030, 59% of the global potential area (fig. S6D). Notably, 90% of the committed area is contributed by LMICs with limited potential area (Fig. 2). Africa, with only 4% of the global potential area (16.0 Mha), has committed 50% of forestation pledges (115.5 Mha) (table S2); this limited potential area is mainly due to biodiversity concerns (*16*, *23*), suggesting that commitments may exceed its potential area by more than six times. The commitments in 38 out of 41 African countries were more than double the estimated potential area (Fig. 2, 200% line). Sudan, the Democratic Republic of the Congo, and Ethiopia account for 43 Mha of overcommitments (Fig. 2), likely relying on lands that we excluded in this study, such as savanna. Commitments ("demand") exceed the estimated potential area ("supply") in 57 out of 105 countries across the globe, including 54 LMICs. This leaves only 119.7 Mha, which is 31% of "supply" or 52% of "demand," to be currently "achievable" from the intersection of committed and suitable land availability, with 84% in LMICs and mostly in Asia (50.5 Mha).

Europe and South America have only committed 12.8 and 16.1% of their potential area, respectively, far less than the global average of 59% (Fig. 2). The top 15 countries, representing 70% of the global potential area, have only committed 31% their estimated potential to forestation (84 Mha) (fig. S6E). Notably, Canada, China, and India have contributed to 62% of the top 15 countries' commitments. By contrast, the last 10 of these 15 countries with high potential area have either made no or extremely limited commitments, totaling 8 Mha.

Global climate change mitigation

Under a hypothetical scenario of worldwide forestation over a period of 10 years starting from 2021, the carbon sequestration in biomass and

soil could reach a theoretical amount of 39.9 Pg C by 2050 on 389 Mha of potential lands estimated in this study (excluding savanna). This would equate to about 1.3 Pg C year^{-1}, which is 63% of the current net annual land sink from 2014 to 2023 (i.e., 2.1 Pg C year^{-1}) (*12*). This amount is much lower than the 91.1 and 63.4 Pg C over 30 years, on 891 and 678 Mha of land, predicted by Bastin *et al.* (*10*) and Griscom *et al.* (*8*), respectively (Fig. 3A). For all these three estimates from the land "supply" perspective, carbon sequestration is dominated by South America, where a large area of previously cleared forests could potentially be regrown (e.g., Brazil, Colombia), followed by Asia, particularly China and India, where potential area is available.

Taking into account only current commitments (materials and methods section M3), the "committed" mitigation potential significantly declines, to 21.4 Pg C by 2050, at a rate of 0.71 Pg C year^{-1} (Fig. 3A). When further considering these commitments along with our land availability limitations, the resulting currently "achievable" mitigation potential drops to 12.5 Pg C, with an average rate of 0.42 Pg C year^{-1}. For comparison, the carbon uptake by land over the period 2014–2023 is 3.2 Pg C year^{-1} (*12*), and halting global forest loss by 2030 could avoid emissions of between 1.1 and 1.3 Pg C year^{-1} (2021–2050) (*17*, *18*). Land-use change causes about 1.1 Pg C year^{-1} carbon emissions globally (*12*) and −0.14 Pg C year^{-1} carbon sink in China (*29*). It should be noted that 95% of the achievable mitigation is contributed by LMICs, mainly Brazil, India, and China. High-income countries only contribute a cumulative 0.9 Pg C, with Canada being the largest contributor (74%) owing to its highly ambitious commitments. Asia alone contributes the most mitigation (5 Pg C, or 40%), with India and China accounting for more than half, followed by South America and Africa (Fig. 3B).

Forestation action: faster, higher, stronger—together

The effectiveness of forestation as a natural climate solution ultimately depends on taking action: acting faster, setting higher targets, prioritizing areas with greater benefits, and coordinating global efforts for the right places with suitable species (*8*, *30*). Actual climate mitigation is heavily influenced by the time taken to reach full potential considering ecological and social delays (*30*). Most current commitments aim to accomplish forestation by 2030, which leaves about 5 years for action starting from 2025. If forestation action is delayed, the potential 21.4 Pg C of sequestration by 2050 would be reduced by 0.41 (0.37–0.45) Pg C for each year of delay (fig. S7). Acting faster could lower the risk of reduced efficacy of human actions (*31*), decreasing the likelihood of severe climate damages (*9*).

Greater mitigation could also be achieved with more ambitious forestation goals. Currently, about 59% of the potential area is designated for global forestation, but only half of that area (120 Mha) aligns with our estimate of land availability. Many high-income countries have not made explicit official commitments to the major forestation projects surveyed here (Fig. 2). To meet a higher target of forestation, like the current global commitment of 59% of the potential area, even more ambitious commitments and actions are urgently needed, especially for countries that have made limited commitments despite their high potential as estimated in Qin24 (Fig. 2). In addition, even with the same area of committed land, prioritizing locations with a stronger capability to sequester carbon can maximize the overall mitigation magnitude. Across the 229.7 Mha of committed global area, total carbon sequestration could be 20% higher if prioritizing locations with stronger rather than lower sequestration rates (fig. S7). Tropical regions, Brazil for example, could boost their mitigation, given their higher carbon sequestration rate compared with other regions, if the potential area permits (Fig. 1A). In contrast, for regions with high potential areas but smaller carbon sequestration rates, such as the USA and Russia, more ambitious commitments are needed to result in equivalent amounts of mitigation (Fig. 4).

Lastly, the asymmetry in mitigation potential and overall socioeconomic development among countries and regions necessitates global and collective action (Fig. 4). Globally, 83.9% of the 39.9 Pg C mitigation potential (on 389 Mha of land) is concentrated in LMICs, which represent only 37.4% of the global gross domestic product. Despite considerable mitigation potential, many LMICs struggle with limited resources for deploying large-scale forestation projects. Countries such as Indonesia and Brazil have potential for forestation but are still losing large amounts of forest land to agricultural and commodity production (e.g., palm oil, wood fiber) (*3*, *32*). The essential involvement of tropical regions, especially LMICs, requires as much local and national commitment as international involvement in financing projects (*33*, *34*). Maximizing effective climate mitigation requires global cooperation, including financing, technology transfer, carbon markets, and mechanisms such as the United Nations Green Climate Fund (*2*, *34*, *35*). Equitable support must address power asymmetries in different regions (e.g., North–South) and actors, such as the state versus local communities and Indigenous peoples and rights (*2*, *36*, *37*). Global and collective action is needed to ensure projects benefit regional and Indigenous communities, including some LMICs that are least responsible for climate change but with high forestation potential.

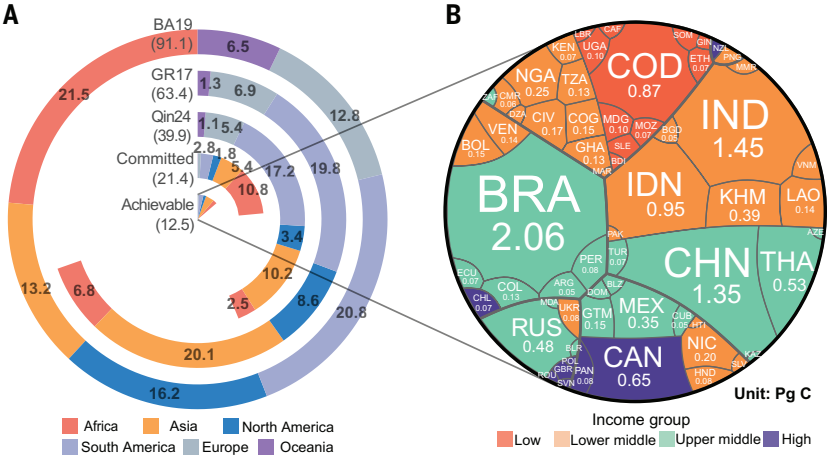


Fig. 3. Estimated mitigation potential from forestation varies with land availability assumptions. (**A**) Global and regional ecosystem carbon sequestration. BA19 (*10*), GR17 (*8*), and Qin24 (*24*) refer to respective assumptions of land potentially available for forestation. The "committed" assumption considers only national commitments, and "achievable" shows currently achievable land limited by availability defined in Qin24 (see materials and methods section M3). (**B**) The currently achievable mitigation potential varies among countries and income groups. Country-specific data are available in data S3 (*24*). Country codes and corresponding income groups can be found in table S6.

Uncertainties and pathways

Global forestation's mitigation potential hinges on carbon sequestration rates and land availability, in addition to finances, governance, and sociopolitical considerations (*8*, *30*, *38*). SOC modeling needs more field data for validation and model development, particularly for longer-term (>30 years), deeper soils (>60 cm) and regions with lack of data such as the Amazon and Congo Basin (*14*, *15*). Land availability assessments require further examination and consolidation to better reflect consistently defined, spatially explicit, biome- and region-specific forestation opportunities (*39*, *40*). Sociopolitical challenges, such as contested land tenure (*41*), power asymmetries, unfairness, and injustice in forest governance (*37*), should be further examined to align local forestation opportunities with global studies (*11*, *42*, *43*). Future mitigation estimates should consider mechanisms and factors missing in current carbon sequestration modeling (*15*, *44*–*46*). Our estimates (see materials and methods section M4) suggest that CO_2 fertilization could boost biomass carbon sequestration and increase estimated mitigation by 11 to 12%. Albedo

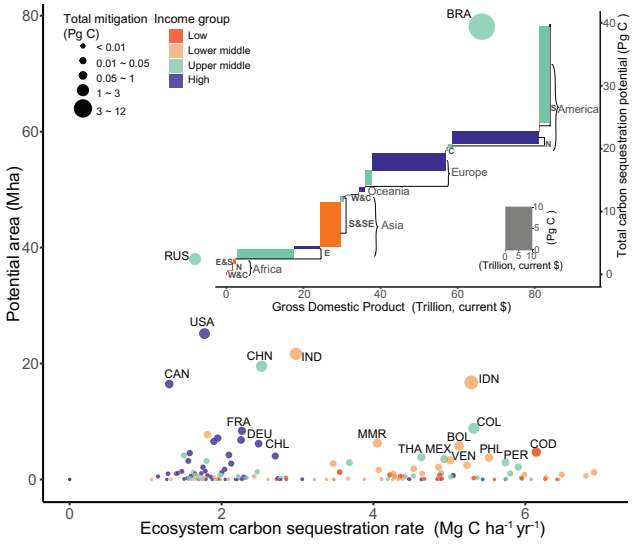


Fig. 4. The ecosystem carbon sequestration rate, potential area, and mitigation potential vary spatially. The countries with large potential areas may not necessarily exhibit high carbon sequestration rates. The inset graph illustrates carbon sequestration potential among regions with varying economic situations. The carbon sequestration is based on potential area estimated in this study (*24*). Country codes are shown for those with mitigation potential exceeding 0.5 Pg C. The ecosystem carbon sequestration rate data are available in data S4 (*24*). Region names, country codes, and income groups can be found in table S6.

change could lead to an overall warming effect, offsetting 1.3 Pg C of achievable mitigation. Studies show that the net impacts of CO_2 fertilization and climate change on biomass productivity vary from a 30% loss to a 60% gain of regional carbon stocks (*47*). Negative impacts such as ozone damage, drought, and nitrogen and phosphorus limitation may often negate CO_2 fertilization effects (*2*, *48*–*50*). Numerous natural (e.g., wildfires, insect herbivores, and fungal pathogens) and anthropogenic disturbances (e.g., catastrophic failure of planted stands) as well as forest harvest or wood extraction may further reduce the potential achievable mitigation (*51*–*53*).

Furthermore, social and economic development, conservation priorities, and the needs and interests of local communities should also be factored into the accounting of regional land supply and productivity, to determine the "practical," not just "theoretical," mitigation potential by country and region (*54*). In the forest sector, halting forest loss and protecting and managing existing forests are just as important as, if not more important than, creating new forests (*7*, *8*, *17*, *18*). From a life-cycle perspective, long-term carbon dynamics (>30 years) involving plantation, wood harvest, and biomass utilization should be explored, especially for net-negative emissions needs beyond ecosystem boundaries (e.g., bioenergy) (*55*–*57*). Natural climate solutions could bridge the gap between national decarbonization pledges (nationally determined contributions, or NDCs) and the efforts required to limit warming within 1.5° or 2°C (*7*, *9*), but rapid decarbonization in industry and other fossil-related sectors remains the priority (*9*, *12*, *58*).

REFERENCES AND NOTES

1. J. Busch *et al.*, *Nat. Clim. Chang.* **9**, 463–466 (2019).
2. J. G. Canadell, M. R. Raupach, *Science* **320**, 1456–1457 (2008).
3. P. G. Curtis, C. M. Slay, N. L. Harris, A. Tyukavina, M. C. Hansen, *Science* **361**, 1108–1111 (2018).
4. Z. Qin *et al.*, *One Earth* **7**, 835–847 (2024).
5. G. M. Domke, S. N. Oswalt, B. F. Walters, R. S. Morin, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 24649–24651 (2020).
6. W. S. Walker *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2111312119 (2022).
7. S. C. Cook-Patton *et al.*, *Nat. Clim. Chang.* **11**, 1027–1034 (2021).
8. B. W. Griscom *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11645–11650 (2017).
9. S. Deng *et al.*, *Glob. Change Biol.* **29**, 289–291 (2023).
10. J.-F. Bastin *et al.*, *Science* **365**, 76–79 (2019).
11. S. Rohatyn, D. Yakir, E. Rotenberg, Y. Carmel, *Science* **377**, 1436–1439 (2022).
12. P. Friedlingstein *et al.*, *Earth Syst. Sci. Data* **17**, 965–1039 (2025).
13. S. C. Cook-Patton *et al.*, *Nature* **585**, 545–550 (2020).
14. L. E. Nave *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 2776–2781 (2018).
15. E. Veldkamp, M. Schmidt, J. S. Powers, M. D. Corre, *Nat. Rev. Earth Environ.* **1**, 590–605 (2020).
16. C. L. Parr, M. Te Beest, N. Stevens, *Science* **383**, 698–701 (2024).
17. Y. Zhu *et al.*, *Innovation (Camb.)* **3**, 100307 (2022).
18. T. Gasser, P. Ciais, S. L. Lewis, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2200519119 (2022).
19. J. Laganière, D. A. Angers, D. Paré, *Glob. Change Biol.* **16**, 439–453 (2010).
20. W. M. Post, K. C. Kwon, *Glob. Change Biol.* **6**, 317–327 (2000).
21. R. Lal, *For. Ecol. Manage.* **220**, 242–258 (2005).
22. S. Hong *et al.*, *Nat. Sustain.* **3**, 694–700 (2020).
23. R. C. R. Abreu *et al.*, *Sci. Adv.* **3**, e1701284 (2017).
24. Z. Qin, Forestation: land availability and climate change mitigation, Figshare (2025); https://doi.org/10.6084/m9.figshare.28377266.
25. G. Hayman, *Science* **383**, 833–834 (2024).
26. N. Hasler *et al.*, *Nat. Commun.* **15**, 2275 (2024).
27. A. J. Hoek Van Dijke *et al.*, *Nat. Geosci.* **15**, 363–368 (2022).
28. P. Potapov *et al.*, *Front. Remote Sens.* **3**, 856903 (2022).
29. Y. Zhu *et al.*, *Nat. Clim. Chang.* **15**, 428–435 (2025).
30. Z. Qin *et al.*, *Glob. Change Biol.* **27**, 215–217 (2021).
31. J. P. Dunne, R. J. Stouffer, J. G. John, *Nat. Clim. Chang.* **3**, 563–566 (2013).
32. M. C. Hansen *et al.*, *Science* **342**, 850–853 (2013).
33. J. T. Erbaugh *et al.*, *Nat. Ecol. Evol.* **4**, 1472–1476 (2020).
34. S. Roe *et al.*, *Nat. Clim. Chang.* **9**, 817–828 (2019).
35. Y. Zeng *et al.*, *Nat. Clim. Chang.* **10**, 842–844 (2020).
36. C. M. Anderson *et al.*, *Science* **363**, 933–934 (2019).
37. D. Kleinschmit, C. Wildburger, N. Grima, B. Fisher, *International Forest Governance: A Critical Review of Trends, Drawbacks, and New Approaches*, vol. 43 of *IUFRO World Series* (International Union of Forest Research Organizations, 2024).
38. L. Mo *et al.*, *Nature* **624**, 92–101 (2023).
39. C. R. Drever *et al.*, *One Earth* **8**, 101177 (2025).
40. L. Yao *et al.*, *Nat. Commun.* **15**, 8398 (2024).
41. O. S. Rakotonarivo *et al.*, *Commun. Earth Environ.* **4**, 179 (2023).
42. C. A. Williams, H. Gu, T. Jiao, *Sci. Adv.* **7**, eaax8859 (2021).
43. M. G. Windisch, E. L. Davin, S. I. Seneviratne, *Nat. Clim. Chang.* **11**, 867–871 (2021).
44. R. B. Jackson *et al.*, *Annu. Rev. Ecol. Evol. Syst.* **48**, 419–445 (2017).
45. F. Tao *et al.*, *Nature* **618**, 981–985 (2023).
46. C. Terrer *et al.*, *Nature* **591**, 599–603 (2021).
47. S. L. Lewis, C. E. Wheeler, E. T. A. Mitchard, A. Koch, *Nature* **568**, 25–28 (2019).
48. L. Xia *et al.*, *One Earth* **4**, 1752–1763 (2021).
49. W. R. Wieder, C. C. Cleveland, W. K. Smith, K. Todd-Brown, *Nat. Geosci.* **8**, 441–444 (2015).
50. W. R. L. Anderegg *et al.*, *Science* **368**, eaaz7005 (2020).
51. L. Peng, T. D. Searchinger, J. Zionts, R. Waite, *Nature* **620**, 110–115 (2023).
52. K.-H. Erb *et al.*, *Nature* **553**, 73–76 (2018).
53. C. T. J. Roebroek, G. Duveiller, S. I. Seneviratne, E. L. Davin, A. Cescatti, *Science* **380**, 749–753 (2023).
54. P. Smith *et al.*, *Nat. Clim. Chang.* **6**, 42–50 (2016).
55. E. J. Forster, J. R. Healey, C. Dymond, D. Styles, *Nat. Commun.* **12**, 3831 (2021).
56. Y. Xu, P. Smith, Z. Qin, *iScience* **27**, 110232 (2024).
57. S. Liang *et al.*, *Sci. Adv.* **11**, eadn7915 (2025).
58. W. Steffen *et al.*, *Science* **280**, 1393–1394 (1998).

ACKNOWLEDGMENTS

We thank B. Griscom for his active contributions to data resources and many valuable discussions during the early stages of our research, and the members of the Qin lab for their many helpful discussions throughout this project. **Funding:** This work was supported by the National Natural Science Foundation of China (U21A6001, 42141020) fund outdated. Y.W. and Y.Z. were partially supported by the Science and Technology Program of Guangdong (2024B1212070012). P.C. acknowledges support from the Carbon Loss In Plants, Soils and Oceans (CALPSO) funded by the Schmidt Sciences. C.Y. was funded by the National Key Research and Development Program of China (2023YFB3907403) and the Second Tibetan Plateau Scientific Expedition and Research Program (2022QZKK0101). **Author contributions:** Conceptualization: Z.Q. Methodology: Z.Q., Y.W. Investigation: Y.W., Z.Q., W.S., W.Z. Resources: Z.Q., S.C.C.-P., P.C., T.L., P.S., W.Y., X.Z., J.G.C. Data curation: Y.W., Y.Z., X.D., Y.X., H.X., C.Y. Formal analysis: Y.W., Z.Q. Visualization: Y.W., Z.Q., Y.Z. Supervision: Z.Q. Project administration: Z.Q. Writing – original draft: Z.Q. Writing – review & editing: All authors. **Competing interests:** S.C.C.-P. serves on the technical advisory board for the Symbiosis Coalition and the science advisory group for Restor and as a science adviser for CTrees. All other authors declare that they have no competing interests. **Data and materials availability:** All data and code needed to replicate and extend the analysis are available on Figshare (*24*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adj6841
Materials and Methods; Figs. S1 to S7; Tables S1 to S8; References (*59*–*342*)

Submitted 24 August 2023; resubmitted 3 December 2024; accepted 11 July 2025
10.1126/science.adj6841

ROBOTICS

Launching by cavitation

Dalei Wang[1]†, Zixiao Liu[2]†, Hongping Zhao[1]†, Huanqi Qin[1]†, Gongxun Bai[3], Chi Chen[2], Pengju Shi[2], Yingjie Du[2], Yusen Zhao[2], Wei Liu[1]*, Dan Wang[4]*, Guoquan Zhou[1]*, Ximin He[2]*, Chaoqing Dai[1]*

Cavitation, characterized by formation of vapor bubbles in a low-pressure or high-temperature region of a liquid, is often destructive, but it can be harnessed for actuators and robots. We exploit cavitation to accumulate substantial energy in superheated liquids by suppressing its immediate release until reaching a stability limit. The energetic, unstable bubbles collapse violently, producing a burst of high power and force that initiates motion. Notably, a millimeter-scale device launched by cavitation can jump to a height of 1.5 meters—reaching a 12 meters per second (m/s) peak velocity, a 7.14×10^4 m/s^2 acceleration, and a 0.64% energy efficiency—and can also swim on water at 12 centimeters per second. Cavitation-based launching works with a broad range of device materials, liquid media, stimuli, and operational environments.

Launching propels a system from a state of rest or low speed into rapid motion by storing substantial energy and swiftly releasing it, generating a potent burst of power and force that initiates motions, such as jumping, sprinting, swimming, and flying. Inspired by biological muscles' elastic energy ($U = \int_V 0.5k\varepsilon^2 dV$) is widely harnessed for artificial launching. Strategies to improve launch performance typically focus on increasing elastic energy storage by optimizing material stiffness k and deformability ε_{max}, or by enhancing the energy release rate dU/dt through mechanical instabilities. For example, snap-through buckling enables considerable energy storage by creating a high energy barrier through tailored mechanical properties and geometric designs (*1*), while its intrinsic instability promotes rapid energy release (*2*, *3*). Additionally, antagonistic structures with self-latching mechanisms, as observed in insect flight, can effectively amplify power output. In these systems, the contraction of an agonist muscle pre-stretches its antagonist, and the sudden relaxation of these muscles explosively releases the stored energy. This mechanism is exemplified by an engineered jumper that combines the compression of a structural frame (agonist; energy density $\Gamma = 1922$ J/kg) with the tension of multiple elastic linkages (antagonist; $\Gamma = 7000$ J/kg), enabling a 0.3-m device to achieve a jump height of ~32 m (*4*). However, solid structures generally exhibit limited energy storage capacity and slow deformation rates, inherently constraining the performance of elastic energy–based launchers.

To overcome these limitations, phase-transition processes capable of delivering higher energy and faster dynamics have increasingly been used. For instance, by loading high-energy propellants and explosives with high Γ (e.g., butane, $\Gamma = 49.5$ MJ/kg) into jumpers and harnessing their rapid, substantial volumetric expansion during combustion (*5*–*7*), insect-scale jumpers can reach a height of 59 cm (*8*). Similarly, inducing water boiling (Γ = 2260 kJ/kg) in hydrogels via photothermal stimulation (*9*, *10*) can trigger bubble expansion that exceeds the hydrogel's fracture strain (*10*). This sudden structural failure rapidly releases energy and amplifies launch power. However, devices based on the fracture-driven mechanism are single-use, and their performance remains suboptimal: The less intense dynamics of boiling restrict both the amount and rate of energy release.

Cavitation is a highly dynamic and energetic liquid-vapor transition process in which vapor bubbles nucleate, explosively expand, and violently collapse in response to local pressure drops or elevated temperatures. Unlike boiling, which releases less energy over a longer duration (several to tens of milliseconds), cavitation can accumulate greater energy in a superheated state, and its highly energetic, unstable nature enables an ultrafast energy release on a microsecond timescale (*11*), producing ultrahigh-pressure output. Cavitation is typically regarded as destructive in industrial contexts, where it can mechanically erode hard alloys within a microsecond by generating pressures of several hundred megapascals (*12*). However, if its immense output and ultrafast dynamics can be effectively harnessed, cavitation may potentially enhance the performance of actuators, robots, and other engineered devices. Nature has evolved ingenious ways to exploit cavitation: fern sporangia utilize cavitation within dehydrated annulus cells to trigger rapid bending, launching spores at high velocities of ~10 m/s (*13*, *14*). Similarly, mantis shrimps swiftly snap their appendages to induce hydrodynamic cavitation, delivering mechanical impact–cavitation shockwave dual strikes with large forces of ~1500 N on their prey (*15*).

We use cavitation as a launching mechanism for rapid motion. The rapid transmission of substantial energy during cavitation produces a burst of power and acceleration, imparting high kinetic energy to the launched objects. This mechanism can be triggered by various stimuli to launch devices made from different materials across a wide range of operational environments.

Jumpers launched by photothermal cavitation

Although cavitation can be induced by various stimuli, including lasers, electrical sparks, and ultrasound, we found that photothermal cavitation delivers the best motion performance while enabling remote activation. Therefore, we selected photothermal cavitation–launched jumpers to analyze the launching mechanism. Various photothermal materials can be used to construct the jumpers, but they generally need to meet three key requirements: (i) high photothermal efficiency for rapid heat generation, (ii) high mechanical robustness and thermal stability to withstand extreme pressure and temperature conditions during cavitation, and (iii) adjustable density for deployment in diverse operational environments. A titanium dioxide–polypyrrole–titanium carbide (TiO_2-PPy-TiC) composite was used [fig. S4A and supplementary materials (SM) sections M1 and M2]. TiC exhibits high near-infrared (NIR) absorbance (fig. S4K), enhancing the system's photothermal efficiency to 84.47% (SM section S1 and fig. S4, I and J) (*16*, *17*)], which enables rapid heating of the surrounding liquid to the superheated state and triggers cavitation. TiO_2 ensures thermal stability above 300°C (figs. S4L and S5), while the branched, interconnected PPy network binds TiC and TiO_2 (fig. S4, B to H), effectively resisting large impact forces. By adjusting TiO_2 concentration to tune the material density, the jumper can either float on or submerge in various liquids. The final dimensions of TiO_2-PPy-TiC-based jumpers are 1 mm by 1 mm by 0.2 mm, with a mass of 0.778 mg.

When deployed on a transparent solid surface moistened with a few water droplets and irradiated from below with an 808-nm near-infrared laser at 0.357 kW/cm^2, the millimeter-scale TiO_2-PPy-TiC jumper achieved a height of 1.5 m (Fig. 1A and movie S1), with a takeoff velocity >12 m/s (Fig. 1B), and an acceleration exceeding 7.14×10^4 m/s^2 (SM section S3). This performance stems from the high impact pressure (280 kPa; SM section S3) delivered by cavitating jets, which accelerate the jumper from rest to 12 m/s within just 0.168 ms [Fig. 1E, (III) to (V)]. The takeoff velocity and acceleration time surpass those observed in the cavitation-assisted spore dispersal of ferns (10 m/s; 1 s) (*14*).

[1]National Key Laboratory for Development and Utilization of Forest Food Resources, Zhejiang A&F University, Hangzhou, China. [2]Department of Materials Science and Engineering, University of California, Los Angeles, Los Angeles, CA, USA. [3]College of Optical and Electronic Technology, China Jiliang University, Hangzhou, China. [4]School of Chemistry and Chemical Engineering, Anhui University of Technology, Ma'anshan, Anhui, China. ***Corresponding author. Email: liuwei@zafu.edu.cn (W.L.); wangdan@ahut.edu.cn (D.W.); 19920006@zafu.edu.cn (G.Z.); ximinhe@ucla.edu (X.H.); dcq424@zafu.edu.cn (C.D.)** †These authors contributed equally to this work.

Launching mechanism

Upon laser irradiation, the local temperature beneath the jumper rapidly exceeds the liquid's boiling point T_b, driving the water into a metastable state. Photothermal energy deposition intensifies molecular motion, raising internal pressure P_{in}. When the temperature reaches the liquid's superheat limit T_t, a cavity nucleates and destabilizes, explosively expanding against the surrounding liquid [Fig. 1E, (I) to (III)] and storing potential energy. This volumetric expansion elevates the jumper [Fig. 1E, (II)] over a duration of 0.624 ms until the bubble attains its maximum radius [Fig. 1E, (III)]. During the bubble growth, its internal pressure P_{in} decreases until it falls below the outside pressure P_{out}, triggering the rapid collapse over the subsequent 0.144 ms [Fig. 1E, (III) and (IV)]. As the bubble contracts, surrounding liquid rushes into the space previously occupied by the bubble at speeds up to 57.1 m/s (fig. S6). The collision of these high-speed flows generates a powerful upward jet [Fig. 1E, (IV) and (V)] that strikes the jumper with a pressure of 280 kPa (SM section S3). The complete cavitation-induced launch process is documented in movie S2. Our TiO_2-PPy-TiC material demonstrates excellent mechanical durability and thermal stability under extreme pressure and thermal conditions during cavitation, enabling over 500 repeated launch cycles without structural disintegration (Fig. 1D). Our cavitation-based launching strategy is applicable to objects made from various materials. For instance, thin platelets of pure candle soot (CS), carbon dots (CD), graphite (GR), or black metals can also be launched with decent performance (Fig. 1C). The photothermal agent TiC can be replaced with alternatives x, such as CS, CD, or GR, to fabricate TiO_2-PPy-x photothermal composites.

We introduced an analytical model to quantitatively analyze the pressure burst, energy conversion, and bubble evolution during cavitation. To clearly observe bubble morphology and jet formation, we temporarily removed the jumper (Fig. 2A and movies S3 and S4). The observed cavitation timescales (Fig. 2B) differ slightly from those in Fig. 1E because the absence of fluid-structure interaction alters the bubble dynamics. The experimentally measured maximum bubble radius $R_{max} \approx 750$ μm was used as the initial condition for the model. The dynamics of the liquid and gas phases were governed by the Navier-Stokes and continuity equations, with the gas described by the Noble-Abel equation of states and the compressible liquid by the Tait equation. Bubble radius evolution $R(t)$ during the collapse was modeled using the Gilmore model (*18*),

$$\left(1-\frac{\dot{R}}{C}\right)R\ddot{R}+\frac{3}{2}\left(1-\frac{\dot{R}}{3C}\right)\dot{R}^2=\left(1+\frac{\dot{R}}{C}\right)H+\left(1-\frac{\dot{R}}{C}\right)\frac{R}{C}\dot{H} \quad (1)$$

where H and C denote the specific enthalpy difference and speed of sound, respectively. Detailed model derivations and variable expressions are provided in SM section S4 and (*18*). Figure 2C shows the simulated variations in bubble radius $R(t)$ and pressure during bubble contraction and collapse. The bubble contracts from R_{max} of 750 μm to a minimum radius R_{min} of 88.4 μm within 72 μs. During contraction, stored potential energy $E_p = \int (P_{out} - P_{in})dV$ is converted into kinetic energy $E_k = 0.5\rho V(dR/dt)^2$ (Fig. 2D). Both contraction rate dR/dt (fig. S6) and the kinetic energy E_k increase as $R(t)$ decreases, with $|dR/dt|$ peaking at 57.1 m/s just before the bubble reaches R_{min}. The subsequent bubble collapse generates a jet with a peak pressure of 10.2 MPa (Fig. 2C), which decays to an impact pressure of 280 kPa on the jumper, as determined experimentally (SM section S3). This decay is attributed to liquid drag and energy losses during momentum exchange at impact. During the bubble collapse, the kinetic energy E_k drops from its peak to zero almost instantaneously (within 8.4 μs), converting into a burst of potential energy E_p (Fig. 2D). Finite volume simulations using the volume of fluid method in OpenFOAM validated the bubble's morphological changes and jet formation [movie S4 and Fig. 2E, (I) to (IV)].

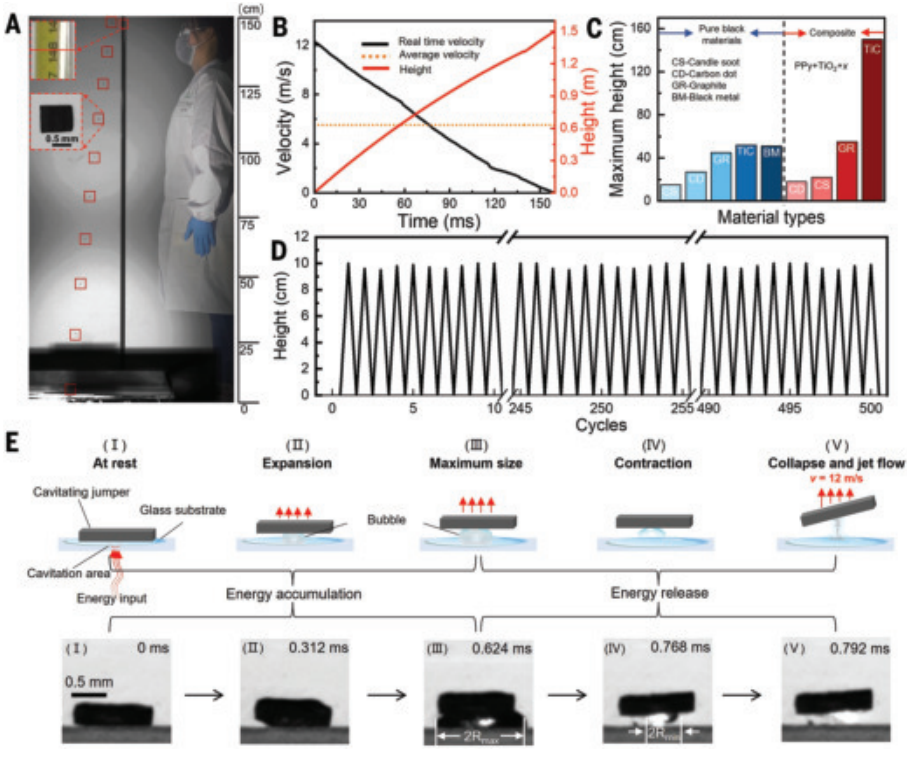


Fig. 1. Jumpers launched by photothermal cavitation. (**A**) Superimposed images depict a jump launched by photothermal cavitation reaching a peak height of 1.5 m (movie S1). Insets provide a magnified view of the ruler scale used for height measurement and a close-up of the TiO_2-PPy-TiC–based photothermal jumper. (**B**) Real-time measurements of velocity and displacement, along with average velocity of jumping. The takeoff velocity can reach 12 m/s. (**C**) Maximum jumping heights of cavitation-launched jumpers irradiated by a laser with the same intensity (0.357 kW/cm^2). Blue bars represent devices constructed from various pure black materials, whereas red bars indicate jumpers fabricated using TiO_2-PPy with different photothermal agents (x). (**D**) Durability test results for the TiO_2-PPy-TiC jumper demonstrate no structural disintegration over 500 launch cycles (experimental setup shown in fig. S1). (**E**) Schematic of the launching mechanism based on cavitation, accompanied by corresponding high-speed photography (movie S2).

Influence of liquid properties and environmental factors on jumping performance

We launched our TiO_2-PPy-TiC jumpers in various liquids (SM section M6) to study how liquid properties—such as compressibility, available nucleation sites, viscosity, and surface tension—affect launch performance (Fig. 2F). In more compressible liquids such as ethanol, jumpers perform less efficiently than in water primarily because water's lower compressibility resists volume change during bubble collapse, enabling greater energy accumulation and a more violent release. Additionally, preexisting nucleation sites influence bubble dynamics: In saline solutions, ion-dipole interactions between salt ions and water molecules form hydration shells, reducing free water concentration and air solubility, which may introduce noncondensable gas bubbles. These bubbles serve as nucleation sites for cavities, promoting premature cavitation (*19*). Besides, the noncondensable gas also acts as a cushion that softens bubble collapse (*20*, *21*),

which further diminishes jumping performance. We also increased liquid viscosity by dissolving polyvinyl alcohol (PVA) in water. The dynamic viscosities of 2.5 weight % (wt %) and 4.5 wt % PVA solutions were measured to be 0.05 and 0.08 Pa·s, respectively, which are much higher than that of water (~0.001 Pa·s). Jumpers in these PVA solutions exhibited reduced performance compared with those in pure water. Our numerical simulations indicated that cavitation pressure remains similar across viscosities, suggesting that the compromised performance primarily stems from increased liquid drag during upward jumping rather than cavitation pressure output. Lastly, jumpers were tested in 0.1 and 0.3 wt % soap water, which exhibited equilibrium surface tensions of 0.036 and 0.027 kg/s^2, respectively. The ultrafast nature of cavitation limits surfactant adsorption at the bubble-liquid interface, so dynamic surface tension needs to be considered (fig. S7). However, our analytical model shows that Laplace pressure from surface effects is several orders of magnitude lower than the bubble's internal pressure (fig. S8), which explains the negligible difference in jumping performance between water and soap solutions.

Liquid depth and laser intensity also influence launch and motion performance. Greater liquid depth requires the jumper to expend more energy overcoming liquid drag, thereby reducing its jumping height (fig. S9). Additionally, a minimum laser intensity I_0 (>0.087 kW/cm^2) is necessary to deposit enough energy to shift the system from a metastable to an unstable state and initiate cavitation (Fig. 2G). At I_0 = 0.357 kW/cm^2, water reaches its superheat limit (303.1°C; fig. S10), yielding optimal jumper performance. However, further increases in I_0 may lead to excessively intense bubble expansion, causing premature collapse before the bubble reaches its theoretical maximum size, which reduces the pressure

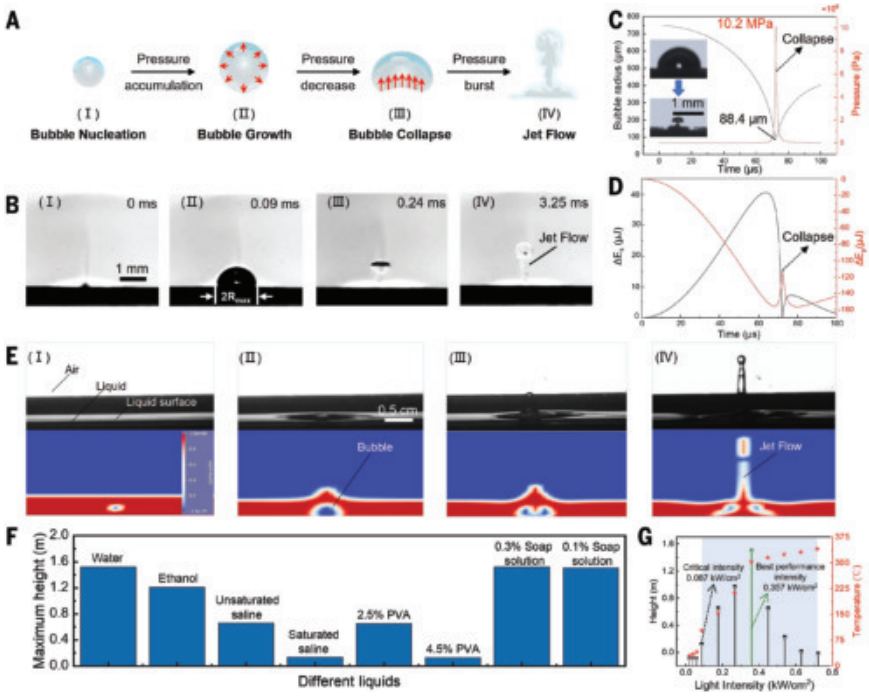


Fig. 2. Mechanism of cavitation-triggered launching and influence factors. (**A**) Schematic diagram and (**B**) high-speed photography (movie S3) capturing the cavitation process in detail, including bubble morphology evolution and jet flow generation. (**C**) Simulation results showing the evolution of bubble radius $R(t)$ and internal pressure $P_{in}(t)$ during bubble contraction and collapse. Insets display the experimental observations of bubble morphologies at maximum radius R_{max} and the jet formation upon collapse. (**D**) Simulated energy conversion between potential energy $E_p = \int (P_{out} - P_{in})dV$ and kinetic energy $E_k = 0.5\rho V(dR/dt)^2$ during bubble cavitation. (**E**) High-speed photography (movie S4), illustrating jet flow ejection from the water surface, alongside finite volume simulation results. (**F**) Jumping performance when launched from different liquids. (**G**) Jumping performance under varying light intensities.

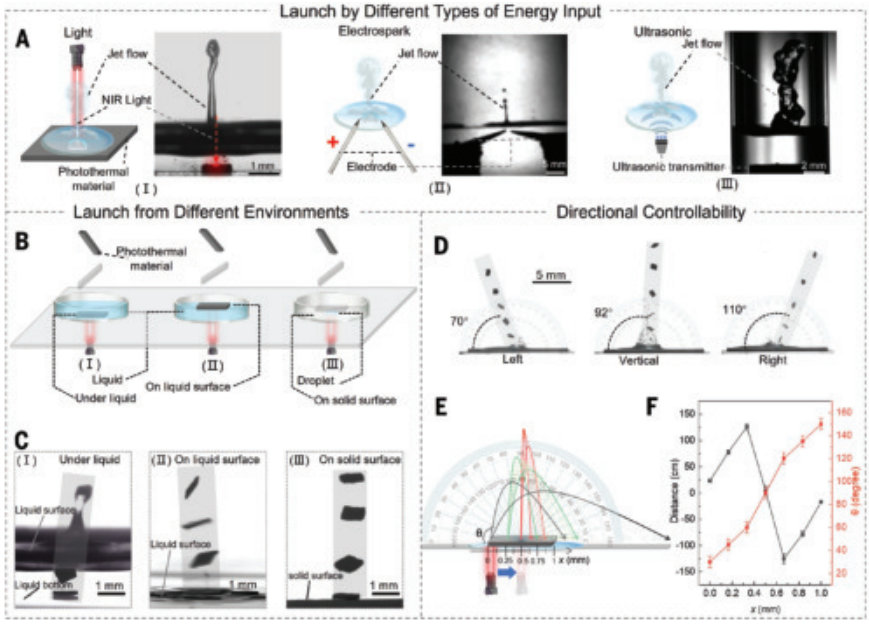


Fig. 3. Generality and controllability of cavitation-triggered launching. (**A**) Schematic and high-speed photography (movie S5) illustrate cavitation jets induced by different stimuli: (I) laser, (II) electrical spark, and (III) ultrasound. (**B**) Schematic diagram and (**C**) high-speed photography (movie S6) show the launches from diverse environments: (I) under liquid, (II) on liquid surfaces, and (III) on wet solid surfaces. (**D**) High-speed photography (movie S7) demonstrates controlled directional launches—leftward, vertical, and rightward—achieved by targeting laser irradiation on the right, center, and left portions of the jumper, respectively. (**E**) Schematic representation of the strategy for precisely controlling launch angle and horizontal launch distance. (**F**) Correlation between horizontal launch distance, launch angle, and laser irradiation position.

output and performance. We also tested jumpers of other shapes, including cones with different diameter-to-height ratios and hemispheres. Their uneven mass distribution increases airborne tumbling compared to flat platelets, dissipating more kinetic energy and reducing jumping height (fig. S11).

Generality and controllability

In addition to photothermal stimulation [Fig. 3A, (I)], cavitation can also be activated by electrical sparks and ultrasound (movie S5). Electrical sparks ionize water molecules, forming plasma that rapidly expands to generate spark-induced cavitation. This cavitation produces a jet flow [Fig. 3A, (II)], which can launch a photothermally nonresponsive polytetrafluoroethylene (PTFE) platelet to a height of 0.93 m (fig. S12F). Spark-induced cavitation does not rely on the material's photothermal properties or substrate transparency to allow light to pass through, making it a viable alternative to photothermal cavitation. Moreover, compact sparking electrodes can be easily embedded in robots as onboard power sources. Ultrasound transmits high-intensity compression and rarefaction waves into the liquid, reducing local pressure and forming acoustic cavitation bubbles. Notably, ultrasound creates a large low-pressure region that can

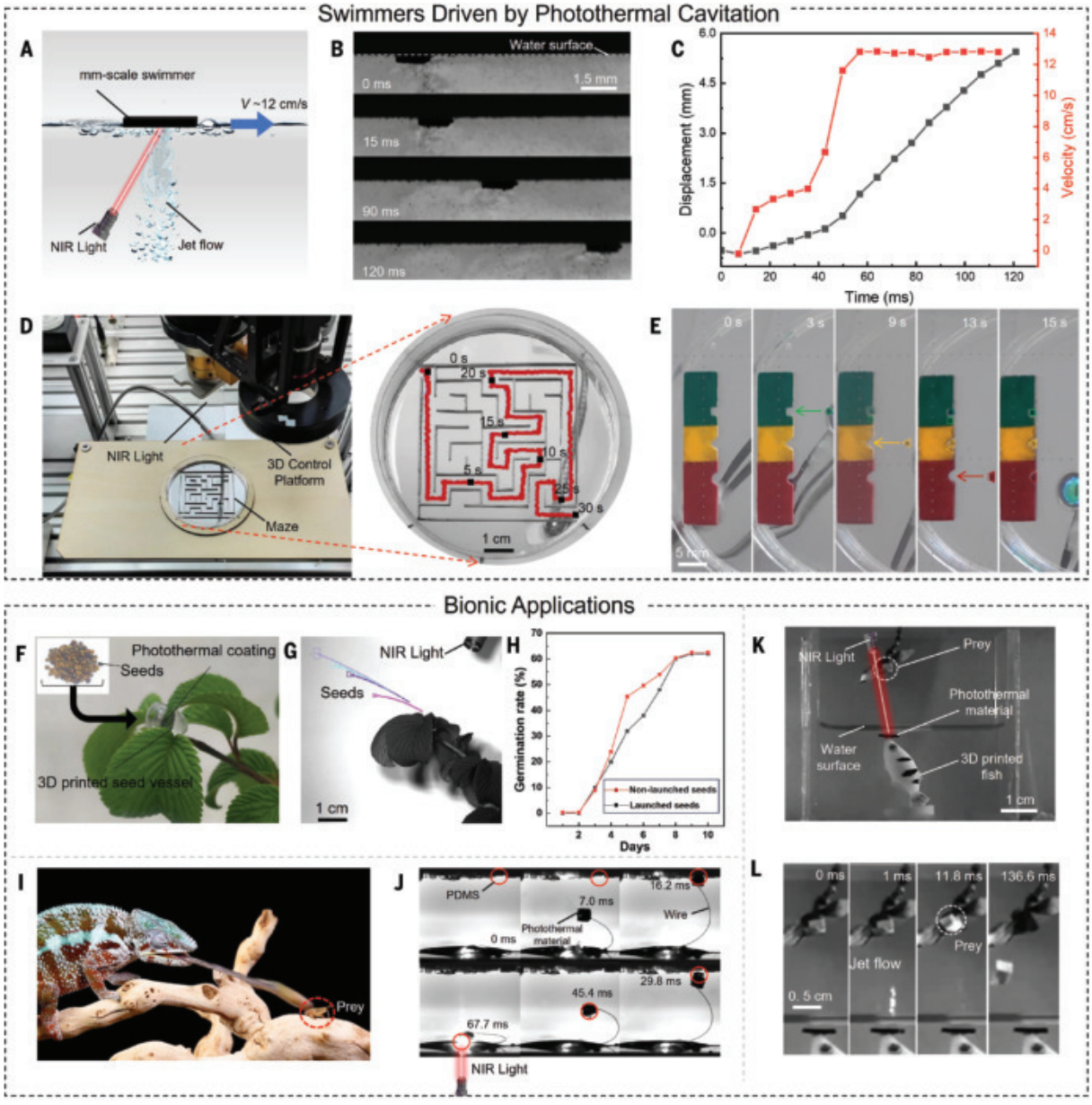


Fig. 4. Swimmers propelled by cavitating jets and bionic applications of cavitation-triggered launching. (**A**) Schematic of a swimmer driven by photothermal cavitation, achieving a swimming speed of ~12 cm/s. (**B**) Photographic sequence capturing the swimming within 120 ms. (**C**) Swimming speed and distance versus time. (**D**) A swimmer navigated a complex maze, closely following the laser movement (movie S8). The inset displays the maze layout, with the swimmer's trajectory indicated by red lines and black squares marking its positions at different times (0 to 30 s). (**E**) Swimmers of various shapes (square, triangular, and trapezoidal shapes) being directed toward and precisely fitting into corresponding notches (movie S8). (**F**) Photograph of the cavitation-assisted seeding system and (**G**) trajectories of launched seeds (movie S9). (**H**) Comparable germination rates between seeds launched by photothermal cavitation and untreated seeds, demonstrating the potential for launching delicate objects by cavitation. (**I**) Chameleons ballistically project their tongues to swiftly and precisely capture prey, which inspires (**J**) the design of an object retrieval system (movie S9). PDMS, polydimethylsiloxane. (**K**) Schematic of an archerfish-inspired liquid jetting system. (**L**) Sequential snapshots from movie S9, depicting the jetting process over 136.6 ms.

simultaneously induce multiple bubbles (fig. S12D). Therefore, acoustic cavitation [Fig. 3A, (III)] may have the potential to launch large-scale objects (fig. S12E) or simultaneously launch multiple objects.

The versatility of our launching strategy is further demonstrated by its adaptability to various environments. Our photothermal jumpers can be deployed under liquids, on liquid surfaces, and on solid surfaces (Fig. 3, B and C, and movie S6), accommodating diverse application scenarios. The jumpers perform best on wet solid surfaces, whereas their performance diminishes as they are submerged in liquids (fig. S13). Because initiating single-bubble cavitation requires only a small amount of liquid, even residual solvent within the jumper material is sufficient to induce cavitation (fig. S13D), thereby minimizing reliance on external liquid environments. However, because cavitation occurs within the launcher without the buffering effect of surrounding liquid, repeated launches may compromise structural integrity.

The substantial force generated by cavitation may raise concerns about precise control over launch direction and trajectory. By adjusting the position of laser irradiation on the jumper (Fig. 3, D and E), the point of force application can be altered, enabling launching over a wide three-dimensional (3D) range (Fig. 3F). For instance, irradiating the right portion of the jumper produces cavitation that generates a counterclockwise rotational moment, elevating its right end and launching the jumper leftward (Fig. 3D and movie S7). Adjusting the launch angle yields different horizontal distances, with a maximum of 1.26 m achieved at angles of 155° and 25° (Fig. 3F).

Swimmers propelled by photothermal cavitation

Cavitation is capable of launching a range of motions beyond simple jumping. By controlling fluid-structure interactions during cavitation near a solid, a millimeter-scale swimmer driven by photothermal cavitation can be engineered. A TiO_2-PPy-TiC platelet was positioned on the water surface. When irradiated from below at an oblique angle and guided by a laser moving at up to 12 cm/s (Fig. 4A), the swimmer accelerates to match the laser's speed and closely follows its trajectory (Fig. 4B and movie S8). A 1.5-mm-long cavitation-driven swimmer can achieve a high swimming speed of ~12 cm/s (Fig. 4C). The moving, obliquely angled laser creates an asymmetric temperature field that initiates a downward cavitating jet at an angle. This jet rebounds toward the water surface and forms an asymmetric loop flow that not only propels the swimmer but also confines it within the irradiated region. Surface tension at the water-air interface prevents the swimmer from being launched into the air. The swimming motion is highly controllable, enabling navigation through complex, confined environments such as mazes (Fig. 4D and movie S8) and microfluidic channels. Additionally, swimmers of various shapes can be guided to fit seamlessly into corresponding notches (Fig. 4E and movie S8), highlighting potential applications in precision assembly, circuit repair, micromanipulation, and beyond (*22*). The millimeter-scale TiO_2-PPy-TiC platelet can also be deployed on the water's bottom and swim at 3 cm/s (fig. S14 and movie S8).

Bioinspiration and potential applications

Biological organisms have evolved various launching mechanisms for rapid movement and powerful actions essential for hunting, reproduction, and other physiological functions. For example, ferns exploit hydrodynamic cavitation within their sporangia to forcefully eject spores for wide dispersal. Inspired by this, we investigated whether photothermal cavitation could similarly propel delicate objects like seeds without compromising their structural integrity or functionality, despite the high local pressures and temperatures. *Oxalis* seeds (1 to 2 mm in diameter) were placed in a custom-built, water-filled container with its bottom coated in our TiO_2-PPy-TiC material (Fig. 4F). Upon irradiation, photothermal cavitation successfully launched the seeds (Fig. 4G and movie S9) to distances >0.7 m (fig. S15). The germination rates of the cavitation-launched seeds were comparable to those of untreated seeds (Fig. 4H). This model scenario thus demonstrates the potential of photothermal cavitation for safely transporting fragile objects—such as microcircuitry, sensors, and signal transmitters—into confined and inaccessible environments. To retrieve objects from such environments, we developed a microscale cargo retrieval device inspired by the chameleon's ballistic tongue (Fig. 4I and fig. S16A). A sticky TiO_2-PPy-TiC gripper was attached to a fixed substrate via a flexible wire. Upon irradiation, the gripper is precisely launched toward the target position to retrieve cargo (Fig. 4J, fig. S17, and movie S9).

Inspired by the archerfish, we also developed a cavitation-powered liquid jetting system by affixing a TiO_2-PPy-TiC material to an underwater substrate (a 3D-printed archerfish in Fig. 4K). Upon cavitation, a high-speed jet with an initial velocity of ~1.8 m/s (fig. S16D) was generated, reaching a maximum height of ~10 cm (fig. S18 and movie S9). The rapid energy transmission of cavitation enables continuous jetting, with intervals of less than 10 ms between successive jets. This high-speed, high-pressure (Fig. 2C) liquid jet has the potential to penetrate biological tissues and inject drugs, providing a promising alternative to needle-based drug injections (*23*).

In conclusion, our study demonstrates that cavitation can serve as an efficient launching mechanism. The rapid transmission of substantial energy during cavitation produces a burst of power and acceleration, imparting high kinetic energy to the launched objects. The jumpers and swimmers launched by cavitation exhibit decent performance. With compatibility across a wide range of device materials, liquid media, external stimuli, and operational environments, our cavitation-based launching strategy can be tailored to meet the demands of different applications.

REFERENCES AND NOTES

1. J. Hu *et al.*, *Angew. Chem. Int. Ed.* **62**, e202218227 (2023).
2. Y. Kim, J. van den Berg, A. J. Crosby, *Nat. Mater.* **20**, 1695–1701 (2021).
3. H. Lee, C. Xia, N. X. Fang, *Soft Matter* **6**, 4342–4345 (2010).
4. E. W. Hawkes *et al.*, *Nature* **604**, 657–661 (2022).
5. M. Loepfe, C. M. Schumacher, U. B. Lustenberger, W. J. Stark, *Soft Robot.* **2**, 33–41 (2015).
6. W. A. Churaman, L. J. Currano, C. J. Morris, J. E. Rajkowski, S. Bergbreiter, *J. Microelectromech. Syst.* **21**, 198–205 (2012).
7. N. W. Bartlett *et al.*, *Science* **349**, 161–165 (2015).
8. C. A. Aubin *et al.*, *Science* **381**, 1212–1217 (2023).
9. M. Li, X. Wang, B. Dong, M. Sitti, *Nat. Commun.* **11**, 3988 (2020).
10. X. Wang *et al.*, *Nat. Mater.* **23**, 1428–1435 (2024).
11. V. P. Carey, *Liquid-Vapor Phase-Change Phenomena: An Introduction to the Thermophysics of Vaporization and Condensation Processes in Heat Transfer Equipment* (CRC Press, 2020).
12. M. Koch, thesis, Göttingen, Georg-August Universität (2020).
13. A. L. King, *Proc. Natl. Acad. Sci. U.S.A.* **30**, 155–161 (1944).
14. X. Noblin *et al.*, *Science* **335**, 1322–1322 (2012).
15. N. A. Alderete *et al.*, *Science* **387**, 659–666 (2025).
16. D. K. Roper, W. Ahn, M. Hoepfner, *J. Phys. Chem. C Nanomater. Interfaces* **111**, 3636–3641 (2007).
17. Y. Liu *et al.*, *Adv. Mater.* **25**, 1353–1359 (2013).
18. M. Koch *et al.*, *Comput. Fluids* **126**, 71–90 (2016).
19. Y. Huang, X. Qin, *Cryogenics* **146**, 104016 (2025).
20. T. Trummler, L. Freytag, S. J. Schmidt, N. A. Adams, in *Proceedings of the 10th International Symposium on Cavitation CAV* (ASME Press, 2018), pp. 656–659.
21. L. Tian *et al.*, *Int. Commun. Heat Mass Transf.* **145**, 106815 (2023).
22. Y. Sun *et al.*, *Angew. Chem. Int. Ed.* **59**, 1098–1102 (2020).
23. G. Arrick *et al.*, *Nature* **636**, 481–487 (2024).

ACKNOWLEDGMENTS

Funding: National Natural Science Foundation of China (grant no. 12261131495 and 12475008) (C.D.), National Natural Science Foundation of China (grant no. 12374281) (G.Z.), Scientific Research and Developed Fund of Zhejiang A&F University (grant no. 2021FR0009) (C.D.). **Author contributions:** Conceptualization: D.L.W., X.H., Z.L., W.L., C.D. Methodology: D.L.W., Z.L., X.H., H.Z., H.Q., G.B., W.L., D.W., G.Z., C.D. Investigation: D.L.W., Z.L., H.Q., H.Z., H.Q. Visualization: D.L.W., Z.L., H.Q., H.Z., H.Q. Funding acquisition: D.L.W., W.L., C.D. Project administration: X.H., Z.L., W.L., D.W., G.Z., C.D. Supervision: X.H., W.L., D.W., G.Z., C.D. Writing – original draft: Z.L., X.H., D.L.W., H.Z., W.L., D.W., CD. Writing – review & editing: Z.L., X.H., D.L.W., G.B., H.Z., C.C., P.S., Y.D., W.L., D.W., G.Z., C.D. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adu8943
Materials and Methods; Supplementary Text; Figs. S1 to S19; Movies S1 to S9

Submitted 26 November 2024; accepted 20 June 2025

10.1126/science.adu8943

QUANTUM NETWORKS

Classical-decisive quantum internet by integrated photonics

Yichi Zhang[1], Robert Broberg[2], Alan Zhu[3], Gushu Li[3], Li Ge[4,5], Jonathan M. Smith[3], Liang Feng[1,2]*

Classical and quantum technologies have traditionally been viewed as orthogonal, with classical systems being deterministic and quantum systems inherently probabilistic. This distinction hinders the development of a scalable quantum internet even as the global internet continues expanding. We report a classical-decisive quantum internet architecture in which the integration of quantum information into advanced photonic technologies enables efficient entanglement distribution over a commercially deployed fiber network. On-chip precise synchronization between classical headers and quantum payloads enables dynamic routing and networking of high-fidelity entanglement guided by classical light. The quantum states are preserved through real-time error mitigation, relying solely on classical signal readout without disturbing quantum information. These classical-decisive features demonstrate a practical path to a scalable quantum internet using existing network infrastructure and operating systems.

Rapid advancements in quantum technologies are building the foundation for the future information infrastructure, making computing faster (*1*, *2*) and sensing technology more intelligent (*3*) than ever before. Nevertheless, quantum systems operate mainly as standalone systems and have yet to be interconnected to achieve distributed quantum information processing, which takes advantage of nonlocal quantum entanglement to further enhance quantum capabilities (*4*–*6*). The major obstacle to this is the absence of fully standalone quantum interconnects, necessitating the development of new switching and routing protocols (*7*) capable of handling quantum information while maintaining its coherence and quantum correlation. Operations on quantum states often require premeasurements, but these measurements erase the statistical characteristics of quantum states, thereby potentially destroying quantum signals. This demands resource-intensive centralized management, with entanglement distribution paths typically pre-allocated and static (*8*–*12*). Moreover, a quantum internet is highly susceptible to noise and photon loss, demanding real-time error mitigation to preserve the integrity of quantum information (*11*).

A possible solution is to explore classical-decisive quantum systems. At the physical layer, both quantum and classical information converge on the same electromagnetic guiding mechanism. Such a hybrid system can leverage existing classical interconnect hardware, circumventing the need for entirely new quantum-specific infrastructure while simultaneously providing flexibility to incorporate quantum functionalities. Recent attempts have demonstrated quantum key distribution and entanglement distribution over fiber links alongside classical data traffic (*13*–*17*). New protocols incorporating different multiplexing techniques have been tested to efficiently code the quantum-classical hybrid data internet protocol (IP) packet stream for the optimization of quantum network management (*18*–*21*). In this scenario, classical signals carry labels and routing instructions to guide quantum data with precision and stability while avoiding direct measurement and manipulation of quantum states. However, large-scale explorations assessing classical-decisive quantum information processing in deployed classical fiber networks are still lacking. This gap exists because quantum-classical hybrid data co-control, noise suppression, and other unpredictable environmental disturbances in a deployed network are far more complex than in fiber setups in laboratory settings (*22*, *23*). These complexities demand precise orchestration of the quantum-classical hybrid data IP packet to enable scalability across the globally deployed classical fiber internet infrastructure.

To address this challenge, we demonstrate the integration of quantum information into advanced photonic technologies (*24*–*26*), enabling the synchronization of quantum and classical data on the same photonic chip with well-controlled accuracy and latency. The control plane integrates into a commercially deployed fiber network. Controlled by the photonic chip, classical light orchestrates entangled photon distribution and qubit operations in deployed fiber networks, facilitating efficient quantum-classical hybrid internet management following the widely used IP architecture. The precision and stability of chip-controlled classical light improves the preparation, manipulation, and control of complex quantum states, leading to real-time quantum error mitigation for robust quantum operations against inevitable real-world environmental disturbances.

Chip-powered quantum-classical hybrid networking framework

The core of the quantum-classical hybrid networking framework is an IP packet with hybrid data that includes a leading classical header followed by a quantum payload (Fig. 1A). The IP packet encapsulates the quantum payload carrying qubits and the classical header containing metadata for systematic quantum internetwork management. Using this packet format, entangled photons are produced and enclosed inside the quantum payload. The classical header consists of multiple segments, each assigned a specific classical-decisive function in the quantum internetwork, including an IP header, control of quantum data readout, labeling, duration of the quantum payload, and real-time monitoring of quantum state errors. Additional segments can be easily inserted to support more classical-decisive quantum controls as needed, providing excellent scalability for our framework to handle more sophisticated quantum tasks. Because hybrid data packets are a key enabler of quantum information distribution, it is crucial to minimize interference from high-intensity classical light on the delicate quantum payload and to package all packets with high precision and control. The precise synchronization and continuous generation and manipulation of these hybrid data IP packets rely on a quantum-classical hybrid photonic server chip that precisely transmits classical light-guided quantum information through on-chip multiplexing.

After being transmitted from the chip server, hybrid data IP packets propagate through existing commercial network infrastructure composed of routers, nodes, and fiber links (Fig. 1B). At each router, the IP header containing origin and destination addresses is used for relaying incoming traffic to the correct outgoing fiber. Once determined, the router controller rapidly reprograms the switching network for the upcoming quantum payload, dynamically routing entanglement distribution between end points. Because the router only processes classical data without interacting with the quantum payload, existing commercial fiber routers can support this protocol with network interface–specific hardware modifications to ensure the maximal integrity of the quantum payload. Therefore, a quantum-classical hybrid version of IP can be conducted on existing communication infrastructure.

[1]Department of Materials Science and Engineering, University of Pennsylvania, Philadelphia, PA, USA. [2]Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, USA. [3]Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA, USA. [4]Department of Physics and Astronomy, College of Staten Island, CUNY, Staten Island, NY, USA. [5]The Graduate Center, CUNY, New York, NY, USA.
*Corresponding author. Email: fenglia@seas.upenn.edu

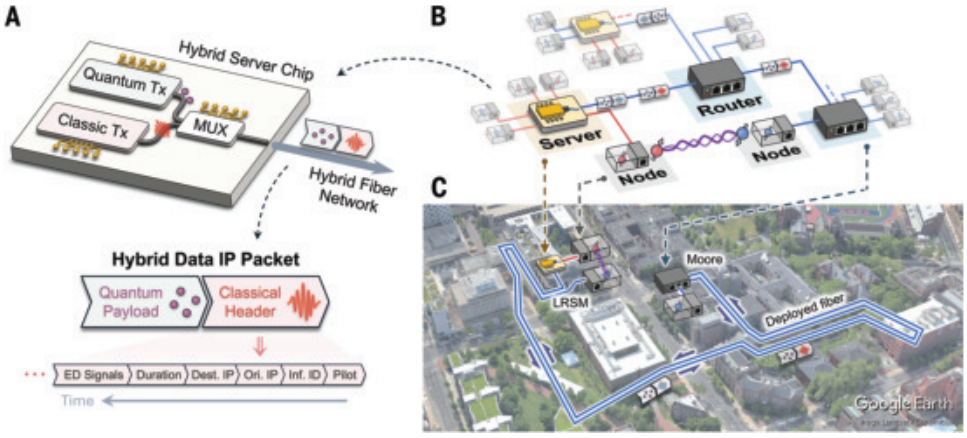


Fig. 1. Schematic of the classical-decisive quantum internet. (**A**) The quantum internet is driven by a hybrid server chip containing a classical transmitter (Tx), a quantum Tx, and a WDM multiplexer (MUX). The classical header contains multiple important information segments related to the routing and monitoring of quantum payload. These segments, in our protocol, include a header "Pilot" as the indicator for data readout, an "Information ID" (Inf. ID) labeling each hybrid data IP packet, Ori. IP and Dest. IP addresses, a "duration" specifying the quantum payload duration, and "error detection" (ED) signals for real-time monitoring of state errors along the propagating path. (**B**) Schematic of a commercial fiber infrastructure with quantum-classical hybrid internet working on it. Deployed fiber links interconnect servers, routers, and nodes to dynamically distribute hybrid data IP packets in which the classical header facilitates quantum payload routing and management. Links with two colors represent the path for two server-generated entangled photons. (**C**) A multinode campus-level testbed at the University of Pennsylvania. An integrated chip server is deployed at the Laboratory for Research on the Structure of Matter (LRSM), a router is deployed at the Moore Building (Moore), and multiple fiber links connect the two buildings.

To experimentally validate this framework, a multinode quantum networking testbed was established using a deployed fiber network (Fig. 1C) (*27*). Multiple underground commercial fiber links connect two buildings, with each link spanning ~1 km. The insertion loss for a single link is ~4.3 dB due to the presence of several switching racks along the path. This setup realistically reflects real-world conditions, in which communication paths typically contain multiple interconnected fibers (and are thus subject to photon loss and unpredictable environmental disturbances) rather than a single intact fiber running from start to end. The designed hybrid data IP packet, traveling through the fiber-optical network, uses the classical header to determine the switching path and required operations for the quantum payload, enabling efficient distribution of entangled photons (*27*). Quantum state analysis and detection nodes are also placed in the network for real-time error monitoring and mitigation. This testbed can be further upscaled to a larger network by activating additional fiber links, routers, and nodes.

The photonic server chip (Fig. 2A) was designed and fabricated on a Si_3N_4-on-insulator platform (figs. S1 and S2) that incorporates multiple thermo-optic tunable components for information encoding and multiplexing, facilitating precise quantum-classical hybrid data co-control (*27*). A continuous-wave laser beam at 1550.92 nm is coupled into the chip and split by a Mach-Zehnder interferometer (MZI) power control unit, which balances the power between the classical and quantum transmitters. For the classical header, on-off keying modulation encodes the classical IP header, whereas for the quantum payload, the laser beam acts as a pump light, generating entangled photon pairs through spontaneous four-wave mixing in high-quality-factor microring resonators (*28*, *29*). The generated broadband photon pairs are symmetrically distributed with respect to laser wavelength, with signal photons centered at 1547.72 nm and idler photons at 1554.13 nm. A subsequent wavelength demultiplexer made of band-pass microring filters separates the path-entangled photon pairs. By leveraging the intrinsically distinct wavelengths of the nonlinear-generated photon pairs from the pump light, wavelength division multiplexing (WDM) minimizes the interference from the high-intensity classical signal on quantum payloads when forming the hybrid data IP packet. The idler photons propagate locally to an on-chip state analyzer, whereas signal photons are combined with the prepared classical header through WDM. Additionally, cointegration and co-control of both classical and quantum transmitters allow for precise time division multiplexing, ensuring that the classical header precedes the quantum payload. The application of well-established multiplexing techniques on the chip guarantees minimal overlap between classical and quantum signals to preserve the integrity of quantum information. All optical signals then couple off-chip into fibers, with the idler photons entering a single-photon detection module directly and the IP packets traversing the network after being converted from path to polarization degree of freedom, ultimately reaching the node for state analysis and detection (fig. S3).

Before conducting classical-decisive quantum networking, we characterized the generated quantum-classical hybrid data stream. The entangled photon pair sources achieved a measured coincidence-to-accident ratio exceeding 80 and a coherence time of ~0.54 ns (Fig. 2B). The generated biphoton Bell state $|\Phi^+\rangle = (|HH\rangle + |VV\rangle)/\sqrt{2}$ was verified by quantum state tomography, demonstrating a high fidelity of $F = 0.983$ and a high purity of $P = 0.979$ (Fig. 2C), matching theoretical predictions (Fig. 2D). The temporal profile of an exemplary final hybrid data IP packet is shown in Fig. 2E, where the photon pair coincidence counts span the whole quantum payload duration and different information segments were accurately encoded onto the classical header (Fig. 2F), confirming the power of integrated photonic chips to precisely encode and transmit IP packets.

Multinode hybrid IP networking for efficient entanglement distribution

Facilitated by the hybrid server and existing fiber network infrastructure, a proof-of-principle hybrid IP networking experiment was demonstrated. Throughout the test, the independently preprogrammed server, router, and nodes worked collaboratively under our designed protocol and effectively eliminated the need for a resource-demanding top-tier manager (*27*).

Initially, packet switching was verified on a simple network consisting of one "Origin IP" (Ori. IP) address {1} and two "Destination IP" (Dest. IP) addresses {1, 2} connected through one router relaying data among three deployed fiber links (Fig. 3A). The IP addresses encoded on the classical header signal the router to switch the subsequent quantum payload to the desired destination. To mimic quantum internet traffic in which quantum information is transmitted on request by users, each IP packet randomly encodes the destination and contains a random duration for the quantum payload. In experiments, entanglement distribution is recorded and labeled by the biphoton correlation time delay ΔT, which reveals the routing pattern (Fig. 3B). For example, when the "1–1" (from origin 1 to destination 1) path is selected, correlated biphoton signals only appear at time delay ΔT_2, with no signal observed at ΔT_1 and vice versa for the "1–2" (from origin 1

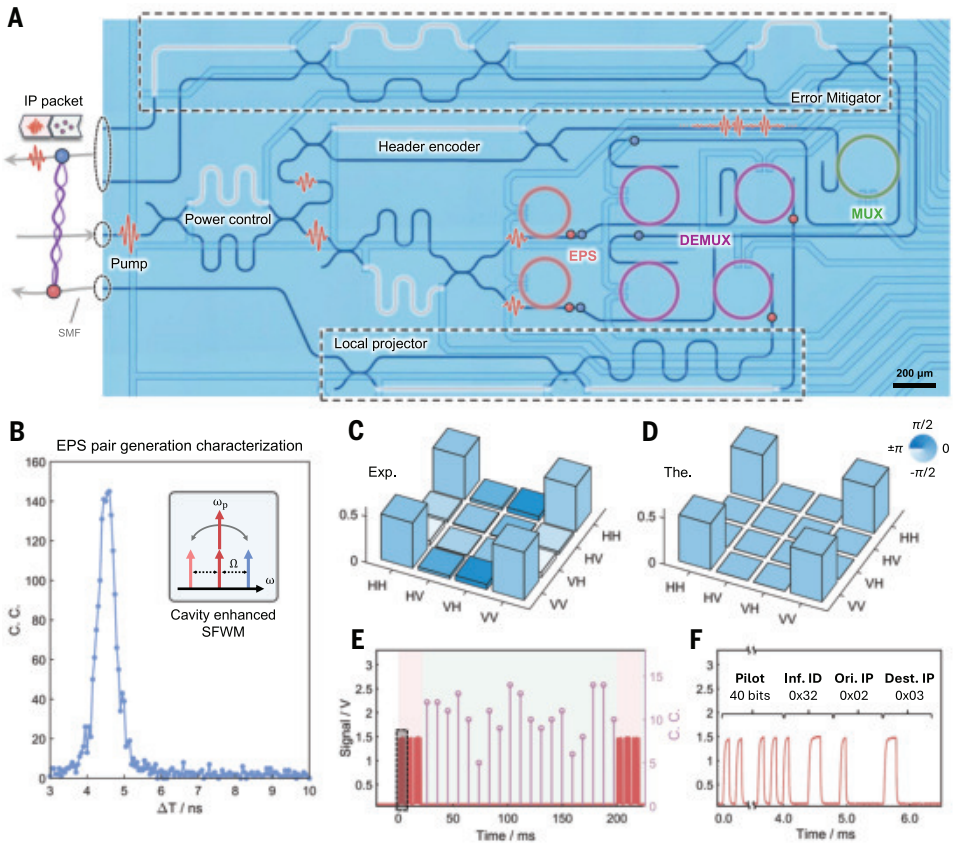


Fig. 2. The hybrid server chip. (**A**) Optical microscopic photo of a fabricated photonic chip. The pump light enters through an MZI to control the power between classical Tx (upper path) and quantum Tx (lower path). The classical Tx encodes classical header with an MZI (header encoder), and the quantum Tx contains an entangled photon source (EPS), which generates path-entangled photon pairs through two microring resonators facilitating cavity-enhanced spontaneous four-wave mixing (SFWM). The generated signal and idler photons are then separated by a WDM demultiplexer (DEMUX). Although idler photons propagate locally to a local projector for state analysis, signal photons are combined with classical headers through a multiplexer (MUX) to form hybrid data IP packets before coupling off-chip into single-mode fibers (SMFs). (**B**) Coincident counting characterization of a photon-pair source. Here, two pump photons with frequency ω_p are simultaneously absorbed and converted to a signal photon at $\omega_p + \Omega$ and an idler photon at $\omega_p - \Omega$ (inset), creating a biphoton correlation peak at a specific time delay ΔT. (**C** and **D**) Measured and theoretical biphoton density matrices of Bell-state $|\Phi^+\rangle$. (**E**) Generated hybrid data IP packets with classical header signals (red) and a time-domain histogram of heralded signal photons (purple). (**F**) Magnified view of the initial classical header marked by the gray box in (E).

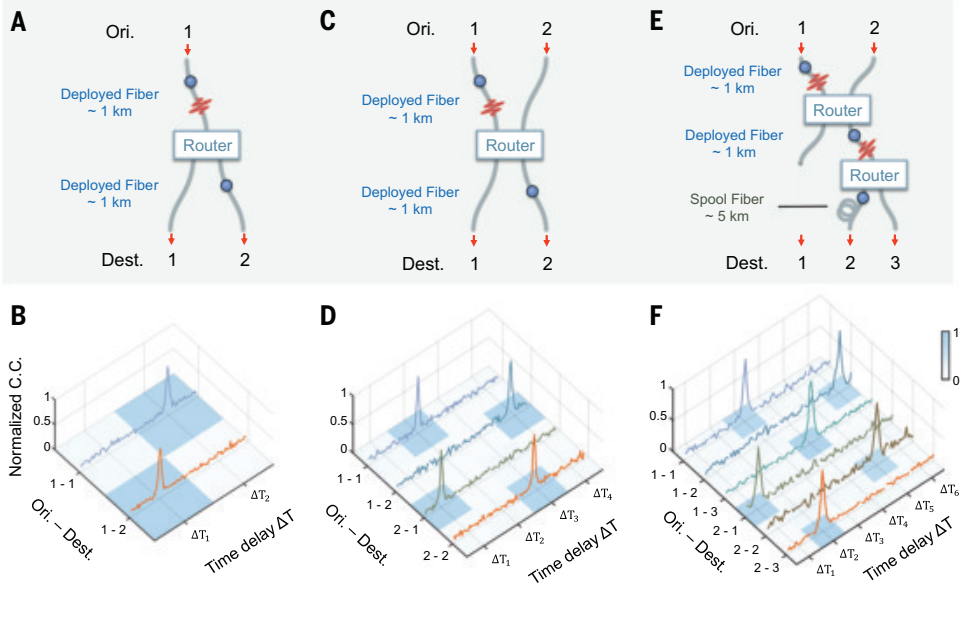


Fig. 3. Hybrid IP networking. (**A**) Packet-switching verification with a simple network in which one Ori. IP address and two Dest. IP addresses are connected by one router. (**B**) The resulting normalized temporal distribution of coincidence count (C.C.) for two origin-destination path configurations. The calculated 2×2 transfer matrix (shown beneath the C.C. curves) shows a quantum routing precision of $p_{R1} = 0.971$. (**C**) Further validation through a network containing two Ori. IP addresses and two Dest. IP addresses. (**D**) Normalized temporal distribution of C.C. for four path configurations and the resulting 4×4 transfer matrix with a quantum routing precision of $p_{R2} = 0.962$. (**E**) Double-layer packet-switching test in which two Ori. IP addresses and three Dest. IP addresses connect through two routers, representing a more authentic communication situation. (**F**) Normalized temporal distribution of C.C. for six path configurations and the resulting 6×6 transfer matrix with a quantum routing precision of $p_{R3} = 0.916$.

to destination 2) path configuration. The difference in ΔT ($\Delta T_2 - \Delta T_1 \approx 50$ ns) matches the length difference between two paths. The resulting 2×2 transfer matrix shows a quantum routing precision (*27*) of $p_{R1} = 0.971$ (Fig. 3B). The slight deviation from the ideal scenario of $p'_{R1} = 1.0$ primarily arises from quantum noise throughout measurements, convincingly showing the viability of the router hardware and the routing protocol for quantum networking.

The hybrid IP networking was further validated in more complicated fiber networks. An additional 1×2 optical switch placed after the hybrid server chip directs the hybrid data IP packets into two paths, facilitating two Ori. IP addresses, which, alongside the two Dest. IP addresses {1, 2} associated with a router, create four possible network connections (Fig. 3C). Biphoton correlation signals appear at four distinct time delays, in which the measured 4×4 transfer matrix

exhibits a quantum routing precision of p_{R2} = 0.962 (Fig. 3D). Moreover, in a double-layer packet switching setup, an additional router leads to three Dest. IP addresses {1, 2, 3}, forming six network connections (Fig. 3E). The measured 6 × 6 transfer matrix shows a quantum routing precision of p_{R3} = 0.916, where the slight drop compared with p_{R1} and p_{R2} is not caused by the increased network complexity but rather is due to the extra insertion loss from the 5-km-long spool fiber after the second router. These experiments collectively show that our hybrid IP networking is scalable and practical for larger entanglement network deployments. The demonstrated protocol can be easily extended to align with standard IPv4 or IPv6 formats (table S1) (*27*) because it has no inherent limitations on bit counts or switch scaling. Therefore, in the context of a large-scale quantum internet, this mechanism can facilitate the dynamic and flexible interconnection of hundreds or even thousands of nodes.

Entanglement distribution with real-time error mitigation

Although hybrid IP networking is capable of flexibly distributing correlated photons among multiple nodes, the real-world quantum networks demand robust preservation of delicate quantum states under environment-induced disturbances. For example, in a fiber network, such disturbances can be induced by vibrations, temperature fluctuations, and geometrical deformations and rotations (*22*, *30*), causing the polarization state of photons to experience an unknown general $SU(2)$ transformation $U_F(\alpha, \beta, \gamma) = R_z(\alpha) \cdot R_y(\beta) \cdot R_z(\gamma)$, where $R_i(\theta)$ is an $SU(2)$ rotation matrix along the i-th axis with a rotation angle of θ_i (Fig. 4A). Without in situ monitoring and correction of polarization variation, substantial errors would emerge at the nodes and consequently deteriorate quantum information integrity.

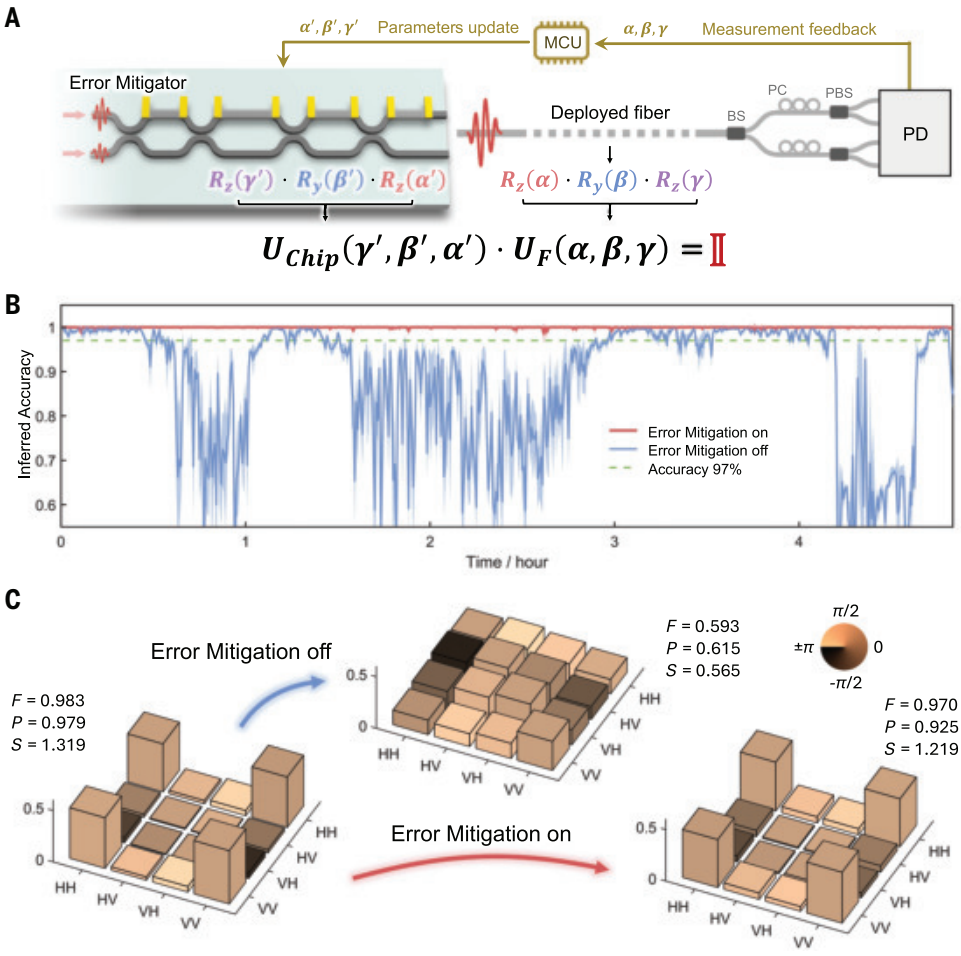


Fig. 4. Entanglement distribution protected by real-time error mitigation. (**A**) Schematic of the error-mitigation algorithm. The error-detection (ED) signals, primarily comprising controllable initial polarization states, are placed inside classical headers. At the end point, a state analyzer and a photodiode (PD) detect the polarization state to extract critical parameters of the unknown $SU(2)$ matrix associated with the fiber for the on-chip error mitigator to perform a precorrection to eliminate the polarization errors. BS: beam splitter; PC: polarization controller; PBS: polarization beam splitter. (**B**) Long-term performance of quantum information accuracy with and without error mitigation. Each error mitigation cycle takes ~3.2 ms. The shadow area below each curve represents 1σ of the accuracy oscillation. (**C**) Quantum state tomography of distributed entangled photons. With error mitigation on, the acquired density matrix demonstrates fidelity F = 0.970 and purity P = 0.925 very close to the original characterized one, whereas without error mitigation, the biphoton state becomes highly mixed with fidelity F = 0.593 and purity P = 0.615. The contrast on mutual information (S = 1.219 versus S = 0.565) also verifies the critical role of our hybrid strategy for preserving quantum bipartite entanglement.

Our hybrid networking framework actively harnesses guiding classical signals to address this challenge by detecting and correcting any quantum state errors through an active real-time feedback and optimization algorithm based on the error-detection signals embedded in the classical header (*27*). The algorithm is effectively enabled by dynamical reconfiguration of the on-chip error mitigator consisting of four MZIs (Fig. 4A). From a fundamental perspective, both classical and quantum signals obey the same electromagnetic mechanisms so they experience identical environmental disturbances. Therefore, by sending and checking different classical polarization states, $U_F(\alpha, \beta, \gamma)$ can be reconstructed, retrieving α, β, and γ. Subsequently, a microcontroller unit processes these parameters and updates the voltages applied onto the error mitigator to form a precorrected $SU(2)$ transformation $U_{\text{Chip}}(\gamma', \beta', \alpha') = R_z(\gamma') \cdot R_y(\beta') \cdot R_z(\alpha')$. By leveraging the feedback flow and optimization algorithm (fig. S7), a perfect error cancellation is possible under the condition $U_{\text{Chip}}(\gamma', \beta', \alpha') \cdot U_F(\alpha, \beta, \gamma) = \text{I}$. In this way, the classical header enables error mitigation of the quantum payload, creating a robust path for precise entanglement distribution. All measurements performed are on classical polarization states without interfering with quantum states, thereby substantially increasing detection and analysis efficiency.

During a 5-hour long-term comparison between error mitigation on and off (Fig. 4B), a clear difference in classical-light inferred quantum key distribution accuracy can be observed. The two cases both started from a set of parameters with corrected U_F. With error mitigation enabled, the inferred accuracy remained >97% for the duration of the experiment. By contrast, without error mitigation, the inferred accuracy significantly dropped due to accumulated polarization drift, and severe oscillations caused by random environmental vibrations during certain periods. The quantum error mitigation was further confirmed by characterizing entanglement fidelity (Fig. 4C). Without error

mitigation, the entangled state experiences severe decoherence and evolves into a highly mixed state. However, with continuous error mitigation every 100 ms, the original biphoton Bell state can be successfully retrieved with an almost unaffected entangled state fidelity of $F = 0.970$. The substantial improvement convincingly demonstrates the effectiveness of classical-decisive quantum information processing in real-time high-efficiency quantum error mitigation.

Conclusions

Integrating quantum information processing with advanced photonic technologies allows for the precise encoding of a quantum-classical hybrid data stream, enabling a classical-decisive quantum internet that seamlessly integrates with existing commercial communication infrastructures. In the designed hybrid IP networking protocol, the classical signal plays a critical role in efficiently guiding entanglement distribution while preserving the integrity of quantum information. Additionally, robust entanglement distribution is achieved through real-time error feedback driven by an error detection and feedback algorithm encoded in classical light, which can in principle be extended to support other error mitigation modalities such as phase stabilization (*27*, *31*). When integrated with state-of-the-art high-speed photonic components (*32*–*34*) and combined with advanced multiplexing techniques (*8*, *11*, *27*, *35*), our demonstrated classical-decisive quantum architecture can be further enhanced to increase the hybrid data rate, improving the overall efficiency of quantum networking. As quantum memories become available (*12*, *36*–*38*), they can readily be supported by our system, facilitating a practical and resilient quantum internet for large-scale distributed quantum information systems (*27*).

REFERENCES AND NOTES

1. F. Arute *et al.*, *Nature* **574**, 505–510 (2019).
2. H.-S. Zhong *et al.*, *Science* **370**, 1460–1463 (2020).
3. C. L. Degen, F. Reinhard, P. Cappellaro, *Rev. Mod. Phys.* **89**, 035002 (2017).
4. H. J. Kimble, *Nature* **453**, 1023–1030 (2008).
5. S. Wehner, D. Elkouss, R. Hanson, *Science* **362**, eaam9288 (2018).
6. A. S. Cacciapuoti *et al.*, *IEEE Netw.* **34**, 137–143 (2020).
7. Y. Lee, E. Bersin, A. Dahlberg, S. Wehner, D. Englund, *NPJ Quantum Inf.* **8**, 75 (2022).
8. S. Wengerowsky, S. K. Joshi, F. Steinlechner, H. Hübel, R. Ursin, *Nature* **564**, 225–228 (2018).
9. Y.-A. Chen *et al.*, *Nature* **589**, 214–219 (2021).
10. M. Pompili *et al.*, *NPJ Quantum Inf.* **8**, 121 (2022).
11. Y. Zheng *et al.*, *Science* **381**, 221–226 (2023).
12. J.-L. Liu *et al.*, *Nature* **629**, 579–585 (2024).
13. N. A. Peters *et al.*, *New J. Phys.* **11**, 045012 (2009).
14. J. M. Thomas, G. S. Kanter, P. Kumar, *Opt. Express* **31**, 43035–43047 (2023).
15. I. A. Burenkov *et al.*, *Opt. Express* **31**, 11431–11446 (2023).
16. J. M. Thomas *et al.*, *Optica* **11**, 1700–1707 (2024).
17. Z.-Q. Zhong *et al.*, *Optica* **11**, 1056–1061 (2024).
18. S. DiAdamo, B. Qi, G. Miller, R. Kompella, A. Shabani, *Phys. Rev. Res.* **4**, 043064 (2022).
19. R. Mandil, S. DiAdamo, B. Qi, A. Shabani, *NPJ Quantum Inf.* **9**, 85 (2023).
20. S. J. B. Yoo *et al.*, *IEEE Commun. Mag.* **62**, 76–81 (2024).
21. M. B. On *et al.*, *J. Lightwave Technol.* **42**, 3504–3514 (2024).
22. D.-D. Li *et al.*, *Opt. Express* **26**, 22793–22800 (2018).
23. D. Bunandar *et al.*, *Phys. Rev. X* **8**, 021009 (2018).
24. J. Wang, F. Sciarrino, A. Laing, M. G. Thompson, *Nat. Photonics* **14**, 273–284 (2020).
25. C. Gao, P. Gaur, D. Almutairi, S. Rubin, Y. Fainman, *Nat. Commun.* **14**, 4421 (2023).
26. H. Zhao *et al.*, *eLight* **4**, 10 (2024).
27. Materials and methods are available as supplementary materials.
28. X. Lu *et al.*, *Nat. Phys.* **15**, 373–381 (2019).
29. R. Chen *et al.*, *Phys. Rev. Lett.* **133**, 083803 (2024).
30. G. B. Xavier, G. Vilela de Faria, G. P. Temporão, J. P. von der Weid, *Opt. Express* **16**, 1867–1873 (2008).
31. M. Pittaluga *et al.*, *Nat. Photonics* **15**, 530–535 (2021).
32. C. Wang *et al.*, *Nature* **562**, 101–104 (2018).
33. C. Wang *et al.*, *Nature* **629**, 784–790 (2024).
34. X. Feng *et al.*, *Nat. Photonics* **19**, 264–270 (2025).
35. Y. Zhang *et al.*, *Phys. Rev. X* **15**, 011024 (2025).
36. A. I. Lvovsky, B. C. Sanders, W. Tittel, *Nat. Photonics* **3**, 706–714 (2009).
37. M. F. Askarani *et al.*, *Phys. Rev. Lett.* **127**, 220502 (2021).
38. C. M. Knaut *et al.*, *Nature* **629**, 573–578 (2024).
39. Y. Zhang *et al.*, Data for: Classical-decisive quantum internet by integrated photonics, version 1, Figshare (2025); https://doi.org/10.6084/m9.figshare.29199629.v1.

ACKNOWLEDGMENTS

Funding: This work was supported by the Gordon and Betty Moore Foundation (grant GBMF12960 and DOI 10.37807 to L.F. and Y.Z.); the Office of Naval Research (ONR grant N00014-23-1-2882 to L.F.); the National Science Foundation (NSF grant DMR-2323468 to L.F. and A.Z.); an Olga and Alberico Pompa endowed professorship (J.M.S. and R.B.); and by PSC-CUNY (grant ENHC-54-93 to L.G.). **Author contributions:** Chip design and fabrication: Y.Z.; Conceptualization: L.F., Y.Z., R.B., J.M.S.; Data analysis: Y.Z., R.B., A.Z., L.G., L.F.; Investigation: Y.Z., A.Z.; Supervision: L.F.; Writing – original draft: Y.Z., L.F.; Writing – review & editing: Y.Z., L.F., R.B., G.L., A.Z., L.G., J.M.S. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main manuscript or supplementary materials or have been deposited at Figshare (*39*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adx6176
Materials and Methods; Supplementary Text; Figs. S1 to S8; Table S1; References (*40*–*48*)

Submitted 21 March 2025; accepted 23 June 2025

10.1126/science.adx6176

Growing into mentorship

Jasmine Gabriel Hughes

As I waited for the Teams meeting to begin, I started to question myself. "Wait, who am I to be mentoring someone?" I thought. I was just a first-year graduate student who still regularly sought guidance myself; what advice could I have to offer? Months earlier, I had applied to be a mentor through a program at my university that provides free support for potential Ph.D. applicants from groups that are historically underrepresented in science. I am passionate about helping students from backgrounds like mine, and I was eager to pay forward the guidance I had received earlier in my journey. But now that the moment was here, I was overcome with doubt.

As an undergraduate, I had no idea how to become a scientist. The process felt opaque and overwhelming, particularly to a first-generation college student. I always felt as though I was behind my peers, simply because I did not know how to access certain resources or get involved. Still, I pushed forward, learning the hard way through trial and error.

Things began to turn around at my first meeting with the professor who would become my lab supervisor. I was extremely nervous, but she was welcoming and understanding, genuinely interested in learning about me and my career goals. Throughout college, she provided support, professionally and personally, bolstering my confidence, helping me understand it is OK to take time away from lab for family, and more.

Being a mentor at my Ph.D. university seemed a great opportunity to do the same for others. I enthusiastically applied and was excited to be selected and matched with a mentee. But as our first meeting drew close, uncertainty crept in. There was no guidebook to follow. How should I structure our meetings? What if she asks a question that I have no idea how to answer? How could I be ready for this type of leadership role, when I still had so far to go myself?

That day of our first meeting, I was terrified. But once my mentee joined the call, seeming very enthusiastic about meeting me, and started to talk about herself, I had a flashback to my own college experience. I remembered struggling to navigate getting into a research lab and applying to summer internships and graduate school. The fellow first-generation student on the other side of the screen was probably going through something similar—feeling both uncertainty and a fierce determination to figure it out and achieve her professional goals.

What mattered, I realized, was not to be some imaginary perfect mentor with all the answers, but to get to know my mentee, including her hopes and ambitions, and offer whatever guidance and support I could based on my own experiences. We ended the meeting having set some practical goals for the year—including writing her personal statement and practicing research presentations—and just as important, laid the grounds for an authentic, personal relationship.

As our sessions continued, I still went into each one worried I would not be prepared to solve every problem my mentee encountered. But over time, I realized I could help in practical ways. I could equip her with the skills to tackle obstacles, such as answering difficult questions during interviews and research presentations. Just like my mentee, I had dreaded the "tell me about yourself" prompt; where do you start and how much should you tell? In my case, I had found a happy balance by explaining how being diagnosed with mixed connective tissue disease during college had driven me to pursue a Ph.D.—but I didn't go into details that would have felt invasive and draining. I described my approach to my mentee, so she could adopt the parts that resonated with her.

When I did not know the answer to an issue she raised, I was honest about it and did my best to listen, provide feedback and guidance, and allow her to determine her best course of action. Sometimes I was there simply to provide a safe space to vent. I could help even when I did not have a solution.

A few months after submitting her graduate school applications, my mentee sent an email thanking me for my support and guidance, which she said helped increase her confidence. She probably doesn't know that she helped increase my confidence, too. □

Jasmine Gabriel Hughes is a Ph.D. student at the University of Cambridge.
She thanks Professor Elizabeth Ables for her mentorship.

ILLUSTRATION: ROBERT NEUBECKER

Science AAAS

CHEN TIANQIAO & CHRISSY INSTITUTE

CO-HOST ETH zürich

AI & MENTAL HEALTH

SEPTEMBER 8–9, 2025 | ZURICH

Mental health is a vast and growing worldwide problem. In addition to the suffering experienced by individuals and their affected family members it also has an enormous economic impact. Because it can affect people at a young age they will often need treatment for many years and during that period they will not be able to participate to their full potential in society.

REGISTER TODAY!