A new kind of organ donation stirs an ethical storm p. 572

A fast radio burst with a flipping magnetic field pp. 580 & 599

Does fishing policy drive underfishing? p. 585

Science

$15
12 MAY 2023
science.org

AAAS



HOLD YOUR BREATH

Deep-diving hammerheads close their gill slits to keep warm pp. 583 & 651

EDITORIAL

STEM must meet people where they are



Sethuraman Panchanathan is director of the US National Science Foundation, Alexandria, VA, USA. spanchan@nsf.gov

Ideas in science and engineering drive innovation, entrepreneurship, and economic growth—this is true around the world, inspiring people to want to be part of this enterprise. I hear this across the United States. Everywhere I go, people tell me that they want opportunities for themselves and their families to learn and solve real-world problems, to revitalize their communities, and to find paths to build a better life without necessarily having to move away from home. In short, people want to pursue possibilities where they live. Fortunately, this is feasible and finally starting to happen.

In the United States, the number of students earning science and engineering doctoral degrees has doubled since 1985, but a PhD is not always needed to participate in the science, technology, engineering, and mathematics (STEM) workforce. In the skilled technical workforce, there are 20 million people without a bachelor's degree who contribute. STEM jobs at all levels proved more resilient during the pandemic, with lower unemployment across the STEM labor force compared to the non-STEM labor force. STEM workers also earn more than non-STEM workers, regardless of sex, race, ethnicity, or disability status. It is therefore unsurprising that many working people, particularly in lower-paid, non-STEM jobs in the service industry (such as customer service and sales), are eager to acquire new or improved skills to transition into the STEM workforce. But what is also true is that most people looking to attain a good STEM job are not necessarily able to relocate to where the opportunities are because of personal situations such as financial constraints, housing availability, and caretaking responsibilities.

> "...people looking to attain a good STEM job are not necessarily able to relocate..."

Offshoring (which erased 5 million middle class manufacturing jobs), the 2008 economic recession, and COVID-19 have particularly affected women, people of color, and rural communities—those largely missing from the scientific enterprise and who deserve STEM opportunities. Including a diversity of ideas, experiences, contexts, talents, and geographies is therefore crucial to expanding innovation capacity and meeting the demand for a skilled STEM workforce. Currently, just 15 states supply 76% of the skilled domestic workforce in knowledge and technology industries. Half of all US innovation jobs in STEM fields are concentrated in just 41 counties.

As the nation's largest independent federal funding agency for basic research, the US National Science Foundation (NSF) is trying to meet people where they are by building realistic paths locally for them to enter the scientific workforce. It launched a new Directorate for Technology, Innovation and Partnerships (TIP) in 2022 aimed at connecting millions to STEM opportunities in the form of education, training, and jobs everywhere. For example, the agency is partnering with companies such as Micron, Intel, and IBM to develop a robust workforce in semiconductor design and manufacturing through student and teacher development programs in community colleges and minority-serving institutions. This will create avenues for high-tech jobs in more places so that people can contribute their ideas and talent where they live.

And rather than trying to turn the Midwest into a Silicon Valley or the South into a Kendall Square, the agency launched its Regional Innovation Engines program last year, which aims to bring together governments, institutes of higher education, employers, labor unions, and community-based organizations to spur regional innovation and development. The idea is to leverage local assets and innovation potential specific to each region and unleash entrepreneurship, accelerate commercialization, retain value locally, and create a virtuous, sustainable cycle that attracts follow-on regional investments. The first awards for this program will soon be announced.

There are other promising programs with similar goals. For example, the Department of Commerce is launching regional technology hubs to expand economic opportunity across the United States. And as part of the Microelectronics Commons, a national network, the Department of Defense is establishing regional capabilities to create direct pathways to commercialization for US microelectronics researchers and designers.

The most important US science and engineering legislation in a generation, the landmark CHIPS and Science Act of 2022, passed Congress with bipartisan support. Building on this Act, the Fiscal Year 2023 enacted budget for NSF is $9.9 billion, which is an excellent start to rapidly scale local STEM pathways. If the United States is to vigorously compete with other countries, including China, then it must ensure that the CHIPS and Science Act authorizations become definitive budget appropriations. There are good initiatives in the works right now. Let's accelerate progress.

–Sethuraman Panchanathan

PHOTO: STEPHEN VOSS

10.1126/science.adi5538

NEWS

$30,000,000,000

Estimated cost to plug 14,000 inactive oil and gas wells in the U.S. Gulf of Mexico and inland waters that may be adding to carbon emissions. The government could hold oil and gas producers liable. (*Nature Energy*)

IN BRIEF

Edited by **Jeffrey Brainard**



Workers funded by the EcoHealth Alliance collected bats for study from a cave in south China in 2019.

PANDEMICS

NIH restarts halted bat virus grant

Three years after then-President Donald Trump pressured the U.S. National Institutes of Health (NIH) to shut down a research grant to a group studying how bat coronaviruses jump to people, the agency has restarted the award. The new 4-year grant providing $576,000 per year is a stripped-down version of a 2019 grant to the EcoHealth Alliance, a nonprofit in New York City. That grant included a subaward to China's Wuhan Institute of Virology (WIV), which some conservative commentators allege started the COVID-19 pandemic. The restarted grant omits WIV, sampling of bats and people in China, and controversial experiments with live, hybrid viruses. Instead, it funds studies of bat virus' genomes and characteristics that could cause human disease and lab work with viral proteins and "pseudoviruses" that can't cause disease. NIH has also imposed extensive new accounting rules on EcoHealth. In January, a federal audit found that EcoHealth had misreported nearly $90,000 in expenses for several grants dating to 2014 and that NIH had erred by not justifying its termination (later changed to a suspension) of the 2019 grant.

FDA approves first RSV vaccine

INFECTIOUS DISEASES | A decadeslong effort to develop a vaccine against respiratory syncytial virus (RSV), a respiratory pathogen that targets infants and the elderly, succeeded last week when the U.S. Food and Drug Administration (FDA) green-lighted one made by manufacturer GlaxoSmithKline. The vaccine, the first ever approved, will become available to people ages 60 and older if, as expected, an advisory committee at the Centers for Disease Control and Prevention recommends it next month. A similar vaccine by pharmaceutical giant Pfizer is also expected to win approval for use in older people this month. And FDA has committed to deciding by August whether to allow use of the Pfizer vaccine in pregnant people, who pass protective antibodies to their newborns. Annually, RSV kills an estimated 33,000 people ages 60 and older in hospitals in high-income countries. It also kills an estimated 46,000 infants younger than 7 months old, most of them in developing countries.

Minisatellites eye cyclones

ATMOSPHERIC SCIENCE | A pair of small satellites funded by NASA to study how hurricanes rapidly intensify launched this week from New Zealand, carried on a Rocket Lab Electron rocket. The Tropics satellites, each about the size of a shoebox, sport miniature microwave sensors like those typically carried on larger, more expensive weather satellites, allowing them to see through clouds to measure water vapor, precipitation, and temperature. The two CubeSats will be joined by a second pair later this month; the constellation will fill in the 6-hour gaps in data from larger weather satellites. The first two Tropics satellites were lost in June 2022 when the rocket company Astra failed to deliver them into orbit. That prompted NASA to select Rocket Lab for the two remaining launches.

CDC chief to step down

LEADERSHIP | Rochelle Walensky, who directed the U.S. Centers for Disease Control and Prevention (CDC) through a turbulent pandemic era, will leave her job at the end of June, she announced last week. Walensky took up her post in January 2021 after

leading the infectious diseases division at Massachusetts General Hospital. During the COVID-19 pandemic, CDC was buffeted by criticism, including that it bowed to former President Donald Trump's administration's political agenda; after President Joe Biden named Walensky director, she made communications missteps seen as further eroding public trust in the agency. Last summer, she launched a reorganization that included a focus on making CDC's public communications better and faster. The timing to nominate a successor is uncertain. Beginning on 20 January 2025, the U.S. Senate will need to confirm any new nominee for the position.

Human 'pangenome' released

BIOMEDICINE | Researchers presented this week the first draft of the human pangenome—a collection of sequenced genomes belonging to individuals of diverse ancestry that aims to eventually represent as many as possible of the DNA sequences found across our species. Scientists have been using a single reference human genome—mostly representing people of European ancestry—since it was unveiled 20 years ago. The Human Pangenome Reference Consortium diversified that reference by sequencing the genomes of 47 individuals from all continents except Antarctica. Published in a series of papers in *Nature*, the pangenome adds 119 million base pairs to the current reference genome and doubles the known number of structural variants. Researchers hope to identify potentially harmful variants present only in some populations, enabling more equitable health care.

WHO fired COVID-19 investigator

#METOO | The World Health Organization (WHO) has dismissed Peter Ben Embarek, the scientist who led its probe of the origins of COVID-19, following findings he committed sexual misconduct. The dismissal occurred last year but was reported by the *Financial Times* only last week. The allegations concern incidents in 2015 and 2017 and were first received by WHO investigators in 2018, and the organization followed due process, a spokesperson says. WHO did not disclose additional details. Ben Embarek told Reuters that he "contests the qualification of harassment" and that he was "quite hopeful in the defense of my rights." In March, WHO adopted a revised policy to stop sexual harassment and misconduct. It aimed to address what it called shortcomings in its response to allegations that its workers sexually exploited and abused women and girls during an Ebola outbreak in the Democratic Republic of the Congo from 2018 to 2020.



A "rocking wall" strip helped protect a wood building shaken by an earthquake simulator (not visible) beneath it.

PHOTO: DAVID BAILLOT/JACOBS SCHOOL OF ENGINEERING/UNIVERSITY OF CALIFORNIA, SAN DIEGO

ENGINEERING

Tall wood building is shaken, but not scathed

A 10-story wooden building survived two severe, simulated earthquakes intact this week as scientists sought to show that wood can rival steel and concrete to safely support tall buildings. The structure is the tallest ever tested at the University of California, San Diego's earthquake simulator. As part of a project funded by the U.S. National Science Foundation, the machine subjected the building to the equivalent of the 1994 6.7 magnitude Northridge, California, quake. A few minutes later it re-created a 1999 7.7 magnitude temblor that struck Taiwan. The building is reinforced by narrow wooden panels running from the ground to the top of each of its four sides. These "rocking walls" are pinned in place by metal rods anchored to the ground and running though the panels to the top. They enable the walls to sway during a quake, then return to vertical. Scientists say using wood to construct tall buildings could lower their environmental footprint.



COVID-19

End of COVID-19 emergencies sparks debate

Moves by WHO and U.S. dial back data collection, reporting mandates

By **Kai Kupferschmidt** *and* **Meredith Wadman**

The World Health Organization (WHO) last week declared an end to the emergency phase of the COVID-19 pandemic, just as a similar emergency in the United States expires on 11 May. Both moves usher the world into a new phase of disease monitoring, with scaled back surveillance and fewer resources to fight SARS-CoV-2.

WHO's emergency committee recommended ending the Public Health Emergency of International Concern (PHEIC), the highest alert level WHO can declare, that has been in effect since 30 January 2020. "It's therefore with great hope that I declare COVID-19 over as a global health emergency," WHO's director-general, Tedros Adhanom Ghebreyesus, said at a press conference in Geneva.

The pandemic had been on a downward trajectory for a year, Tedros said, allowing most countries to return to life as it was before the new coronavirus emerged. Tedros emphasized that this declaration does not mean COVID-19 is no longer a threat. "The worst thing any country could do now is to use this news as a reason to let down its guard, to dismantle the systems it has built, or to send the message to its people that COVID-19 is nothing to worry about," he said.

Both the U.S. wind-down and the WHO decision indicate that "the emergency and acuity part is in the rearview mirror" for governments and organizations, says Beth Blauer, a data and public policy expert at Johns Hopkins University. "This need to have COVID-19 at the fore as a dedicated declared emergency is no longer considered the best use of collective resources."

Since the pandemic's start 3 years ago, WHO has recorded close to 7 million deaths from COVID-19, and the real death toll may be three times that. Although deaths are now lower than they have been since March 2020, a few thousand are still being reported to the agency each week. In his statement, Tedros highlighted COVID-19's ongoing impact, emphasizing not just the deaths, but also that thousands of people around the world are still being treated for the disease in intensive care units and millions of others are dealing with the prolonged aftereffects of SARS-CoV-2 infections. "This virus is here to stay," he said. "It's still killing and it's still changing. The risk remains of new variants emerging that cause new surges in cases."

Coming just a week later, the end of the U.S. public health emergency will have immediate policy impacts. The flow of migrants at the southern border is likely to increase with the end of pandemic-era restrictions, and the federal government's provision of free rapid antigen tests will end.

In a press conference explaining some of the changes, officials from the U.S. Centers for Disease Control and Prevention (CDC) conceded they will also lose some insight into the pandemic as their data collection powers are dialed back. When the U.S. emergency ends, CDC will no longer be able to compel testing laboratories to report their COVID-19 test results, or states to report numbers of vaccinations, although many are expected to continue to report the latter voluntarily. Community levels, case numbers, and test positivity rates—the percentage of positive tests among all tests taken—will no longer appear on CDC's popular data tracker. The changes in CDC's capabilities come as roughly 1100 people in the U.S. are dying weekly of COVID-19, most of them people 65 and older.

Still, senior CDC officials insist the remaining data will be adequate. "We have the right data for this phase of COVID-19," said Nirav Shah, CDC's principal deputy director. Shah said the agency's primary surveillance metric in the new regime will be hospitalization data, which will be reported from all hospitals weekly and not daily as it currently is. In an era in which people don't test themselves or report results when they do, "Hospitalizations provide the best national level view of COVID trends," he said.

Tools such as wastewater testing and genomic sequencing will still be used, giving CDC a better grasp of SARS-CoV-2's distribution and evolution than for other common respiratory pathogens, Shah stressed. COVID-19–associated deaths will be published weekly as a percentage of all U.S. deaths.

But some public health experts disagree

that the new data landscape will be adequate. "It is a mistake to unwind the data collection and reporting infrastructure too aggressively," says David O'Connor, a virologist at the University of Wisconsin-Madison. "I worry that … will leave us more vulnerable the next time."

WHO's decision to end the PHEIC carries risks, too, said Didier Houssin, chair of the agency's emergency committee, comparing the situation to a hermit crab transitioning from one shell to another. The risks the committee debated included that a new variant will take the world by surprise, that countries will lower their guard, and that access to vaccines will become more difficult. But those risks had to be balanced with a realistic picture of the waning pandemic, Houssin said.

And keeping a PHEIC in place too long also carries risks, says Lawrence Gostin, director of the O'Neill Institute for National and Global Health Law at Georgetown University. "The public loses trust in WHO and other public health agencies." Prolonged emergencies also dilute the power of such declarations, he says. The danger is that, "The public will not readily support declaring an emergency for future crises."

An ongoing WHO review of the International Health Regulations should also address the uncertainty about when to end a PHEIC, says Alexandra Phelan, a global health expert at the John Hopkins Bloomberg School of Public Health. "We have an opportunity to provide greater clarity about the end of PHEICs: How do we transition out of an emergency without losing political, financial, and technical commitment to addressing the structural and long-term impacts to public health?"

In announcing the end of the PHEIC, Tedros said WHO will establish a review committee tasked with developing "long-term, standing recommendations for countries on how to manage COVID-19 on an ongoing basis." But many places have stopped doing much to combat the disease, says Gregg Gonsalves, a public health expert at the Yale School of Public Health. "Almost everywhere we've declared mission accomplished and we have no appetite for doing anything more to fight this disease," he says. "We're willing to bake in a huge amount of morbidity and mortality to get back to normal [and] it doesn't bode well for facing our future."

In his press conference, Tedros also urged all countries to do more to prepare for future pandemics. The tools to detect and respond to pandemics better are available, he said, "But globally, a lack of coordination, a lack of equity, and a lack of solidarity meant that those tools were not used as effectively as they could have been. Lives were lost that should not have been." ■

ANIMAL RESEARCH

U.S., European researchers face monkey shortage crisis

Two new reports highlight growing demand and shrinking availability of nonhuman primates for biomedical studies

By **David Grimm**

The supply of monkeys for research is shrinking in the United States and Europe, and access to remaining animals is becoming increasingly unreliable. That's the dire message of two new reports, which conclude that the situation is compromising critical biomedical research now—and will continue to do so well into the future.

The U.S. report, released last week by the National Academies of Sciences, Engineering, and Medicine, is the strongest official statement yet on the precarious state of monkey research. Sponsored by the National Institutes of Health (NIH) and based on information from government agencies, National Primate Research Centers, academic institutions, and more than 200 scientists who use nonhuman primates, it finds that nearly two-thirds of U.S. researchers have reported challenges in obtaining monkeys. This has delayed studies, forced scientists to use fewer and less ideal animals (heavier or younger than the norm, for example), and, in some cases, has led to projects being abandoned. Monkey prices have also skyrocketed, putting these primates out of reach for many researchers.

The shortage has developed as Asian sources of monkeys have dried up and domestic primate centers have been unable to expand their breeding programs enough to keep up with growing biomedical demand, the U.S. report says. The analysis also suggests that China is competing with the U.S. for the purchase of nonhuman primates from Southeast Asia, further straining supply. "Relying on importing these animals from other countries is unsustainable," the report argues.

A European report released last month reached similar conclusions. The demand for research monkeys is so high in EU nations, and the supply is so low, that the EU will not be able to meet its goal of creating sustainable domestic breeding colonies of these animals. "Unless a solution is found, nonhuman primate research in Europe will dwindle and even stop in certain places," says Kirk Leech, executive director of the European Animal Research Association. "We're moving into a very uncertain world."

The reports come during perhaps the most tumultuous 3-year period in the history of monkey research. The U.S. uses about 70,000 monkeys per year in studies of the brain, infectious disease, and aging; the EU about 5000; and the U.K. about 2000. In the early days of COVID-19, China stopped exporting research monkeys, shutting off a key pipeline for the U.S. and U.K. The vast ma-



PHOTO: KATHLEEN FLYNN/REUTERS

Rhesus macaques at the Tulane National Primate Research Center, which breeds monkeys for U.S. researchers.

that the new data landscape will be adequate. "It is a mistake to unwind the data collection and reporting infrastructure too aggressively," says David O'Connor, a virologist at the University of Wisconsin-Madison. "I worry that … will leave us more vulnerable the next time."

WHO's decision to end the PHEIC carries risks, too, said Didier Houssin, chair of the agency's emergency committee, comparing the situation to a hermit crab transitioning from one shell to another. The risks the committee debated included that a new variant will take the world by surprise, that countries will lower their guard, and that access to vaccines will become more difficult. But those risks had to be balanced with a realistic picture of the waning pandemic, Houssin said.

And keeping a PHEIC in place too long also carries risks, says Lawrence Gostin, director of the O'Neill Institute for National and Global Health Law at Georgetown University. "The public loses trust in WHO and other public health agencies." Prolonged emergencies also dilute the power of such declarations, he says. The danger is that, "The public will not readily support declaring an emergency for future crises."

An ongoing WHO review of the International Health Regulations should also address the uncertainty about when to end a PHEIC, says Alexandra Phelan, a global health expert at the John Hopkins Bloomberg School of Public Health. "We have an opportunity to provide greater clarity about the end of PHEICs: How do we transition out of an emergency without losing political, financial, and technical commitment to addressing the structural and long-term impacts to public health?"

In announcing the end of the PHEIC, Tedros said WHO will establish a review committee tasked with developing "long-term, standing recommendations for countries on how to manage COVID-19 on an ongoing basis." But many places have stopped doing much to combat the disease, says Gregg Gonsalves, a public health expert at the Yale School of Public Health. "Almost everywhere we've declared mission accomplished and we have no appetite for doing anything more to fight this disease," he says. "We're willing to bake in a huge amount of morbidity and mortality to get back to normal [and] it doesn't bode well for facing our future."

In his press conference, Tedros also urged all countries to do more to prepare for future pandemics. The tools to detect and respond to pandemics better are available, he said, "But globally, a lack of coordination, a lack of equity, and a lack of solidarity meant that those tools were not used as effectively as they could have been. Lives were lost that should not have been." ■

ANIMAL RESEARCH

U.S., European researchers face monkey shortage crisis

Two new reports highlight growing demand and shrinking availability of nonhuman primates for biomedical studies

By **David Grimm**

The supply of monkeys for research is shrinking in the United States and Europe, and access to remaining animals is becoming increasingly unreliable. That's the dire message of two new reports, which conclude that the situation is compromising critical biomedical research now—and will continue to do so well into the future.

The U.S. report, released last week by the National Academies of Sciences, Engineering, and Medicine, is the strongest official statement yet on the precarious state of monkey research. Sponsored by the National Institutes of Health (NIH) and based on information from government agencies, National Primate Research Centers, academic institutions, and more than 200 scientists who use nonhuman primates, it finds that nearly two-thirds of U.S. researchers have reported challenges in obtaining monkeys. This has delayed studies, forced scientists to use fewer and less ideal animals (heavier or younger than the norm, for example), and, in some cases, has led to projects being abandoned. Monkey prices have also skyrocketed, putting these primates out of reach for many researchers.

The shortage has developed as Asian sources of monkeys have dried up and domestic primate centers have been unable to expand their breeding programs enough to keep up with growing biomedical demand, the U.S. report says. The analysis also suggests that China is competing with the U.S. for the purchase of nonhuman primates from Southeast Asia, further straining supply. "Relying on importing these animals from other countries is unsustainable," the report argues.

A European report released last month reached similar conclusions. The demand for research monkeys is so high in EU nations, and the supply is so low, that the EU will not be able to meet its goal of creating sustainable domestic breeding colonies of these animals. "Unless a solution is found, nonhuman primate research in Europe will dwindle and even stop in certain places," says Kirk Leech, executive director of the European Animal Research Association. "We're moving into a very uncertain world."

The reports come during perhaps the most tumultuous 3-year period in the history of monkey research. The U.S. uses about 70,000 monkeys per year in studies of the brain, infectious disease, and aging; the EU about 5000; and the U.K. about 2000. In the early days of COVID-19, China stopped exporting research monkeys, shutting off a key pipeline for the U.S. and U.K. The vast ma-



PHOTO: KATHLEEN FLYNN/REUTERS

Rhesus macaques at the Tulane National Primate Research Center, which breeds monkeys for U.S. researchers.

jority of these were cynomolgus macaques (or cynos), a species mostly used by private industry for drug and vaccine research. (Rhesus macaques, the primary monkey species used in U.S. academic labs, are mostly bred domestically in the U.S.)

The U.S. and other countries were able to make up for some of the shortfall by importing more monkeys from Southeast Asia and the Indian Ocean island of Mauritius. But the pandemic spiked demand, sucking up monkeys for vaccine research. "COVID left the cupboard bare, and we're still struggling to recover," says Jonah Sacha, an immunologist at Oregon Health & Science University who uses rhesus macaques and cynos to study various infectious diseases.

Last year, Air France became the final major airline to refuse to transport monkeys for research, increasing the difficulty of getting these animals to labs. The International Union for Conservation of Nature also classified cynos as endangered, prompting fears that the animals would become even harder to import.

Then, in November 2022, Cambodia—which had largely made up for the shuttered Chinese pipeline by supplying more than 29,000 cynos in 2020, the vast majority to the U.S.—was hit with a major smuggling scandal. U.S. investigators charged several individuals with illegally exporting wild-caught cynos to the U.S. and falsely labeling them as captive-bred.

Since then, two of the world's largest suppliers of cynos—Inotiv and Charles River Laboratories—have stopped their imports from Cambodia. Both companies tell *Science* they are working on new guidelines that would ensure any monkeys they import have a legitimate provenance.

Domestic breeding facilities in the U.S. and Europe have also struggled to meet demand. "There are certain categories of research we've had to abandon," says JoAnne Flynn, a microbiologist at the University of Pittsburgh who studies tuberculosis in both cynos and rhesus macaques. The high price of monkeys is limiting the number of studies she can do, she says, because her NIH grants don't allow for the extra expense. "It's an enormous problem."

Sacha says the situation is especially bad for cynos, which he must purchase from private suppliers. "Before the pandemic, you could get them almost instantaneously for about $2000. Now, it's taking at least a year and half, and I'm competing with others willing to pay more than $19,000 per animal." To combat supply problems, the National Academies report says the U.S. should expand domestic breeding programs and increase coordination between the various domestic monkey suppliers. The country should also continue to invest in nonanimal research tools such as "organ on a chip" technology, which the report says are looking more promising than ever, though not quite ready for prime time.

Joyce Cohen, associate director of animal resources at the Emory National Primate Research Center, says even if all of the money and infrastructure for a domestic breeding program materialized tomorrow, it could take 4 to 5 years to start providing even some of the animals academic scientists need. Expanding current primate research centers to meet that need could cost hundreds of millions of dollars. Still, Cohen hopes these challenges won't dissuade NIH from seriously considering the investment. "You have to start somewhere."

Sacha says he feels vindicated by the National Academies report. "It goes right to the heart of what we've been saying for the past 5 to 10 years," he says. "This is where our medicine comes from. We have to invest in this. This is the health of America."

Flynn agrees. "China is cornering the market on monkey research," she says. "If we don't maintain our scientific infrastructure in the future, who's going to own the vaccines?"

The EU's problems are even more stark, Leech says. Its solution also lies in more domestic breeding. But unlike the U.S., the EU cannot legally import wild-caught monkeys to start or supplement breeding colonies. That's a huge obstacle, he says. "I don't see any short-term solution to this problem."

Eric Kleiman, a researcher at the Animal Welfare Institute, an animal advocacy group, says he hopes the supply crisis will force scientists and government agencies to more seriously consider—and invest in—alternatives to animals in research. "Increased domestic breeding is not a panacea."

A U.S. consultant on industry and academic monkey research says major pharmaceutical and biotech companies are getting "a lot more serious" about investing in domestic breeding. "There is a rising level of concern that this is a major issue," says the source, who has worked in the field for decades but asked not to be named because of concerns about damaging relationships with his clients. "There have been discussions about whether and how they should all work together to ensure supply."

But as the crisis intensifies, he says companies may need to start asking themselves whether the amount of money and effort it's going to take to meet demand is worth it—or even possible. Depending on the answer, the biomedical industry may need to rethink its priorities, he says: "Are we in the drug discovery business, or are we in the monkey breeding business?" ■



PUBLISHING

New tools show promise for tackling paper mills

Fake papers—a symptom of academia's "publish or perish" pressure—are alarmingly common, new findings suggest

By **Jeffrey Brainard**

When neuropsychologist Bernhard Sabel put his new fake-paper detector to work, he was "shocked" by what it found. After screening some 5000 papers, he estimates up to 34% of neuroscience papers published in 2020 were likely made up or plagiarized; in medicine, the figure was 24%. Both numbers, which he and colleagues report in a medRxiv preprint this week, are well above levels they calculated for 2010—and far larger than the 2% baseline estimated in a 2022 publishers' group report.

"It is just too hard to believe" at first, says Sabel of Otto von Guericke University Magdeburg and editor-in-chief of *Restorative Neurology and Neuroscience*. It's as if "somebody tells you 30% of what you eat is toxic."

His findings underscore what was widely suspected: Journals are awash in a rising tide of scientific manuscripts from paper mills—secretive businesses that allow researchers to pad their publication records by paying for fake papers or undeserved authorship. "Paper mills have made a fortune by basically attacking a system that has had no idea how to cope with this stuff," says Dorothy Bishop, a University of Oxford psychologist who studies fraudulent publishing practices. A 2 May announcement from the publisher Hindawi underlined the threat: It shut down four of its journals it found were "heavily compromised" by articles from paper mills.

Sabel's tool relies on just two indicators—

jority of these were cynomolgus macaques (or cynos), a species mostly used by private industry for drug and vaccine research. (Rhesus macaques, the primary monkey species used in U.S. academic labs, are mostly bred domestically in the U.S.)

The U.S. and other countries were able to make up for some of the shortfall by importing more monkeys from Southeast Asia and the Indian Ocean island of Mauritius. But the pandemic spiked demand, sucking up monkeys for vaccine research. "COVID left the cupboard bare, and we're still struggling to recover," says Jonah Sacha, an immunologist at Oregon Health & Science University who uses rhesus macaques and cynos to study various infectious diseases.

Last year, Air France became the final major airline to refuse to transport monkeys for research, increasing the difficulty of getting these animals to labs. The International Union for Conservation of Nature also classified cynos as endangered, prompting fears that the animals would become even harder to import.

Then, in November 2022, Cambodia—which had largely made up for the shuttered Chinese pipeline by supplying more than 29,000 cynos in 2020, the vast majority to the U.S.—was hit with a major smuggling scandal. U.S. investigators charged several individuals with illegally exporting wild-caught cynos to the U.S. and falsely labeling them as captive-bred.

Since then, two of the world's largest suppliers of cynos—Inotiv and Charles River Laboratories—have stopped their imports from Cambodia. Both companies tell *Science* they are working on new guidelines that would ensure any monkeys they import have a legitimate provenance.

Domestic breeding facilities in the U.S. and Europe have also struggled to meet demand. "There are certain categories of research we've had to abandon," says JoAnne Flynn, a microbiologist at the University of Pittsburgh who studies tuberculosis in both cynos and rhesus macaques. The high price of monkeys is limiting the number of studies she can do, she says, because her NIH grants don't allow for the extra expense. "It's an enormous problem."

Sacha says the situation is especially bad for cynos, which he must purchase from private suppliers. "Before the pandemic, you could get them almost instantaneously for about $2000. Now, it's taking at least a year and half, and I'm competing with others willing to pay more than $19,000 per animal." To combat supply problems, the National Academies report says the U.S. should expand domestic breeding programs and increase coordination between the various domestic monkey suppliers. The country should also continue to invest in nonanimal research tools such as "organ on a chip" technology, which the report says are looking more promising than ever, though not quite ready for prime time.

Joyce Cohen, associate director of animal resources at the Emory National Primate Research Center, says even if all of the money and infrastructure for a domestic breeding program materialized tomorrow, it could take 4 to 5 years to start providing even some of the animals academic scientists need. Expanding current primate research centers to meet that need could cost hundreds of millions of dollars. Still, Cohen hopes these challenges won't dissuade NIH from seriously considering the investment. "You have to start somewhere."

Sacha says he feels vindicated by the National Academies report. "It goes right to the heart of what we've been saying for the past 5 to 10 years," he says. "This is where our medicine comes from. We have to invest in this. This is the health of America."

Flynn agrees. "China is cornering the market on monkey research," she says. "If we don't maintain our scientific infrastructure in the future, who's going to own the vaccines?"

The EU's problems are even more stark, Leech says. Its solution also lies in more domestic breeding. But unlike the U.S., the EU cannot legally import wild-caught monkeys to start or supplement breeding colonies. That's a huge obstacle, he says. "I don't see any short-term solution to this problem."

Eric Kleiman, a researcher at the Animal Welfare Institute, an animal advocacy group, says he hopes the supply crisis will force scientists and government agencies to more seriously consider—and invest in—alternatives to animals in research. "Increased domestic breeding is not a panacea."

A U.S. consultant on industry and academic monkey research says major pharmaceutical and biotech companies are getting "a lot more serious" about investing in domestic breeding. "There is a rising level of concern that this is a major issue," says the source, who has worked in the field for decades but asked not to be named because of concerns about damaging relationships with his clients. "There have been discussions about whether and how they should all work together to ensure supply."

But as the crisis intensifies, he says companies may need to start asking themselves whether the amount of money and effort it's going to take to meet demand is worth it—or even possible. Depending on the answer, the biomedical industry may need to rethink its priorities, he says: "Are we in the drug discovery business, or are we in the monkey breeding business?" ■



PUBLISHING

New tools show promise for tackling paper mills

Fake papers—a symptom of academia's "publish or perish" pressure—are alarmingly common, new findings suggest

By **Jeffrey Brainard**

When neuropsychologist Bernhard Sabel put his new fake-paper detector to work, he was "shocked" by what it found. After screening some 5000 papers, he estimates up to 34% of neuroscience papers published in 2020 were likely made up or plagiarized; in medicine, the figure was 24%. Both numbers, which he and colleagues report in a medRxiv preprint this week, are well above levels they calculated for 2010—and far larger than the 2% baseline estimated in a 2022 publishers' group report.

"It is just too hard to believe" at first, says Sabel of Otto von Guericke University Magdeburg and editor-in-chief of *Restorative Neurology and Neuroscience*. It's as if "somebody tells you 30% of what you eat is toxic."

His findings underscore what was widely suspected: Journals are awash in a rising tide of scientific manuscripts from paper mills—secretive businesses that allow researchers to pad their publication records by paying for fake papers or undeserved authorship. "Paper mills have made a fortune by basically attacking a system that has had no idea how to cope with this stuff," says Dorothy Bishop, a University of Oxford psychologist who studies fraudulent publishing practices. A 2 May announcement from the publisher Hindawi underlined the threat: It shut down four of its journals it found were "heavily compromised" by articles from paper mills.

Sabel's tool relies on just two indicators—



authors who use private, noninstitutional email addresses, and those who list an affiliation with a hospital. It isn't a perfect solution, because of a high false-positive rate. Other developers of fake-paper detectors, who often reveal little about how their tools work, contend with similar issues.

Still, the detectors raise hopes for gaining the advantage over paper mills, which churn out bogus manuscripts containing text, data, and images partly or wholly plagiarized or fabricated, often massaged by ghost writers. Some papers are endorsed by unrigorous reviewers solicited by the authors. Such manuscripts threaten to corrupt the scientific literature, misleading readers and potentially distorting systematic reviews. The recent advent of artificial intelligence tools such as ChatGPT has amplified the concern.

To fight back, the International Association of Scientific, Technical, and Medical Publishers (STM), representing 120 publishers, is leading an effort called the Integrity Hub to develop new tools. STM is not revealing much about the detection methods, to avoid tipping off paper mills. "There is a bit of an arms race," says Joris van Rossum, the Integrity Hub's product director. He did say one reliable sign of a fake is referencing many retracted papers; another involves manuscripts and reviews emailed from internet addresses crafted to look like those of legitimate institutions.

Twenty publishers—including the largest, such as Elsevier, Springer Nature, and Wiley—are helping develop the Integrity Hub tools, and 10 of the publishers are expected to use a paper mill detector the group unveiled in April. STM also expects to pilot a separate tool this year that detects manuscripts simultaneously sent to more than one journal, a practice considered unethical and a sign they may have come from paper mills. Such large-scale cooperation is meant to improve on what publishers were doing individually and to share tools across the publishing industry, van Rossum says.

"It will never be a [fully] automated process," he says. Rather, the tools are like "a spam filter … you still want to go through your spam filter every week" to check for erroneously flagged legitimate content.

STM hasn't yet generated figures on accuracy or false-positive rates because the project is too new. But catching as many fakes as possible typically produces more false positives. Sabel's tool correctly flagged nearly 90% of fraudulent or retracted papers in a test sample. For every 56 true fakes it detected, however, it erroneously flagged 44 genuine papers, so results still need to be confirmed by skilled reviewers. Other paper mill detectors typically have a similar tradeoff, says Adam Day, founding director of a startup called Clear Skies who consulted with STM on the Integrity Hub. But without some reliance on automated methods, "You either have to spot check randomly, or you use your own human prejudice to choose what to check. And that's not generally very fair."

Scrutinizing suspect papers can be time-consuming: In 2021, Springer Nature's post-publication review of about 3000 papers suspected of coming from paper mills required up to 10 part- and full-time staffers, said Chris Graf, the company's director of research integrity, at a U.S. House of Representatives subcommittee hearing about paper mills in July 2022. (Springer Nature publishes about 400,000 papers annually.)

Newly updated guidelines for journals issued in April may help ease the workload. They may decide to reject or retract batches of papers suspected of having been produced by a paper mill, even if the evidence is circumstantial, says the nonprofit Committee on Publication Ethics, which is funded by publishers. Its previous guidelines encouraged journals to ask authors of each suspicious paper for more information, which can trigger a lengthy back and forth.

Some outsiders wonder whether journals will make good on promises to crack down. Publishers embracing gold open access—under which journals collect a fee from authors to make their papers immediately free to read when published—have a financial incentive to publish more, not fewer, papers. They have "a huge conflict of interest" regarding paper mills, says Jennifer Byrne of the University of Sydney, who has studied how paper mills have doctored cancer genetics data.

The "publish or perish" pressure that institutions put on scientists is also an obstacle. "We want to think about engaging with institutions on how to take away perhaps some of the [professional] incentives which can have these detrimental effects," van Rossum says. Such pressures can push clinicians without research experience to turn to paper mills, Sabel adds, which is why hospital affiliations can be a red flag.

Day sees some hope: Flagging journals suspected of being targeted by paper mills can quickly deter additional fraudulent submissions. He points to his analysis of journals that the Chinese Academy of Sciences (CAS) put on a public list because of suspicions they contained paper mill papers. His company's Papermill Alarm detector showed that before the CAS list came out, suspicious papers made up the majority of some journals' content; afterward, the proportion dropped to nearly zero within months. Journals could drive a similar crash by using automated detectors to flag suspicious manuscripts, nudging paper mills to take them elsewhere, Day says.

Some observers worry paper mill papers will merely migrate to lower impact journals with fewer resources to detect them. But if many journals act collectively, the viability of the entire paper mill industry could shrink.

It's not necessary to catch every fake paper, Day says. "It's about having practices which are resistant to their business model." ■



Pollution, warming waters, and introduced fish have reduced wild axolotl numbers.

CONSERVATION

In Mexico, the world's last wild axolotls face extinction

Scientists push to save the unusual salamander

By **Richard Stone**, *in Mexico City*

It's not yet noon on a Wednesday at Lake Xochimilco, a mosaic of ponds and canals to the south of this sprawling metropolis, but revelers on a brightly colored tourist boat have already broken out the beer and are whooping it up. On another boat, a mariachi band tunes up. Carlos Uriel Sumano Arias, paddling a flat-bottomed *chalupa* belonging to the National Autonomous University of Mexico (UNAM), slips into a quiet canal and lays up next to a *chinampa*, an artificial island for growing crops—a farming system invented by the Aztecs. "I'd like to show you an axolotl," he says.

A darling of regenerative medicine, the axolotl (*Ambystoma mexicanum*) can regrow severed limbs, its eyes, its frilly external gills, and even brain tissue. About 1 million are in captivity in labs and aquariums worldwide. Cute to some, grotesque to others, the salamander has also made its mark in pop culture, appearing as a character in online games. In Mexico, its image graces the 50 peso note. The axolotl is omnipresent—and at the same time, vanishing. The only remaining wild population is making its last stand here, in canals isolated from a former habitat that has become too hostile to sustain it.

Now, this haven is also in jeopardy. Entrepreneurs have been buying up *chinampas*, converting some to soccer fields and building pavilions on others for torch-lit feasts. "This gentrification phenomenon is a huge threat," as the businesses are less invested than the farmers in keeping the canals clean, says Luis Zambrano González, a UNAM biologist who has spent 20 years trying to stave off the axolotl's extinction.

The animal's best shot may be aggressive habitat restoration. Zambrano hopes to expand the number of canal refuges from the current 20, totaling 5 kilometers of waterways, to 200—enough, he says, to sustain a viable population. "A species is not a species if it's not in its environment," he argues. Others see salvation in breeding axolotls with closely related species, which could introduce genetic variation that might make wild or captive populations more resilient. "Other *Ambystoma* populations have recently swapped genes with the axolotl and are genetically similar," says biologist David Weisrock of the University of Kentucky. "All hope is not lost."

The 15 *Ambystoma* species found in central Mexico share a genetic heritage with the tiger salamander (*A. tigrinum*). Like most amphibians, it has gills and lives in the water as a larva before moving onto land as an air-breathing adult. But when axolotl ancestors colonized isolated lakes across the region, starting about 1 million years ago, ample food and few predators spurred a remarkable adaptation: Salamanders in some lakes began to spend their entire lives underwater, retaining juvenile features. "If there's no pressure to get out … you can become sexually mature without having to leave the pond," Weisrock says. The axolotl is one of four central Mexico species that rarely, if ever, metamorphose in the wild.

The Aztecs revered the axolotl, believing it to be the incarnation of Xolotl, a god of death and transformation. But after the Spanish founded Mexico City, they drained nearby lakes, shrinking axolotl habitat. The amphibian suffered another blow in the 1970s and '80s, when authorities introduced carp and tilapia into Xochimilco as food for the area's burgeoning population. Both species prey on axolotl eggs and young. Compounding the salamander's woes, the shallow waters have grown warmer and more polluted. A 1998 survey tallied 6000 axolotls per square kilometer. But the latest census, completed in 2015, estimated just 36 per square kilometer.

By now the salamander appears to have vanished from most of Xochimilco. Over the past 2 years, Alejandro Maeda-Obregón, a Ph.D. student at University College London (UCL), analyzed DNA from the lake's water, looking for signs of axolotls and other endangered species. Maeda-Obregón, working with UCL's Julia Day and Elizabeth Clare at York University, has so far failed to detect mitochondrial axolotl sequences in Xochimilco's touristed stretches. "That's a collapsed ecosystem," he says. They have now turned to the refuges, probing for axolotls known from past surveys to be living there.

The axolotl's close kinship with other *Ambystoma* salamanders could allow biologists to reinvigorate the species by breeding it with relatives. Scientists are investigating the source of this unusual genetic similarity, says UNAM geneticist Gabriela Parra Olea. She suspects that the plateau tiger salamander (*A. velasci*) may have been "the promoter of genetic uniformity." As a terrestrial adult, it may have crawled from lake to lake, mating with local salamanders.

Zambrano hopes such interventions won't be necessary for Xochimilco's axolotls. He and his team are working to create more havens, free of predatory fish, by closing off canals with nets or rock barriers. In one such refuge, the water is clear and cool. "Axolotls can do fine here," Sumano says. Nearby, in the shade, are two water tanks. He leans over one and points out an axolotl larva, a few centimeters long, paddling along the edge. In the other swim mature axolotls that are 10 times bigger. The team is waiting on permits to release the captives.

Zambrano hopes to "turn our gentrification problem into an opportunity" by raising money from developers to create refuges. He recently launched an "axolotl adoption" campaign to restore habitats and support farmers who monitor the refuges. "The next 5 years are critical," he says. For the salamander of the gods, time is running out. ■

BIOMEDICINE

Latest Alzheimer's antibody is 'not a miracle drug'

Results on Eli Lilly's donanemab highlight stark risk-benefit trade-offs

By **John Travis**

Is Alzheimer's disease about to become truly treatable? Some scientists raised the possibility last week after Eli Lilly and Co. announced positive results from its key phase 3 trial of donanemab, an antibody designed to prevent or reduce the accumulation of a protein called beta amyloid in the brain. It was only the second experimental drug to clearly stem the cognitive and functional decline of people with early signs of Alzheimer's. Following comparable results a year ago for a similar antibody, lecanemab, the announcement raises the prospect of widespread use of the drugs, expensive monoclonal antibodies that must be given by infusion.

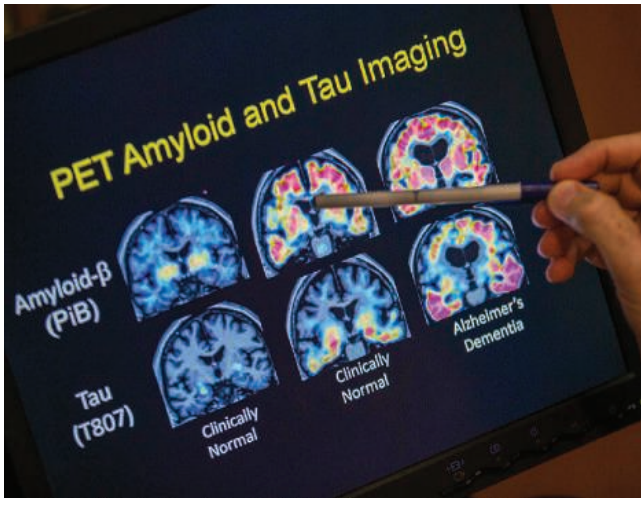
"This is monumental and aligns very closely to lecanemab as far as the amount of benefit," says Donna Wilcock, an Alzheimer's disease scientist at Indiana University. "In fact, it looks like donanemab may have greater benefit, although the trial designs do not allow direct comparison."

But the preliminary donanemab results, presented so far only in a company press release, also reveal a sobering risk of brain swelling and hemorrhaging. The side effects may be linked to three deaths in the clinical trial and echo hazards seen with lecanemab, developed by Eisai and Biogen. With the U.S. Food and Drug Administration considering full approval for lecanemab next month and Eli Lilly vowing to quickly submit donanemab for FDA review, patients, families, and caregivers may soon face difficult conversations about whether to risk immediate harm for the benefits of these therapies, which appear modest at best so far.

Donanemab is "not a miracle drug," cautions Sorbonne University neurologist Nicolas Villain, who helped test lecanemab and believes Eli Lilly's antibody may carry even higher risks. Still, many Alzheimer's researchers see the trial results for both drugs as vindication for the long-held but contested hypothesis that amyloid buildup in the brain drives the fatal neurodegenerative disorder. They hope the drugs' partial success inspires better, safer interventions to fight amyloid buildup.

Both donanemab and lecanemab bind to beta amyloid, which forms extracellular deposits known as plaques. Brain scans taken during Eli Lilly's phase 3 trial of more than 1700 people showed that donanemab cleared the protein effectively. In more than half of participants, amyloid plaques fell to a level Eli Lilly had set as a criterion for stopping the monthly antibody infusions.

More importantly, over 18 months, cognitive and physical functions declined more slowly in the treatment group than in the placebo group. The antibody slowed the disease's expected progression by 35% based on a scale Eli Lilly developed and by 36% on the more widely used Clinical Dementia Rating-Sum of Boxes, or CDR-SB, scale, which ranks dementia severity on an 18-point scale. The results echo those for lecanemab, which showed a 27% slowing after 18 months.



Eli Lilly has shown that an antibody helps clear amyloid protein from the brain, even in Alzheimer's patients with high levels of the tau protein.

PHOTO: EVAN VUCCI/AP

Eli Lilly also reported that 47% of the treated trial participants showed no worsening of symptoms based on the CDR-SB, compared with 29% among the placebo group. And on average treated individuals maintained a greater ability to perform key daily activities such as using the toilet independently and dressing themselves.

Eli Lilly says it plans to reveal more data at a meeting in July and in a forthcoming publication, but some outside researchers question how meaningful the clinical benefits are. The absolute difference between treatment and placebo groups on the CDR-SB was just under 0.7; some researchers argue it takes at least a full point on that scale to represent a clear difference in disease severity. But others say the slowing could be enough to make a difference for patients and caregivers.

Eli Lilly's trial also looked at another brain protein, tau, that many researchers have linked to neurodegeneration. Unlike beta amyloid, which forms plaques outside brain cells, tau malfunctions inside neurons, forming clumps known as tangles. Their presence tends to indicate more advanced disease. Eli Lilly expected that people with less tau would be more responsive to the antibody treatment, and they were, suggesting that early treatment may be critical. But it appeared to have some benefit even in those with high tau levels.

Other trial results are less heartening. Lecanemab has been linked to several deaths and cases of severe brain damage, as well as to brain swelling and hemorrhaging, and Eli Lilly's antibody comes with similar hazards. A form of brain swelling known as ARIA-E occurred in 24% of people who received it, with about 6% reporting symptoms. A more severe condition involving microhemorrhages, ARIA-H, affected 31.4% of antibody-treated people, versus 13.6% of the placebo groups. (Both forms of ARIA are known symptoms of Alzheimer's disease.) Eli Lilly says most ARIA cases were "mild to moderate" and stabilized after "appropriate management," but the company acknowledges the incidence of "serious" ARIA was 1.6%. It did not disclose details about the three deaths.

"Looks as dangerous as lecanemab," tweeted Robert Howard, a psychiatrist at University College London who has tested treatments for dementia.

Many questions about the drugs remain to be answered, such as whether the antibody benefits endure and whether even rich countries can afford the treatments. The weighing of hope and caution is just beginning. ■

With reporting by Charles Piller.



GIVING HEART

A new procedure for donating hearts and other organs is saving lives. But for some it challenges the definition of death

By **Jennifer Couzin-Frankel**

On a chilly holiday Monday in January 2020, a medical milestone passed largely unnoticed. In a New York City operating room, surgeons gently removed the heart from a 43-year-old man who had died and shuttled it steps away to a patient in desperate need of a new one.

More than 3500 people in the United States receive a new heart each year. But this case was different—the first of its kind in the country. "It took us 6 months to prepare," says Nader Moazami, surgical head of heart transplantation at New York University (NYU) Langone Health, where the operation took place. The run-up included oversight from an ethics board, education sessions with nurses and anesthesiologists, and lengthy conversations with the local organization that represents organ donor families. Physicians spent hours practicing in the hospital's cadaver lab, prepping for organ recovery from the donor. "We wanted to make sure that we controlled every aspect," Moazami says.

That's because this donor, unlike most, was not declared dead because of loss of brain function. He had been suffering from end-stage liver disease and was comatose and on a ventilator, with no hope of regaining consciousness—but his brain still showed activity. His family made the wrenching choice to remove life support. Following that decision, they expressed a wish to donate his organs, even agreeing to transfer him to NYU Langone Health before he died so his heart could be recovered afterward.

In individuals declared brain dead, organs can be recovered before life support is disconnected, as these people have already died; such machinery keeps organs oxygenated and healthy prior to transplant. But for this man the donation process would be al-



Tony Donatelli (right, bottom) was the first person ever to receive a heart, liver, and kidney through a new donation procedure. "I cannot tell you how lucky I am," he says.

PHOTO: SANDY HUFFAKER

tered: Life support had to be withdrawn for death to occur. His heart stopped, and his circulation with it.

As is customary regardless of whether organs will be donated, physicians waited 5 minutes to ensure that the heart didn't start beating again on its own. It did not, and the man was declared dead. The baton then passed to the organ recovery and transplant team. They clamped blood vessels running from the torso to the brain and reconnected his body to machines that circulated oxygenated blood, causing the heart to begin pumping again.

These two interventions—initiating a heartbeat after death is declared and taking steps to prevent blood flow to the brain—are at the core of a raging debate about the ethics of such donations. To some people, the approach risks disrupting the dying process; to others, it allows that process to continue as the family desires, while also honoring individual or family wishes for organ donation.

The debate touches on the definition of death, Moazami says. "When the heart stops, we say, 'time of death, 5:20 a.m.'" But, "The fact of the matter is, death is a process. Death is not a time point." Cells can take hours to die. Sophisticated machinery can induce a heartbeat hours after death, but does that make a person "alive"?

An expanding number of hospitals and organ procurement organizations (OPOs), which work with donor families, support this novel category of donations, and the number performed in the U.S. is growing. "I had about 3 months, tops," left to live, says Tony Donatelli, 41, who lives near San Diego with his wife and two young children, and who developed a rare disease that causes a dangerous buildup of protein in the body. On Valentine's Day 2022, he became the first person in the world known to receive a heart, liver, and kidney from a donor whose organs were perfused after circulatory death. Donatelli is back to surfing, woodworking, and wrestling on the floor with his sons. "I cannot tell you how lucky I am," he says.

Yet professional groups have expressed dueling views about the organ donation strategy, and a paper in press urges more research. Some countries are holding off on these organ donations, whereas others embrace them. One OPO says families who welcome donation do so without regard for the organ recovery technique, as such gifts can bring comfort after a terrible loss; another worries that without more research and greater attention to legal and ethical questions, there's a risk fewer people may volunteer to be organ donors. Meanwhile, surgeons say this category of donors could increase heart transplants by up to 30%, saving lives with organs that would otherwise go unused.

"There is definitely that initial reaction that there's something different" about this, says Anji Wall, an abdominal transplant surgeon and bioethicist at Baylor University Medical Center. Although Wall acknowledges the complexities, she supports such transplants and has performed them herself. "At the end of the day, the donor is dead," she says. "What you do does not make them alive again."

ORGAN TRANSPLANTATION has evolved and flourished from its first success in 1954, when a 23-year-old in Boston donated a kidney to his identical twin. In the years since, the number of transplants has surged, but demand invariably outstrips supply. In the U.S., which performs more transplants than any other country, about 104,000 people are awaiting a new organ, and, on average, 17 die each day before they get one. "We are a system that has always operated with scarcity," says Alexandra Glazier, an attorney and president and CEO of New England Donor Services. Her system is one of 56 OPOs, each covering a geographic region across the U.S., that coordinate organ donations by working with hospitals and donor families.

The transplant system relies on public trust and the generosity of these families at an excruciating and disorienting time. In 2013, Emily Stillman, a 19-year-old college student in Michigan, was rendered brain dead by a meningitis infection. When her mother, Alicia Stillman, was approached about donating Emily's organs, her first reaction was one of horror. "I said, 'Absolutely not, tell them to stay away.' … I remember screaming."

But she quickly had second thoughts, believing her daughter would have wanted organ donation; a call to the family's rabbi also helped. Emily's organs were donated to five people, and the family bonded with four of them. The heart recipient, a young physician in Ohio, named his baby girl after Emily. On the anniversary of Emily's death in February, the mothers of the heart and kidney recipients reached out to Alicia, "telling me, mother to mother, how grateful they are that they had these 10 years,"

she says. Gifting Emily's organs "was a huge part of our healing. It always gave us something positive to grab onto."

Until recently, virtually all organ donors in the U.S. were like Emily. Following a grievous injury or some other catastrophe, they were left brain dead—which is defined as lacking any brain function, including the ability to breathe on one's own. Their organs, however, can be protected by keeping donors on supportive machinery.

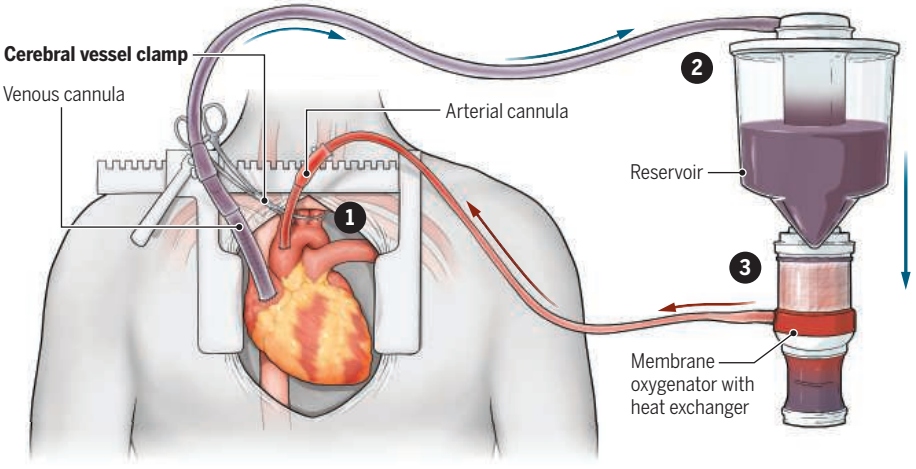
But in the 1990s, doctors grew interested in another potential category of donors: people who retained some brain activity after a serious illness or accident but who died when their circulation ceased—normally because, like the heart donor at NYU, their families had opted to withdraw life support when there was no hope of meaningful recovery. The lungs, liver, and kidneys, surgeons learned, could be recovered and function after transplant. This became known as "donation after circulatory death," or DCD donation. Initially uncommon, the number of DCD donors has soared; today, about one-quarter of the kidneys transplanted in the U.S. are from DCD donors.

The heart was another story. Circulatory death could severely injure the organ. To address the problem, companies experimented with machinery that would run blood through a heart after it was removed from the body and stimulate its electrical activity. In 2014, Australia was the first to test one such device, made by the company TransMedics, after circulatory death. Five years later, a TransMedics clinical trial began in the U.S. Regulators there approved the system for this purpose in 2022.

"It gave us access to hearts that no one else was using," says Ashish Shah, chief of cardiac transplantation at Vanderbilt University, who participated in the trial. But using the device costs between $65,000 and $85,000 each time. Recovering organs from DCD donors can also be logistically complex, as surgeons race to remove them before they succumb to a lack of oxygen. Sometimes, one organ is recovered but another can't be saved.

Heart of the matter

Typically, organs are recovered after brain death. A newer approach allows donations from people whose relatives choose to withdraw life support, following an unsurvivable illness or injury. After death is declared due to loss of circulation, machines are connected to oxygenate the blood and allow the heart to restart, while clamps prevent blood from reaching the brain. The circulation keeps organs healthy until they are recovered for transplant.



1 A clamp is placed across three arteries that supply blood to the brain, to focus perfusion on the organs being recovered and to avoid interfering with the dying process.

2 Deoxygenated blood travels from right atrium to an external reservoir.

3 Blood is warmed and oxygenated before returning to the body.

Shah, like Moazami, had been scrutinizing reports from colleagues in Europe and the United Kingdom about another kind of DCD donation. It entailed initiating oxygenated blood flow to the organs intended for transplant while they were still in the donor's body. For the heart, that meant starting it beating again after a declaration of death.

The strategy, called normothermic regional perfusion–DCD (NRP-DCD) and sometimes abbreviated to NRP, was yielding promising results. In 2020, a team at Royal Papworth Hospital in Cambridge, England, published outcomes on three categories of heart recipients: those who received a heart from a donor after brain death, those whose heart was from a DCD donor and placed on an external device, and those whose donor organ was recovered after NRP, with perfusion within the body. All 22 people who received NRP hearts were still alive 1 year later. For the group receiving hearts maintained on external machinery, the 1-year survival rate was 86%. For people receiving hearts from donors assessed as brain dead, 1-year survival was 89%. A March study reported comparable outcomes after 157 NRP-DCD heart transplants spanning several countries and 673 heart transplants from donors declared brain dead.

These weren't randomized trials, but nonetheless "the results were great" for NRP, says Stephen Large, a cardiothoracic surgeon at Royal Papworth. Large pioneered these heart transplants in the U.K. beginning in 2015, after years of deliberation by authorities. The inspiration, he recalls, came in 2006 when a family approached him after their 57-year-old wife and mother had suffered a devastating stroke. The relatives intended to remove life support and wished to donate her heart, but DCD heart donations weren't possible at the time. The family reached for the next best option, asking Large whether he wanted to study her heart instead. He did, testing NRP for the first time in a human being.

IN DOING SO, Large learned that the strategy allows a surgeon to, in effect, audition the heart. You could "see how the heart was performing" in the body after restarting perfusion, he says. Moazami, the first to adopt the technique in the U.S., had the same reaction: "I can look at the pressure the heart is generating, what the chambers are doing." In addition, surgeons believe that when oxygenated blood circulates through several organs at once, it can help them recover function lost during the dying process and their time without oxygen.

In September 2021, Moazami's team announced that its first eight NRP-DCD heart recipients were all still alive. Earlier that year, Vanderbilt, one of the biggest heart transplant centers in the world, had launched its own NRP program. "We found ourselves traveling all over the country getting hearts," Shah says. "There was a demand from the donor side—these families want these hearts donated. … Our job is to find a way to use" them. And use them he did. In 2022, Vanderbilt performed 40 heart transplants from NRP donors.

NRP technology is being used for other organs as well. Aleah Brubaker was a new liver transplant surgeon at the University of California, San Diego (UCSD), in the fall of 2021 when she was dispatched to get her first liver from an NRP donor. Immediately, the impact "was very evident to me," she says. Patients are "unquestionably" getting organs



Heather Santiago listens to her son Jordan's heart beat inside the person who received it. After Jordan died in a hit-and-run, his organs were donated to five people.

PHOTO: SOUTHWEST TRANSPLANT ALLIANCE

more quickly, including some who might die waiting—among them Donatelli, for whom Brubaker was on the transplant team.

UCSD has accepted livers from NRP donors in their late 60s, above the usual age cutoff, because perfusion inside the body helped doctors determine that the organs would be usable. Research on NRP kidney and liver recipients shows they have improved organ function and less chance of needing a second transplant than patients who get those organs via conventional DCD. "The results were much better," says Beatriz Domínguez-Gil, director general of the Spanish National Transplant Organization. Spain has used NRP for abdominal organs for many years and began NRP heart donations in 2020. Unlike the heart, the liver and kidneys can be sustained with blood flow just to the abdomen, and without inducing a heartbeat, which eases some ethical concerns.

Transplant programs are fiercely competitive, vying for the lowest wait times and highest survival rates. At the same time, the surgeons participating in NRP donations say they wouldn't touch them without full support from their institutions and confidence that they are ethical. The people who become NRP donors "are in this terrible state," Shah says, most often with a devastating brain injury. They have no prospect of meaningful recovery.

"We have to put ourselves back into the context that this family has already accepted their loved one will not live and wants us to go forward with donation," says Brad Adams, an attorney who is also president and CEO of the Southwest Transplant Alliance. His OPO oversaw seven NRP donations in the first quarter of this year compared with nine in all of 2022.

AS NRP DONATIONS ramp up in the U.S., some other countries are pumping the brakes. In the U.K., the first place use NRP for heart donation, they ground to a halt in 2019. Concern bubbled up at a meeting between U.K. and Canadian physicians about whether, despite the clamping of vessels to the brain before organ recovery, some blood could still reach it.

In an attempt to study this, Large and his colleagues examined three NRP donors, looking for blood in tiny arteries that thread up to large vessels feeding the brain. In one person, there was detectable blood flow in these vessels, estimated at 50 milliliters per minute, about 7% of the normal rate. Whether blood actually reached the brain wasn't tested. "There was a heated argument" among physicians, Large says, about how much blood, if any, could get to the brain, and its significance.

Following Large's observations, the donation and transplant communities paused NRP for heart donations in the U.K. The country continues to support NRP for liver and kidney donors, as vessels there are clamped lower in the abdomen and the chance of blood reaching beyond the torso is considered remote.

A knotty and mind-bending question is whether such flow would matter. "The clamping of the vessels is … a postmortem intervention," says Marat Slessarev, a specialist in critical care and organ donation at Western University in Canada. Like colleagues who work in intensive care units, he's comfortable with the standard of declaring death 5 minutes after the heart stops following a withdrawal of life support. A recent study that included 480 patients whose life support was withdrawn backs this up. Transient heart activity resumed spontaneously in 67, but the longest lag time was 4 minutes and 20 seconds, inves-

tigators reported in *The New England Journal of Medicine* in 2021.

But could blood flow to the brain after circulatory death still spark brain activity or function? Because dying is a process, "the best way to put it is, we don't know," Slessarev says. He wants to prove that circulatory death guarantees rapid brain death, which he suspects is the case, and that clamping guarantees zero blood flow to the brain.

Slessarev and colleagues began to address the first challenge in a pilot study of eight people after the withdrawal of life support. They found that brain activity actually ceased before the heart stopped beating—on average, 78 seconds earlier. "Blood pressure falls below a certain level, then the brain stops, then circulation stops," Slessarev says. "That's sort of the sequence of events." To see whether these results hold up, he's now co-leading a larger effort across Canada that aims to enroll about 90 people. His goal is to inform organ donation policies in his country, which does not permit NRP donations but is weighing them.

Slessarev is also co-leading a Canadian team preparing to probe where there's any blood flow to the brain after clamping. He's reassured by a January 2022 paper from a team in Denmark, showing that in pigs, 8 minutes without circulation followed by clamping prevented all blood flow and brain activity when the heart restarted. Animals that did not receive any clamping showed brain waves on monitors.

Others are conducting similar studies. Moazami recently hunted for cerebral blood flow with a transcranial doppler machine in two NRP heart donors. "We could not detect any," he says, and he plans to examine this in more cases.

In the U.K., a team at Royal Papworth will study NRP donors in the coming months, using a test called a CT angiogram to see whether any blood appears in cerebral arteries and returns through the veins. The latter could indicate perfusion through tissues. "I will be really surprised" if there is cerebral blood flow to any meaningful degree, says Antonio Rubino, an intensivist at the hospital who is leading the trial. But he still wants the work to be done.

FOR NOW, "the global ethics of this have not been resolved," Moazami says, and even in the U.S., controversy is erupting. "The definition of death is kind of broken," says Brendan Parent, an NYU bioethicist who helped his hospital and several others consider the ethics of NRP. Irreversible loss of either all circulation or all brain function qualifies as death in the U.S. But circulation can, in theory, be reinstituted by machinery even many hours after a death, Parent notes, rendering the circulatory death definition, practically speaking, meaningless.

As NRP donations increased in the U.S., debate emerged in spurts and crescendoed this spring. An early salvo came in April 2021, when the American College of Physicians, which represents internal medicine doctors, deemed NRP unethical in part because it can reestablish a heartbeat. Wall, Shah, and others published a response arguing that NRP meets ethical standards for consent and organ donation. On the legal front came a volley from Alexander Capron, a law professor at the University of Southern California, and Glazier of the New England OPO, arguing that NRP is inconsistent with U.S. legal standards because it involves restarting circulation whose permanent absence prompted the declaration of death in the first place. Adams, Parent, and others shot back that NRP is consistent with U.S. legal standards of death, because the technique is limited to perfusing organs and doesn't impact the determination of death.

Glazier, who says her OPO is one of a few that don't yet permit NRP for the organ recoveries they coordinate, emphasizes that her concerns are more about a "misalignment rather than a big violation." Can these potential donors return to what is considered meaningful life with NRP interventions? "Absolutely not," she says. But she still has significant concerns about the strategy.

> **"At the end of the day, the donor is dead. ... What you do does not make them alive again."**
>
> **Anji Wall**, Baylor University Medical Center

Glazier thinks death in the U.S. should be redefined as a permanent loss of brain function, period, with no separate definition for circulatory death. She joined authors from eight countries, including some, such as France and Spain, where NRP has long been practiced, on a paper in press in *Transplantation*. They urged all countries to adopt a brain-based definition of death, which could be determined by a permanent absence of circulation to the brain. This would ease concerns about NRP if studies like those planned in Canada and the U.K. confirm that clamping prevents any blood flow to the brain.

In March, the U.S. National Institutes of Health and the Organ Donation and Transplantation Alliance held separate meetings to discuss NRP. Interest in the topic was so great that the alliance was forced to find a larger venue. Elizabeth Pomfret, chief of transplant surgery at the University of Colorado School of Medicine, which recovers organs with NRP, welcomes the discussion. But Pomfret, who is president-elect of the American Society of Transplant Surgeons, worries about "confusion that's arisen around these questions of the permanence of death. ... This whole conversation is sort of reeling out of control."

Parent stresses that protecting the potential donor's interests is the highest priority. But he adds, "It's not as simple as just trying harder to save that person's life," because doing so can become futile and cause more suffering. For individuals who expressed a wish to be a donor and whose life cannot be saved, "what would be in their best interest is preserving their organs" for others who need them, he says. Parent is now developing a project to study public and donor family perceptions of NRP.

Alicia Stillman and her husband didn't have to make a decision about NRP donation for Emily. But if they had, she imagines that the distinction between circulatory and brain death would not have been so important to her. "Nothing was going to bring Emily back," she says. "There was nothing I could have done to save her." Donor families "should be at the head of the table" helping make these decisions, she believes.

Across the country in Texas, another family agrees. Andrew Santiago doesn't know what he and his wife Heather Santiago would have done in this situation, but he suggests these donations "be positioned as an optional deal," with the family making that call. His 25-year-old son Jordan was left brain dead almost 3 years ago after a hit-and-run. Jordan's organs, like Emily's, were donated to five people, and "that's what's keeping us going," his mother says. Both the Stillmans and the Santiagos launched foundations that include organ donor advocacy.

Early next month, the American Transplant Congress will convene in San Diego, about a 15-minute drive from Donatelli's home near the Pacific Ocean; he may be speaking there. Pomfret, meanwhile, is organizing a symposium for the conference's opening day to develop a formal position statement on NRP and lay plans for national data collection on NRP cases and universal standards for recovering those organs, such as clamping techniques. "We're showing the community that we're responsible," Pomfret says. "This isn't just a free-for-all."

UCSD Jacobs Medical Center, the site of the institute's transplant surgeries, is a short drive up the coast from the convention center where surgeons will gather. In one of the hospital's beds, another triple-organ transplant patient is currently recovering. His new heart, liver, and kidney were given by an NRP donor. ■



ANTHROPOLOGY

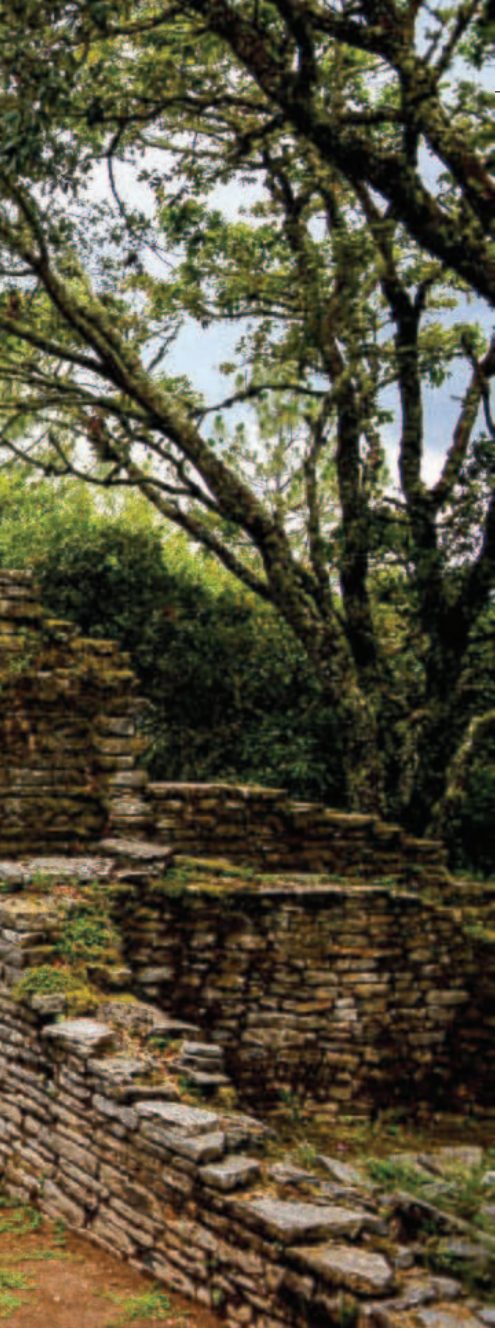
Paleogenomic study of the Mexican past

Ancient DNA analysis of ancestral Mexicans reveals a complex demographic history

By **Bastien Llamas** *and* **Xavier Roca-Rada**

The discovery that DNA survives in archaeological remains revolutionized archaeology, anthropology, and evolutionary biology, leading to the establishment of paleogenomics as a bona fide field of research (*1*). In the context of the Americas, paleogenomic researchers have used ancestral genomic information to reconstruct human history with exquisite detail. However, ethical concerns surrounding practices perceived as extractive and colonizing have sometimes overshadowed this endeavor (*2*). On page 598 of this issue, Villa-Islas *et al.* (*3*) set a standard for how paleogenomics can be conducted in an ethical and sustainable manner. They analyzed ancient DNA retrieved from 27 ancestral humans from eight archaeological sites in Mexico, providing insights into the demographic history of Indigenous populations in Central and North Mexico.

Although Mexico has a rich cultural heritage, paleogenomic studies have primarily considered it as a bridge between North and South America, acting as an obligate transit for early southward human migrations (*4*–*6*). Limited paleogenomic studies applying high-throughput sequencing have addressed Mexican-specific demographic questions, finding that ancient northwest



Paleogenomic approaches to study the demographic history of pre-Hispanic Mexico involve the analysis of ancient human DNA collected from archaeological sites. Shown are ruins in Toluquilla.

Mexico shared genetic affinities with neighboring populations further north, and that Central Mexican cities between 900 and 1500 CE were highly multi-ethnic (*7*, *8*). The study of Villa-Islas *et al.* not only fills a gap in terms of paleogenomic resources in the Americas, but also demonstrates how combining multidisciplinary evidence, including archaeology, anthropology, paleoclimatology, and paleogenomics, can advance our understanding of past human demography. The archaeological record had hinted at an expansion of Aridoamerican cultures (in northern Mexico, inhabited by hunter-gatherers) into Mesoamerican territory (in central and southern Mexico, home to some of the largest pre-Hispanic agriculture-based civilizations in the Americas) around 1100 CE, coinciding with a period of severe droughts. Genomic information can capture the genetic history and diversity of its population, providing insights into population-size changes, migrations, and interactions with other groups. Leveraging the information contained in the ancient genomes, Villa-Islas *et al.* found that droughts likely triggered large-scale migrations, leading to population replacement in the shifting border region of Aridoamerica and Mesoamerica. Interestingly, some Mesoamerican people stayed behind in the Sierra Gorda area, prompting questions about how these ancient cultures adapted to climate change. The results also echo the coincidence of climatic and demographic events in pre-Hispanic Central Andes (*9*) and call for in-depth paleoecological modeling to test causality.

One major challenge in paleogenomic studies is the scarcity of temporal and geographical data points that would enable a detailed reconstruction of the past. For instance, the archaeological record may lack entire regions or time periods, or cultural practices—such as cremation, and the use of cinnabar (mercury mineral ore used for red pigments)—or adverse environmental conditions may hinder DNA preservation. Moreover, in some cases, there may be no data available from relevant present-day populations for comparison. These issues are magnified in regions of the world such as Mexico, where the cultural heritage is incredibly diverse but conditions for ancient DNA preservation are suboptimal, the demographic history is potentially highly complex, and the recent colonial past has substantially altered the genetic makeup of some regions. Consequently, paleogenomicists use modeling to compensate for missing data and may resort to including unsampled "ghost" populations in their statistical inferences. In this regard, Villa-Islas *et al.* not only confirmed the existence of a previously reported ghost population ancestral to present-day Central Mexicans (e.g., Mixe) (*10*), but they also uncovered a second ghost population whose ancestry was last detected in an individual from Cañada de la Virgen, a Mesoamerican site occupied between 540 and 1050 CE.

Recently, scholars from the Global South argued that sustainable paleogenomic research should be based on principles of "knowledge sharing, capacity building, mutual respect, and equitable participation" (*11*, p.1). The study of Villa-Islas *et al.* is a remarkable example of this approach because it targets a Global South region and is led and conducted by (predominantly) local researchers. This is a considerable departure from collaboration between local scholars and laboratories from the Global North, which requires the export of samples and often leads to the relocation of local students and early-career researchers to the Global North for training. Although such collaborations can promote a two-way transfer of knowledge and expertise, they may not foster local development of paleogenomic research owing to economic constraints. This perpetuates global power and wealth imbalances and invariably results in a loss of both control of the traditional narrative and sovereignty over cultural material and data (*12*).

Therefore, developing capacity and capability in paleogenomic research in the countries where studies are conducted is crucial to empower researchers and participating communities. This approach promotes engagement with and involvement of local populations. The study by Villa-Islas *et al.* is an excellent example of high-quality and ethical paleogenomic research. Indeed, assigning a contemporary population as the sole direct descendant of ancestral remains is highly complex in Mexico owing to the sociopolitical and historical context, including the institutionalized imposition of the "mestizaje" rhetoric (which supports that all Mexicans have mixed European and Indigenous ancestry) that undermines the voice of Indigenous peoples, especially in genomic research (*13*). In this context, only Mexican researchers who are sensitive to the social reality of their country can implement best practices when engaging with the communities closest to the archaeological sites, especially given the challenges involved in consulting with Indigenous populations regarding the destructive analysis of ancestral humans in Mexico. ■

Australian Centre for Ancient DNA, School of Biological Sciences, University of Adelaide, Adelaide, SA, Australia. Email: bastien.llamas@adelaide.edu.au

REFERENCES AND NOTES

1. A. Curry, *Science* **378**, 12 (2022).
2. A. D. Cortez, D. A. Bolnick, G. Nicholas, J. Bardill, C. Colwell, *J. Soc. Archaeol.* **21**, 157 (2021).
3. V. Villa-Islas *et al.*, *Science* **380**, eadd6142 (2023).
4. C. L. Scheib *et al.*, *Science* **360**, 1024 (2018).
5. M. Raghavan *et al.*, *Science* **349**, aab3884 (2015).
6. J. C. Chatters *et al.*, *Science* **344**, 750 (2014).
7. A. Y. Morales-Arce *et al.*, *Archaeol. Anthropol. Sci.* **11**, 3459 (2019).
8. A. Y. Morales-Arce *et al.*, *Sci. Rep.* **7**, 18100 (2017).
9. L. Fehren-Schmitz *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 9443 (2014).
10. J. V. Moreno-Mayar *et al.*, *Science* **362**, eaav2621 (2018).
11. M. C. Ávila-Arcos *et al.*, *Front. Genet.* 10.3389/fgene.2022.880170 (2022).
12. B. Yáñez *et al.*, *Am. J. Biol. Anthropol.* **180**, 417 (2022).
13. P. Wade, C. López Beltrán, E. Restrepo, R. Ventura Santos, *Mestizo Genomics: Race Mixture, Nation, and Science in Latin America* (Duke Univ. Press, 2014).

10.1126/science.adh7902

ASTRONOMY

An origin scenario for a fast radio burst

Never-before-observed behavior of a repeating burst suggests its possible origin

By **Zsolt Paragi**

In just a tiny fraction of a second, a fast radio burst (FRB) in the distant cosmos can emit as much energy as the Sun does in several days. The sources of these intense bursts, observed only as radio waves, have been elusive. Studies on FRBs that repeat indicate that their progenitors reside in diverse environments within far-away galaxies. Some "repeaters" also appear to have an associated persistent source, likely a radio nebula, that may be powered by a magnetar (highly magnetized neutron star) or a black hole. The nature of these sources might provide information about the origin of FRBs. On page 599 of this issue, Anna-Thomas *et al.* (*1*) report a highly variable medium with turbulent magnetic fields surrounding a repeating FRB. This observation suggests that the pulses of radio wave emission may come from a compact object accompanied by a binary companion with strong stellar winds.

Magnetars have long been a prime suspect as FRB progenitors because of their similarities to pulsars, strong magnetic fields, and small emitting regions. However, their exact process of formation channel is not known, and there are still viable alternative FRB progenitors. Localizing the sources of FRBs within a host galaxy has been a powerful tool to constrain various progenitor formation models. Radio interferometric methods—especially those involving very long baselines—have shown that FRBs exist in different environments (*2*, *3*). Sometimes an FRB appears near a site of active star formation, but a repeating FRB was also found in a globular cluster (a luminous, nearly spherical collection of hundreds of thousands of stars) containing a very old stellar population (*4*). Precise localization could provide clues about the nature of the environment and constrain the channels of FRB progenitor formation but so far has not provided a conclusive answer. Turning attention to the properties of the bursts themselves may offer clues.

As radio waves travel through plasma in space, their properties change. Dispersion slows down the radio waves at longer wavelengths, and Faraday rotation (a magneto-optical phenomenon) changes their polarization properties. FRB pulses thus have a full history of plasma distribution in space imprinted on them and can provide information about the amount of baryonic matter (atomic matter) in between galaxies (*5*). However, this only works if the effects on radio wave propagation are dominated by this intergalactic matter. Substantial local dispersion and Faraday rotation effects from the immediate environment can constrain FRB progenitor formation models (*6*), but the benefits of this information go beyond just improving models. Understanding individual sources will ultimately allow the applications of FRBs to broader questions in cosmology (such as constraining the distribution of ionized gas in the universe) because we can minimize systematic errors in the data.

"Understanding individual sources will ultimately allow the applications of fast radio bursts to broader questions in cosmology..."

The repeating FRB 20190520B studied by Anna-Thomas *et al.* stands out in that its local environment contributes more to altering the burst emission than the intergalactic plasma and the interstellar gas in the Milky Way combined, resulting in a very large excess dispersion (*7*). It has an associated persistent radio source and resides in a dwarf galaxy—just like the first-ever discovered repeater FRB 20121102A (*8*). Therefore, these two FRBs are sometimes referred to as twins. FRB 20121102A already provided some insight into the nature of its source and its associated persistent radio source: The observed Faraday rotation in this system is particularly high and variable, which may point to an environment around a massive black hole, or in a young supernova remnant (*9*). However, neither interpretation gave a coherent picture of a formation channel for this FRB when considering the data. There was a hint about possible nonuniformities in the environment or in the emitting region, evidenced from the frequency dependence of the burst polarization properties (*10*).

Anna-Thomas *et al.* used some of the largest radio telescopes available to study the bursts seen from FRB 20190520B, and especially to see variations in propagation effects. What the authors found was rather surprising: The Faraday rotation was highly variable and also changing sign, which indicates a rather turbulent environment. A possible scenario is that there is a binary system in which the radio waves travel through the dense and variable stellar wind of a companion star.

Whether the above interpretation holds is still to be seen. The 16.3-day periodicity observed in FRB 20180916B may also support a binary interaction model for repeating FRBs (*11*). Compact objects such as neutron stars or black holes have certainly been found in binary systems, also with companions having massive stellar winds. Perhaps the best known such example is SS433, which is associated with a supernova-like radio nebula, W50. It was proposed that nebulae surrounding hyperaccreting x-ray binaries, similar to the SS433-W50 system, may be the precursors and persistent counterparts to FRBs (*12*).

It has also been proposed that the burst properties are not solely determined by the nature of the progenitor and in some cases can be dominated by interactions with the external medium (*13*). In any case, understanding the effect of the local environment on propagation effects will greatly increase the potentials of using FRBs for precision cosmology. Future monitoring of the time evolution of burst properties, as well as high-spatial-resolution studies of nebulae associated with FRB sources, may provide further clues. These studies will also reveal whether there is only a single FRB population, or if there are different FRB progenitors with various formation channels. ■

Joint Institute for VLBI ERIC (JIVE), 7991 PD Dwingeloo, Netherlands. Email: zparagi@jive.eu

REFERENCES AND NOTES

1. R. Anna-Thomas *et al.*, *Science* **380**, 599 (2023).
2. K. Bannister *et al.*, *Science* **365**, 565 (2019).
3. B. Marcote *et al.*, *Nature* **577**, 190 (2020).
4. F. Kirsten *et al.*, *Nature* **602**, 585 (2022).
5. J.-P. Macquart *et al.*, *Nature* **581**, 391 (2020).
6. D. Michilli *et al.*, *Nature* **553**, 182 (2018).
7. C.-H. Niu *et al.*, *Nature* **606**, 873 (2022).
8. S. Chatterjee *et al.*, *Nature* **541**, 58 (2017).
9. G. H. Hilmarsson *et al.*, *Astrophys. J. Lett.* **908**, L10 (2021).
10. A. Plavin *et al.*, *MNRAS* **511**, 6033 (2022).
11. CHIME/FRB Collaboration, *Nature* **582**, 351 (2020).
12. N. Sridhar, B. D. Metzger, *Astrophys. J. Lett.* **937**, 5 (2022).
13. B. Zhang, *Astrophys. J. Lett.* **836**, L32 (2017).

ACKNOWLEDGMENTS

I thank S. Bhandari and J. W. T. Hessels for helpful discussions.

10.1126/science.adh8099

ECOLOGY

Abandoning land transforms biodiversity

Land abandonment is critical when assessing global biodiversity and conservation

By **Gergana N. Daskalova**[1]
and **Johannes Kamp**[2]

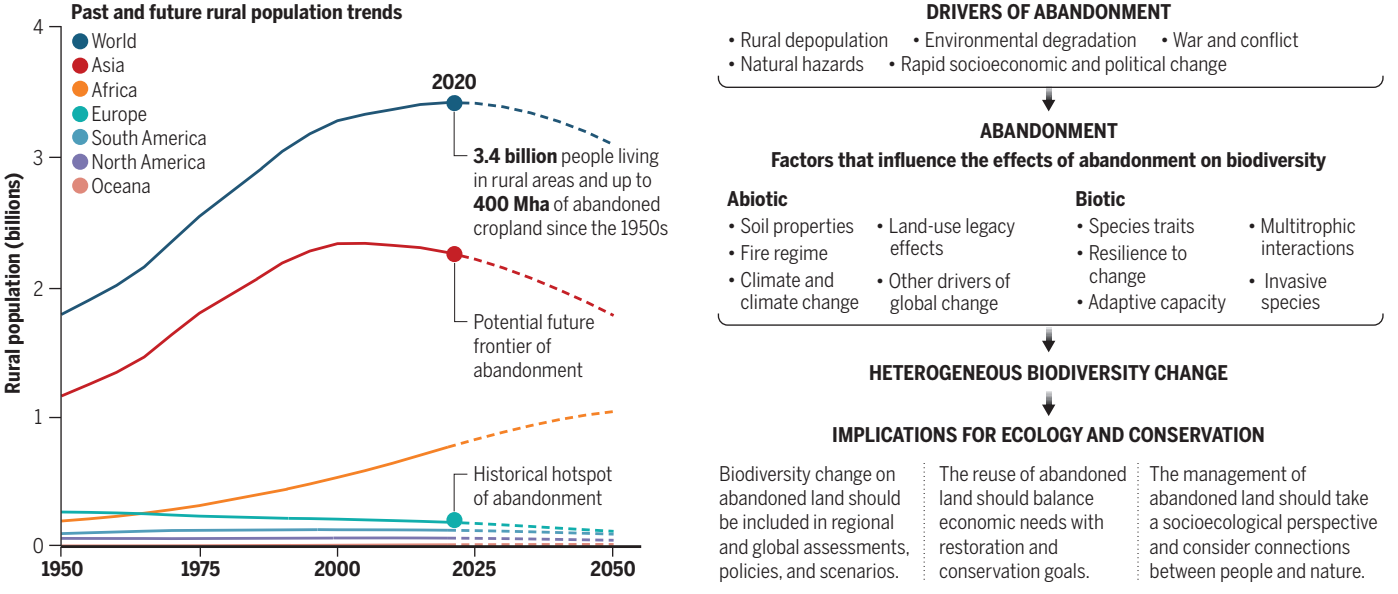
Land abandonment and rural depopulation are accelerating globally. In less than 50 years, the proportion of the human population living in rural areas has decreased by ~25% (*1*). Abandonment takes many shapes, and no single definition has been accepted. Most often, land abandonment refers to a termination of agricultural activities for at least 5 years (to differentiate it from fallow land) and is quantified at the crop-field scale (*2*, *3*). Other types of abandonment have received less attention, such as abandoned pastures, forestry areas, mines, factories, and entire human settlements. Here, "land abandonment" broadly captures the end of human activities. Since the 1950s, abandoned land has accumulated to up to 400 million ha globally (*3*), an area roughly half the size of Australia. Given this scale, there is an urgent need to develop a vision of how to achieve balanced benefits for biodiversity conservation, ecosystem services, and people's livelihoods on abandoned land.

The exact amount of abandoned land is unknown owing to the dynamic nature of abandonment and recultivation, variation in abandonment definitions, and a lack of data. Abandoned land is more widespread in the Northern Hemisphere, and ~117 million ha falls within the former Soviet Union (*4*). Although variable in magnitude geographically, abandonment is a globally important process caused by social, economic, and environmental factors (see the figure). Rural depopulation, the pursuit of urban job opportunities and services, and declining subsistence farming have been driving abandonment since the start of the Industrial Revolution in the 1750s. Rapid socioeconomic and political change, such as the collapse of the Soviet Union in 1991, triggered especially high rates of land abandonment. War and conflict further exacerbate abandonment, and the Russian invasion of Ukraine has already created new abandonment hotspots (*5*). Land-use intensification in lowlands can drive abandonment of marginal land in mountains. Environmental degradation, such as soil erosion and pollution, reduces profits from agricultural land, making abandonment more likely. Finally, natural hazards such as volcano eruptions, earthquakes, and floods can lead to sudden abandonment. The drivers of land abandonment are projected to increase in the future (*1*), highlighting the urgency to understand how abandonment is reshaping the natural world.

The effects of abandonment on biodiversity vary. Direct, positive effects are more likely to occur when intensively managed and biodiversity-poor croplands are abandoned (*6*, *7*). Abandonment benefits forest and mixed- or closed-habitat birds and invertebrates (*6*, *7*). Coupled rural depopulation and abandonment can lead to passive rewilding, particularly for large herbivores and carnivores (*6*). However, complete recovery of species communities on former arable land is often not achieved passively, even after a century (*8*), and not all abandoned lands return to a previous reference state (*9*). Negative or mixed effects of abandonment on biodiversity prevail in cultural landscapes and close-knit socioecological systems that are characterized by low-intensity, often subsistence farming (*10*, *11*). The long coevolutionary history of these landscapes and their people, found in, for example, Eastern Europe, Japan, and parts of the tropics, has created high habitat heterogeneity that can

Abandonment leads to heterogeneous biodiversity trajectories

Land abandonment is increasing as rural populations decrease. Rural population trends over time and projections for the future (dashed lines) are based on United Nations Population Division statistics. Abandonment is driven by a range of factors, and it influences biodiversity in heterogeneous ways, which have implications for ecology and conservation.



GRAPHIC: N. CARY/*SCIENCE* BASED ON G. N. DASKALOVA AND J. KAMP

disappear after abandonment and lead to the loss of locally rare species and to biodiversity homogenization (*10*).

Especially threatened by abandonment is the biodiversity of habitats that are maintained by anthropogenic disturbances that now often replace natural disturbance regimes, such as in temperate grasslands, wood pastures, rice paddies, cocoa agroforestry, and coppice forests where open-habitat specialists become replaced by generalists and closed-forest species (*6*). Abandonment can also result in plant invasions, followed by vertebrate declines (*7*). When some species receive more attention than others, the positive (or negative) effects of abandonment for species in the spotlight may skew the overall perspective of the value of abandonment for biodiversity and mislead policies (*7*). Amid evidence for both gains and losses in species richness and abundance after abandonment, a consistent finding is that of shifts in community composition, which sometime result in new ecosystems that consist of species that have not previously co-occurred in a given location (*7*, *12*).

The heterogeneous picture of both gains and losses in biodiversity after abandonment (*6*, *11*) highlights that abandonment can be both a threat and an opportunity for biodiversity. The impacts of land abandonment depend on the type and intensity of previous land use, climate, soil types, and species' functional traits (*6*, *7*). In Japan, cold-adapted butterflies were more vulnerable to abandonment and human depopulation than warm-adapted ones (*11*). Shifts in fire regimes after the end of agricultural land use can trigger cascades of ecosystem changes that ultimately affect biodiversity, ecosystem services, and the quality of human life through damages to infrastructure and reduced access to services (*13*). On the Eurasian steppes, pasture abandonment triggered by the breakup of the Soviet Union in 1991 and a subsequent accumulation of biomass led to an increase in wildfires, altering plant and animal communities (*13*). Similar processes have occurred in ecosystems of the Mediterranean, a global biodiversity hotspot. Beyond biodiversity, land abandonment affects ecosystem functions and services. Positive effects include the accumulation of soil and aboveground carbon, which thereby partly offset climate change–relevant carbon dioxide emissions (*14*), but there are also negative effects, such as potential soil erosion. The feedback among abandonment, biodiversity, ecosystem services, and people's livelihoods is yet to be investigated at regional to global scales.

The direction and magnitude of the global net effect of abandonment on biodiversity remains unquantified owing to several challenges (*7*). High-resolution maps on the extent, timing, magnitude, and persistence of land abandonment are only available for some hotspots (*4*, *15*) but are not available globally and not for grassland abandonment resulting from destocking or for abandonment in the marine realm, for example, out-of-commission fisheries. High-quality, global-scale remote-sensing imagery is only available from the 1990s onward, which misses big waves of abandonment such as those that occurred during the First and Second World Wars. The regional and national statistics that are available on temporal trends in land use and land abandonment have not yet been fully harnessed and standardized on a global scale. Additionally, few attempts have been made to integrate biodiversity data collected on abandoned land, compiled from global databases such as BioTIME, and obtained from land abandonment maps or to incorporate abandonment in biodiversity scenarios.

"Land abandonment is not a purely social or ecological phenomenon; it is the product of complex, age-old interactions between people and nature."

Land abandonment influences biodiversity on both short- and long-term timescales (*8*). The legacy effects of abandonment on biodiversity, which include extinction debts and colonization credits and take decades to centuries to materialize, are much less studied than immediate, direct effects (*7*, *8*). The same is true for interactions of land abandonment with biotic and abiotic drivers of biodiversity change, such as climate change, nutrient deposition, pollution, and socioeconomic change. A large area of uncertainty is also the self-restoration potential of abandoned land and how it varies by biomes and land-use legacies (*7*, *8*). Lastly, it is unknown what degree of management is needed to ensure that abandoned areas return to biodiversity-rich systems or the original habitats that existed before cultivation, rather than remaining in species-poor "lock-in" stages of a new ecosystem.

Increases in biodiversity after habitat restoration and rewilding on abandoned land (*12*) might be reversed quickly by the recultivation of abandoned land or the repurposing of abandoned land for other anthropogenic uses, which often occurs within less than 15 years (*15*). Abandonment might similarly be less likely to persist in war zones or areas that have experienced natural disasters, after which people and agriculture might eventually return. With increasing global pressure on land, abandoned land is earmarked for large-scale bioenergy, wind energy, and solar energy production (*2*, *3*). Land abandonment can also be followed by industrial activities such as mining owing to cheap land prices and lack of opposition because such activities would largely happen out of sight. Future models and scenarios that aim to predict the net balance of positive versus negative effects of abandonment on biodiversity need to account for variation in the likelihood of abandonment to persist. It is important to quantify the trade-offs between the recultivation of abandoned land for food, fiber, and energy and the contributions that these lands could make toward developing and implementing policies for restoring biodiversity and maintaining ecosystem services.

The recent 15th meeting of the Conference of the Parties to the United Nations Convention on Biological Diversity (COP15) in Montreal in 2022 set new global goals for biodiversity, but land abandonment did not have a prominent place on the agenda and its role in meeting conservation targets remains unexplored. A global synthesis of the conservation potential of abandoned land is still missing, and there are few systematic reviews and meta-analyses. To build toward such work, more local-scale studies and improved remote-sensing products are needed, particularly from underrepresented regions such as Eastern Europe, East Asia, and West and Central Africa. Establishing long-term ecological monitoring sites on diverse types of abandoned land can help to identify immediate and lagged biodiversity responses. Once past and ongoing land abandonment is better understood, better forecasts of where abandonment will occur in the future are possible.

There is an opportunity to integrate abandonment and biodiversity in an interdisciplinary socioecological perspective that combines the conservation of both biodiversity and cultural landscapes. Where passive restoration can result in biodiversity gains and species composition recovery, protection is needed for these newly rewilded ecosystems before they are recultivated or targeted for alternative uses. These abandoned lands could be included in protected area networks, such as Natura 2000, or managed after national and international restoration or rewilding strategies (*12*). Where aban-

[1]Biodiversity, Ecology, and Conservation Group, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria. [2]Department of Conservation Biology, University of Göttingen, Göttingen, Germany. Email: gndaskalova@gmail.com

donment threatens biodiversity, improved land stewardship should be encouraged, for example, through subsidies for traditional farming and policies that improve the quality of rural life. The feedback between abandonment, loss of culture and tradition, and ecological change has received little research attention but could be the missing piece in designing landscapes that improve both human quality of life and biodiversity conservation. Conservation and management increasingly recognize the value of working with local people and Indigenous communities. Land abandonment poses a challenge to such collaborations because abandonment is often paralleled with depopulation, which brings a decline in traditional subsistence farming and threatens to erase traditional knowledge. Historic knowledge of ecosystems and biodiversity is key for establishing baselines to assess the effectiveness of conservation measures and to track progress toward conservation goals. How can land be protected when its stewards are gone?

Nearly 100 years ago, abandoned fields helped develop key ecological theory on succession and the transition from open fields to dense canopy forests. Today, it is important and timely to look at abandoned fields once again but see the full picture—the migrating rural populations, the depopulating villages, and the deserted human infrastructure and crop fields, as well as the changing biodiversity and ecosystems. Land abandonment is not a purely social or ecological phenomenon; it is the product of complex, age-old interactions between people and nature. ■

REFERENCES AND NOTES

1. United Nations, Department of Economic and Social Affairs, Population Division, "World urbanization prospects: The 2018 revision" (ST/ESA/SER.A/420, United Nations, 2019).
2. P. Potapov *et al.*, *Nat. Food* **3**, 19 (2022).
3. C. Field, J. Campbell, D. Lobell, *Trends Ecol. Evol.* **23**, 65 (2008).
4. M. Lesiv *et al.*, *Sci. Data* **5**, 180056 (2018).
5. Y. Ma *et al.*, *Land* **11**, 1810 (2022).
6. J. Kamp *et al.*, *Agric. Ecosyst. Environ.* **268**, 61 (2018).
7. C. Queiroz, R. Beilin, C. Folke, R. Lindborg, *Front. Ecol. Environ.* **12**, 288 (2014).
8. F. Isbell, D. Tilman, P. B. Reich, A. T. Clark, *Nat. Ecol. Evol.* **3**, 1533 (2019).
9. P. H. S. Brancalion *et al.*, *Sci. Adv.* **5**, eaav3223 (2019).
10. J. Fischer, T. Hartel, T. Kuemmerle, *Conserv. Lett.* **5**, 167 (2012).
11. N. Sugimoto *et al.*, *Proc. R. Soc. London Ser. B* **289**, 20212222 (2022).
12. A. Perino *et al.*, *Science* **364**, eaav5570 (2019).
13. M. Freitag *et al.*, *Glob. Change Biol.* **27**, 388 (2021).
14. M. Thibault, E. Thiffault, Y. Bergeron, R. Ouimet, S. Tremblay, *Plant Soil* **475**, 605 (2022).
15. C. L. Crawford, H. Yin, V. C. Radeloff, D. S. Wilcove, *Sci. Adv.* **8**, eabm8999 (2022).

ACKNOWLEDGMENTS

G.N.D. is funded by a Schmidt Science Fellowship. The authors thank I. Myers-Smith and V. Radeloff for helpful comments and C. Crawford and an anonymous reviewer for their feedback, which greatly improved this piece.

10.1126/science.adf1099



The breath-holding strategy of the scalloped hammerhead shark (*Sphyrna lewini*) could be widespread among other deep-diving sharks and fishes.

PHOTO: ADRIANA BASQUES

PHYSIOLOGY

Free-diving sharks

Hammerhead sharks hold their breath when diving to regulate body temperature

By **Mark Meekan**[1] *and* **Adrian Gleiss**[2]

Regular cycles of descent to and ascent from mesopelagic depths (200 to 1000 m) of open oceans are a feature of the ecology of many large epipelagic fishes, including tunas, billfishes, sharks, and rays (*1*). Because these vertical movements typically begin and end in shallow waters, they are commonly referred to as dives, even though these fishes are not compelled to return to the ocean's surface to breathe, unlike marine mammals, reptiles, and birds (*2*). But the parallels between these lung- and gill-breathing animals are more than just a coincidence in vertical movements. On page 651 of this issue, Royer *et al.* (*3*) report that, like air breathers, hammerhead sharks (*Sphryna lewini*) are also "breath holding" during dives. This suggests that oxygen stores, rather than just temperature, could determine dive durations in some epipelagic fishes.

Why these sharks display this remarkable behavior is explained by the combination of ecological opportunities and physiological challenges invoked by diving into cold, deep water. During the day, enormous numbers of small fishes (mostly lanternfishes; Myctophidae) and invertebrates gather to form the deep scattering layer in depths of 200 to 500 m (*4*). The layer is accompanied by predators, such as squid and larger fishes, that inhabit even deeper waters (500 to 1000 m). To access these food sources, epipelagic fishes must regularly descend to these depths, which presents a physiological challenge for species whose primary habitats are relatively warm surface waters.

Although they function to extract oxygen, fish gills are also very efficient at transferring heat, with the consequence that deep, cold water has the potential to rapidly cool the blood and vital organs (*5*). Epipelagic fishes have evolved a variety of physiological, morphological, and behavioral solutions to cope with this problem, which involve adaptations that minimize heat loss through either conductive or convective means. At one end of the scale, slow-swimming whale sharks (*Rhincodon typus*), which are planktivorous filter feeders, grow to immense sizes and have a body plan that reduces conduction of heat when they feed at depth (gigantothermy) (*6*, *7*). At the other end, fast-swimming tunas and lamnid sharks are regional endotherms feeding on other fishes and squid, with heat exchang-

[1]Australian Institute of Marine Science, University of Western Australia, Oceans Institute, MO96, Crawley, WA, Australia. [2]Centre for Sustainable Aquatic Ecosystems, Harry Butler Institute, Murdoch University, Murdoch, WA, Australia. Email: m.meekan@aims.gov.au

donment threatens biodiversity, improved land stewardship should be encouraged, for example, through subsidies for traditional farming and policies that improve the quality of rural life. The feedback between abandonment, loss of culture and tradition, and ecological change has received little research attention but could be the missing piece in designing landscapes that improve both human quality of life and biodiversity conservation. Conservation and management increasingly recognize the value of working with local people and Indigenous communities. Land abandonment poses a challenge to such collaborations because abandonment is often paralleled with depopulation, which brings a decline in traditional subsistence farming and threatens to erase traditional knowledge. Historic knowledge of ecosystems and biodiversity is key for establishing baselines to assess the effectiveness of conservation measures and to track progress toward conservation goals. How can land be protected when its stewards are gone?

Nearly 100 years ago, abandoned fields helped develop key ecological theory on succession and the transition from open fields to dense canopy forests. Today, it is important and timely to look at abandoned fields once again but see the full picture—the migrating rural populations, the depopulating villages, and the deserted human infrastructure and crop fields, as well as the changing biodiversity and ecosystems. Land abandonment is not a purely social or ecological phenomenon; it is the product of complex, age-old interactions between people and nature. ■

REFERENCES AND NOTES

1. United Nations, Department of Economic and Social Affairs, Population Division, "World urbanization prospects: The 2018 revision" (ST/ESA/SER.A/420, United Nations, 2019).
2. P. Potapov *et al.*, *Nat. Food* **3**, 19 (2022).
3. C. Field, J. Campbell, D. Lobell, *Trends Ecol. Evol.* **23**, 65 (2008).
4. M. Lesiv *et al.*, *Sci. Data* **5**, 180056 (2018).
5. Y. Ma *et al.*, *Land* **11**, 1810 (2022).
6. J. Kamp *et al.*, *Agric. Ecosyst. Environ.* **268**, 61 (2018).
7. C. Queiroz, R. Beilin, C. Folke, R. Lindborg, *Front. Ecol. Environ.* **12**, 288 (2014).
8. F. Isbell, D. Tilman, P. B. Reich, A. T. Clark, *Nat. Ecol. Evol.* **3**, 1533 (2019).
9. P. H. S. Brancalion *et al.*, *Sci. Adv.* **5**, eaav3223 (2019).
10. J. Fischer, T. Hartel, T. Kuemmerle, *Conserv. Lett.* **5**, 167 (2012).
11. N. Sugimoto *et al.*, *Proc. R. Soc. London Ser. B* **289**, 20212222 (2022).
12. A. Perino *et al.*, *Science* **364**, eaav5570 (2019).
13. M. Freitag *et al.*, *Glob. Change Biol.* **27**, 388 (2021).
14. M. Thibault, E. Thiffault, Y. Bergeron, R. Ouimet, S. Tremblay, *Plant Soil* **475**, 605 (2022).
15. C. L. Crawford, H. Yin, V. C. Radeloff, D. S. Wilcove, *Sci. Adv.* **8**, eabm8999 (2022).

ACKNOWLEDGMENTS

G.N.D. is funded by a Schmidt Science Fellowship. The authors thank I. Myers-Smith and V. Radeloff for helpful comments and C. Crawford and an anonymous reviewer for their feedback, which greatly improved this piece.

10.1126/science.adf1099

PHOTO: ADRIANA BASQUES



The breath-holding strategy of the scalloped hammerhead shark (*Sphyrna lewini*) could be widespread among other deep-diving sharks and fishes.

PHYSIOLOGY

Free-diving sharks

Hammerhead sharks hold their breath when diving to regulate body temperature

By **Mark Meekan**[1] *and* **Adrian Gleiss**[2]

Regular cycles of descent to and ascent from mesopelagic depths (200 to 1000 m) of open oceans are a feature of the ecology of many large epipelagic fishes, including tunas, billfishes, sharks, and rays (*1*). Because these vertical movements typically begin and end in shallow waters, they are commonly referred to as dives, even though these fishes are not compelled to return to the ocean's surface to breathe, unlike marine mammals, reptiles, and birds (*2*). But the parallels between these lung- and gill-breathing animals are more than just a coincidence in vertical movements. On page 651 of this issue, Royer *et al.* (*3*) report that, like air breathers, hammerhead sharks (*Sphryna lewini*) are also "breath holding" during dives. This suggests that oxygen stores, rather than just temperature, could determine dive durations in some epipelagic fishes.

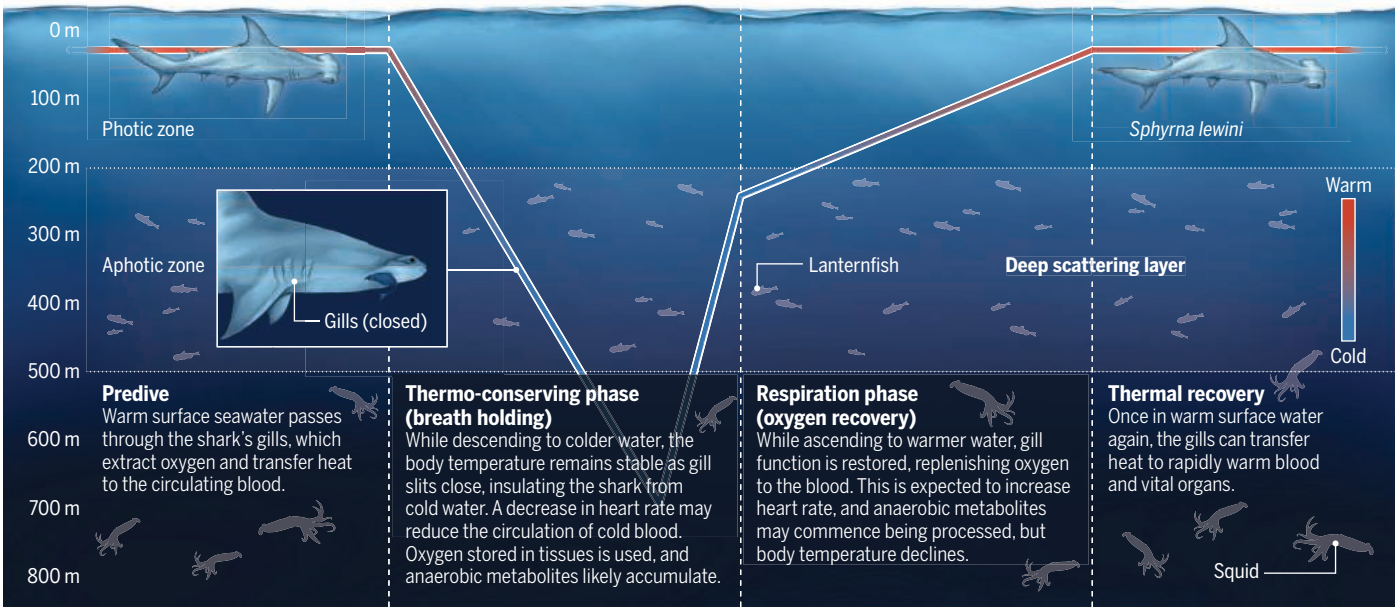
Why these sharks display this remarkable behavior is explained by the combination of ecological opportunities and physiological challenges invoked by diving into cold, deep water. During the day, enormous numbers of small fishes (mostly lanternfishes; Myctophidae) and invertebrates gather to form the deep scattering layer in depths of 200 to 500 m (*4*). The layer is accompanied by predators, such as squid and larger fishes, that inhabit even deeper waters (500 to 1000 m). To access these food sources, epipelagic fishes must regularly descend to these depths, which presents a physiological challenge for species whose primary habitats are relatively warm surface waters.

Although they function to extract oxygen, fish gills are also very efficient at transferring heat, with the consequence that deep, cold water has the potential to rapidly cool the blood and vital organs (*5*). Epipelagic fishes have evolved a variety of physiological, morphological, and behavioral solutions to cope with this problem, which involve adaptations that minimize heat loss through either conductive or convective means. At one end of the scale, slow-swimming whale sharks (*Rhincodon typus*), which are planktivorous filter feeders, grow to immense sizes and have a body plan that reduces conduction of heat when they feed at depth (gigantothermy) (*6*, *7*). At the other end, fast-swimming tunas and lamnid sharks are regional endotherms feeding on other fishes and squid, with heat exchang-

[1]Australian Institute of Marine Science, University of Western Australia, Oceans Institute, MO96, Crawley, WA, Australia. [2]Centre for Sustainable Aquatic Ecosystems, Harry Butler Institute, Murdoch University, Murdoch, WA, Australia. Email: m.meekan@aims.gov.au

Deep dive duration

A hammerhead shark's dive into cold, deep water may be limited not only by the temperature but also by the amount of oxygen stored in its tissues before diving. This is because the shark holds its breath during the descent as a thermoregulation strategy.



ers in the circulatory system that capture heat produced by the swimming muscles, thus reducing convective heat loss (*8*). These strategies can be enhanced through behaviors such as occupying warm surface waters before and after dives to lower energy costs and speed recovery from entry into cold water (*1*).

Using internal tags to monitor physiology, Royer *et al.* show that behavioral thermoregulation in sharks can also include breath holds. When diving to depths of 800 m and water temperatures of only 5°C, hammerhead sharks can substantially delay the onset of reduction in body temperature. Although the exact mechanism responsible for the reduction of convective heat loss is unknown, Royer *et al.* suggest that breath holding, where the shark closes its mouth and/or gill slits, effectively insulates the gills from cold water (see the figure). Other mechanisms could also reduce cooling, such as extreme bradycardia (lowered heart rate). Nonetheless, the only plausible explanation for temperature stability at depth is an interruption in the circulation of newly oxygenated, cold blood.

Breath holding may be an ingenious solution to the problem of heat loss, but it requires cessation of oxygen delivery to tissues. Paradoxically, hammerhead sharks are active foragers during deep dives, implying a high demand for oxygen that could be fueled by stores in the blood and muscle and/or by incurring an oxygen debt that must be repaid once sharks return to the ocean surface. Consequently, the duration of deep dives may not be determined exclusively by temperature but also by oxygen stores and the ability to withstand the accumulation of anaerobic metabolites. If correct, this would be a radical departure from current thinking about the limits to diving by epipelagic fishes in the open ocean (*1*).

Given the effectiveness of breath holding and the strong selective forces that shape behavioral and physiological thermoregulation, this strategy could be widespread among other epipelagic sharks and teleost fishes. Testing this hypothesis will require tags that sense internal physiology across a much wider suite of epipelagic fishes. Archival tags that simultaneously measure internal and external temperatures are an important research tool but, until recently, have mostly been deployed on tunas (*9*, *10*). Investigating physiological thermoregulation will also require advances in "physiologging" tags (*11*) that can be deployed internally or can sense across the skin boundary to monitor the cardiorespiratory responses of fishes to breath holding. The wide-ranging nature of epipelagic fishes makes it extraordinarily challenging to deploy and recover such tags.

Across the world's oceans, global warming is driving the decline of dissolved oxygen and an expansion of vast oxygen minimum zones at depths of 200 to 1000 m (*12*). Although these areas of very low oxygen saturation present a major threat to most macrofauna, the ability of hammerhead sharks to cease respiration while foraging at depth could mean that this species is more likely to cope with such environments. Indeed, this behavior could explain observations of these animals occupying the extremely hypoxic oxygen minimum layer in the Gulf of California (*13*). An ongoing technological revolution in tagging, of which the study of Royer *et al.* is a part, is likely to reveal other insights into the physiology and ecology of sharks that may explain the extraordinary persistence of these animals across 400 million years of changing ocean environments. ■

REFERENCES AND NOTES

1. S. Andrzejaczek, A. C. Gleiss, C. B. Pattiaratchi, M. G. Meekan, *Rev. Fish Biol. Fisheries* **29**, 335 (2019).
2. Y. Y. Watanabe, N. L. Payne, J. M. Semmens, A. Fox, C. Huveneers, *J. Exp. Biol.* **222**, jeb185603 (2019).
3. M. Royer *et al.*, *Science* **380**, 651 (2023).
4. G. C. Hays, in *Migrations and Dispersal of Marine Organisms*, M. B. Jones *et al.*, Eds. (Springer, 2003), pp. 163–170.
5. F. G. Carey, J. M. Teal, J. W. Kanwisher, K. D. Lawson, J. S. Beckett, *Am. Zool.* **11**, 137 (1971).
6. M. G. Meekan, L. A. Fuiman, R. Davis, Y. Berger, M. Thums, *Front. Mar. Sci.* **2**, 64 (2015).
7. I. Nakamura, R. Matsumoto, K. Sato, *J. Exp. Biol.* **223**, jeb210286 (2020).
8. K. A. Dickson, J. B. Graham, *Physiol. Biochem. Zool.* **77**, 998 (2004).
9. B. A. Block *et al.*, *Science* **293**, 1310 (2001).
10. B. A. Block *et al.*, in *Electronic Tagging and Tracking in Marine Fisheries*, J. R. Sibert, J. L. Nielsen, Eds. (Springer, 2001), pp. 65–88.
11. K. Aoki, Y. Watanabe, D. Inamori, N. Funasaka, K. Q. Sakamoto, *Phil. Trans. R. Soc. B* **376**, 20200225 (2021).
12. R. F. Keeling, A. Körtzinger, N. Gruber, *Ann. Rev. Mar. Sci.* **2**, 199 (2010).
13. S. J. Jorgensen, A. P. Klimley, A. F. Muhlia-Melo, *J. Fish Biol.* **74**, 1682 (2009).

ACKNOWLEDGMENTS

We thank R. Harcourt for comments on the manuscript.

10.1126/science.adg8452

POLICY FORUM

MARINE POLICY

Underfished or unwanted?

Much blame cast upon fisheries policy may be misguided

By **Kimberly L. Oremus[1], Eyal G. Frank[2], Jesse Jian Adelman[3], Seleni Cruz[1], Janna Herndon[4], Brad Sewell[5], Lisa Suatoni[5]**

With many fish populations around the world overfished (*1*), countries are increasingly passing laws to prevent overfishing and rebuild depleted stocks (*2*–*4*), an objective adopted in the UN Sustainable Development Goals. However, some elected officials (*5*), scientists (*6*), and industry groups in the US and Europe (*7, 8*) have questioned whether such policies have gone too far, resulting in "underfishing," to the economic detriment of fishing communities. This idea is influencing debate over reauthorizing the US Magnuson-Stevens Act (MSA), which shares key features with fishing policies in the European Union (EU) and Canada and is regarded internationally as a benchmark fishing law. Analyzing two decades of data on 170 US fish stocks, we find that the reasons some species are fished less than others are varied, and the MSA is only sometimes the primary factor. In many cases, fishers are fishing less of a species because they find it unprofitable.

Although thresholds for overexploitation are defined in US, EU, and other major fishing laws, thresholds for underexploitation typically are not. The concept of underexploitation is controversial in the field, for various reasons described below, including the many contested assumptions that would be required to measure things such as "lost value" or "lost yield." Still, the notion that underfishing is a serious problem is gaining public attention and influencing policy debates. In the US, a 2014 Congressional Research Service report on the MSA cited scientific testimony asserting that 77% of US stocks were underfished, with 30 to 48% of the country's potential yield lost (*6*). Although the overarching arguments are nuanced, those underfishing claims have been echoed in more simplistic terms by political leaders (*5*) and the news media (*9*). In Europe, fishing groups have raised concerns that precautionary policies have led to underfishing of some stocks (*7*) and proposed that as a rationale to overfish others (*8*). If policies such as the MSA are blamed for underfishing, other countries could be dissuaded from adopting sustainable fisheries policy. [See section 3.1 of the supplementary materials (SM).]

These arguments highlight a central challenge of fisheries policy, which is how to manage trade-offs between the goals of exploitation and conservation under uncertainty, and between different species that may rely on one another or be caught together. The concept of underfishing could imply that the goal of fisheries policy is to facilitate the maximal long-term harvest of the resource, particularly the most valuable stocks, with anything less being "forgone catch." It could then be implied that this lower yield evinces a failure of policy, perhaps because of overly restrictive management. However, fishery policies may also have conservation goals, such as protecting the most vulnerable stocks, maintaining ecosystem services, consistency with species or wildlife protection laws, or supporting cultural and recreational values.

Despite ongoing rhetoric, the argument that underfishing is a substantial problem has received surprisingly little empirical scrutiny. Our intent is not to adjudicate among competing policy priorities or to formalize the concept of underfishing. Rather, it is to facilitate a more informed discourse by identifying where trade-offs exist between utilization and conservation, and where they may be illusory, as in cases of stocks that would be lightly fished even in the absence of management. We do this by examining empirically the extent to which healthy stocks are being fished below their theoretical long-term sustainable harvest, and investigating the reasons for that on a stock-by-stock basis.

We take the United States as our case study, as the MSA's long history provides ample data for historical review. One previous study (*10*) examined why fishers are not fishing their full quota of groundfish in two US regions. Although we examine a different research question, we share the conclusion that fishing policy resulted in less fishing for most of the subset of stocks studied in (see SM section 3.7 for comparison of results). However, we find that this subset of stocks is not representative of the remaining healthy stocks in these two regions or nationwide.

[1]School of Marine Science and Policy, University of Delaware, Newark, DE, USA. [2]Harris School of Public Policy, University of Chicago, Chicago, IL, USA. [3]Energy Policy Institute, University of Chicago, Chicago, IL, USA. [4]New York State Energy Research and Development Authority, New York, NY, USA. [5]Natural Resources Defense Council, New York, NY, USA. Email: oremus@udel.edu

FISHERY POLICY

Modern fishery policy in the US began with the 1996 Sustainable Fisheries Act (SFA) (*2*), a reauthorization of the 1976 MSA that came with new, science-based requirements for stopping overfishing and rebuilding overfished stocks. The 1976 MSA established the concept of optimum yield (OY), the yield that provides "the greatest overall benefit to the Nation," accounting for economic, social, and ecological factors. The SFA mandated that OY be set at or below maximum sustainable yield (MSY), the greatest level of fishing consistent with sustaining the population over the long term. Fishing mortality (F) is the rate at which a given stock is being depleted by fishing. Overfishing is when F exceeds a defined threshold (often F_{MSY}), whereas overfished is when biomass (B) goes below a defined threshold (often half of B_{MSY}). A stock deemed overfished is subject to a rebuilding plan designed to bring the population back to a biomass level that can produce MSY (B_{MSY}) as quickly as possible (and generally within 10 years). Pursuant to a 2006 reauthorization of the MSA, catch limits to prevent overfishing are required for almost all stocks. Catch limits are typically lowered when $B < B_{MSY}$ to prevent a stock from becoming overfished.

US law does not define "underfished" or "underfishing," nor does it prevent OY from being less than MSY. Previous literature (*10, 11*) has focused on a few main reasons why the catch of a healthy stock might not be maximized: (i) low catch limits are set as a precaution owing to high scientific uncertainty in the population estimate and the stock's status; (ii) the stock cannot be caught without catching another species that is overfished, or near the overfishing limit; or (iii) fishers are choosing not to fish the stock because it is not in sufficiently high demand, making it unprofitable. Other possible reasons a stock may be underfished include ecosystem considerations or compliance with laws such as the Endangered Species Act or the Marine Mammal Protection Act. Understanding which factors are at play for which stocks is a prerequisite for constructive debate about underfishing and its relationship to the MSA.

DEFINING POTENTIALLY LESS UTILIZED AND (LESS) FISHED STOCKS

Per US definitions, whether a fish stock is experiencing overfishing in a given year depends on its utilization (F/F_{MSY}); whether a stock is overfished depends on its biomass.

We focus on stocks that were persistently neither experiencing overfishing nor overfished over most of the study period. We call these stocks "potentially less utilized and (less) fished" (PLUF) rather than experiencing underfishing and underfished, to avoid normative connotations and to signal that further examination is needed to understand why they are not being fished more.

Of the 460 stocks managed in the US, 170 have sufficient information to examine their utilization and status. We include data from 1990 to 2015 because data availability declines sharply after that year, with only 28 stocks having data in 2019 (see SM section 3.4). These 170 represent 85% of US marine fish landings (excluding highly migratory and diadromous fish, which are subject to laws other than the MSA, or in addition to it) (see SM section 2.1.2). Because fish are generally managed on a stock-by-stock basis, it is important to consider potential underfishing on a stock-by-stock basis. This can help us understand whether policy-driven underutilization is a widespread phenomenon or one that is isolated to a few stocks. Stocks, averaged over the post-SFA time period from 1996 to 2015 (see the top left panel of the figure), could be called unhealthy (black dots), either because on-average overfishing was occurring ($F > F_{MSY}$), biomass was not sufficient to produce sustainable yield ($B < B_{MSY}$), or both (see fig. S1 for 1990 and 2015 snapshots). "Unhealthy" does not necessarily equate to an official designation of "overfished," which is a reference point set by regional management councils for each stock, often but not always at half of B_{MSY}. Stocks in the bottom-right quadrant (blue dots) could be called healthy (or potentially underfished and experiencing underfishing), because their average biomass was greater than what would support MSY ($B > B_{MSY}$) and, on average, overfishing was not occurring ($F < F_{MSY}$).

Because averages can mask year-to-year variation in stock status, and to ensure that we examined stocks that were in this bottom-right quadrant for most of the study period, we calculated the percentage of years in which each stock was healthy from 1996 to 2015 (see the top left panel of the figure). Most of the 170 stocks in our dataset were either never healthy (tallest far-left, gray bar) or always healthy (tallest far-right, blue bar). Just over half of the stocks, 88, were healthy for 50% or more of those 20 years. We designate these as PLUF (see SM section 3.4, including alternative PLUF definitions).

Whereas the overall trend in the time period studied shows declining fishing mortality (dark gray lines; see the top right panel of the figure), PLUF stocks stayed relatively flat (blue lines; see the top right panel of the figure). These PLUF stocks, on average, had low fishing mortality in 1990, prior to SFA, and continued to have low fishing mortality as of 2015, post-SFA. Mean fishing mortality declined only slightly post-1996 for PLUF stocks (dashed blue line; see the top right panel of the figure). And weighting these stocks by revenue (solid blue line; see the top right panel of the figure) reveals that mortality held steady for the most economically important PLUF stocks, rather than declining as one would expect if the SFA were a major driver of underfishing. (See SM section 3.8).

PLUF stocks fished just below F_{MSY} and just above B_{MSY} could perhaps be classified as optimally utilized and fished. This study does not attempt to classify stocks as optimally utilized and fished because the term lacks a clear definition. However, there are reasons for setting OY or catch limits below the overfishing limit (OFL). For example, the federal National Standard 1 (NS1) Guidelines recommend setting OY below MSY and catch limits below OFL in cases of scientific and management uncertainty with regard to measuring or surpassing MSY and OFL (*12*). Furthermore, managers may lower catch limits to reduce bycatch of stocks in rebuilding.

EXAMINING REASONS

We reviewed the stock assessments and fishery management plans and interviewed stock assessors to discuss why each of the 88 stocks were PLUF. The stock assessments provide quantitative, peer-reviewed data; the fishery management plans are official policy implementation documents; and the stock assessor interviews are qualitative. Our approach synthesizes the information to provide descriptive evidence for the determinants of PLUF stocks, rather than causal linkages. We classified these stocks into four categories, corresponding to the primary reason they were not fished more: economics, bycatch considerations, scientific uncertainty, and other (see the middle panel of the figure). (See "One-page Summaries" in the SM.)

Forty-one PLUF stocks were classified in the economics category, meaning that they were not fished more because fishers found it unprofitable to do so. Stock assessors and fishery management plans indicate that fishers were choosing to fish well below these stocks' catch limits owing to their low prices. Of those 41 stocks, 40 have catch levels consistently below their catch limits (fig. S4). These stocks are frequently considered nontargeted or bycatch species (see SM 3.2, 3.6, and the "One-page Summaries" for details).

"Bycatch considerations" refers to stocks that cannot be profitably caught without also catching other species that are more tightly managed. For example, yelloweye rockfish is common bycatch in the China rockfish fishery on the Pacific Coast; yelloweye rockfish is considered overfished and subject to restrictive catch limits, which indirectly limited catch of China rockfish. Nine of the 23 PLUF stocks in the bycatch category were constrained by bycatch regulations to prevent Pacific halibut from being overfished under an international treaty. This treaty is not part of the MSA (see SM section 2.2.3). The remaining 14 stocks had other constraining stocks that were either at risk of overfishing, or already overfished and required by the MSA (*2*) to go into recovery plans (table S9).

Of the 14 PLUF stocks in the scientific uncertainty category, seven had updated reference points that changed their status from fished near F_{MSY} or overfished to PLUF. This implies that they were thought to be unhealthy, potentially triggering lower catch limits, but more data or improved models showed retroactively that they were healthy—suggesting unreliability in the earlier data. The other seven stocks were identified by stock assessors as having high uncertainty in their MSY, which could lead to overfishing if MSY were overestimated. Consistent with NS1 guidelines, managers can set catch limits below the OFLs to account for scientific and management uncertainties (*12*). These buffers could make a stock appear to be underutilized when in reality it is not.

Finally, 10 PLUF stocks were categorized as "other." Reasons for this categorization included compliance with laws other than the MSA, such as the Endangered Species Act or a non-MSA cap on Alaskan pollock catch; stocks' particular circumstances; or a combination of more than one of the categories previously described. (See "One-page Summaries" and SM section 2 for details.)

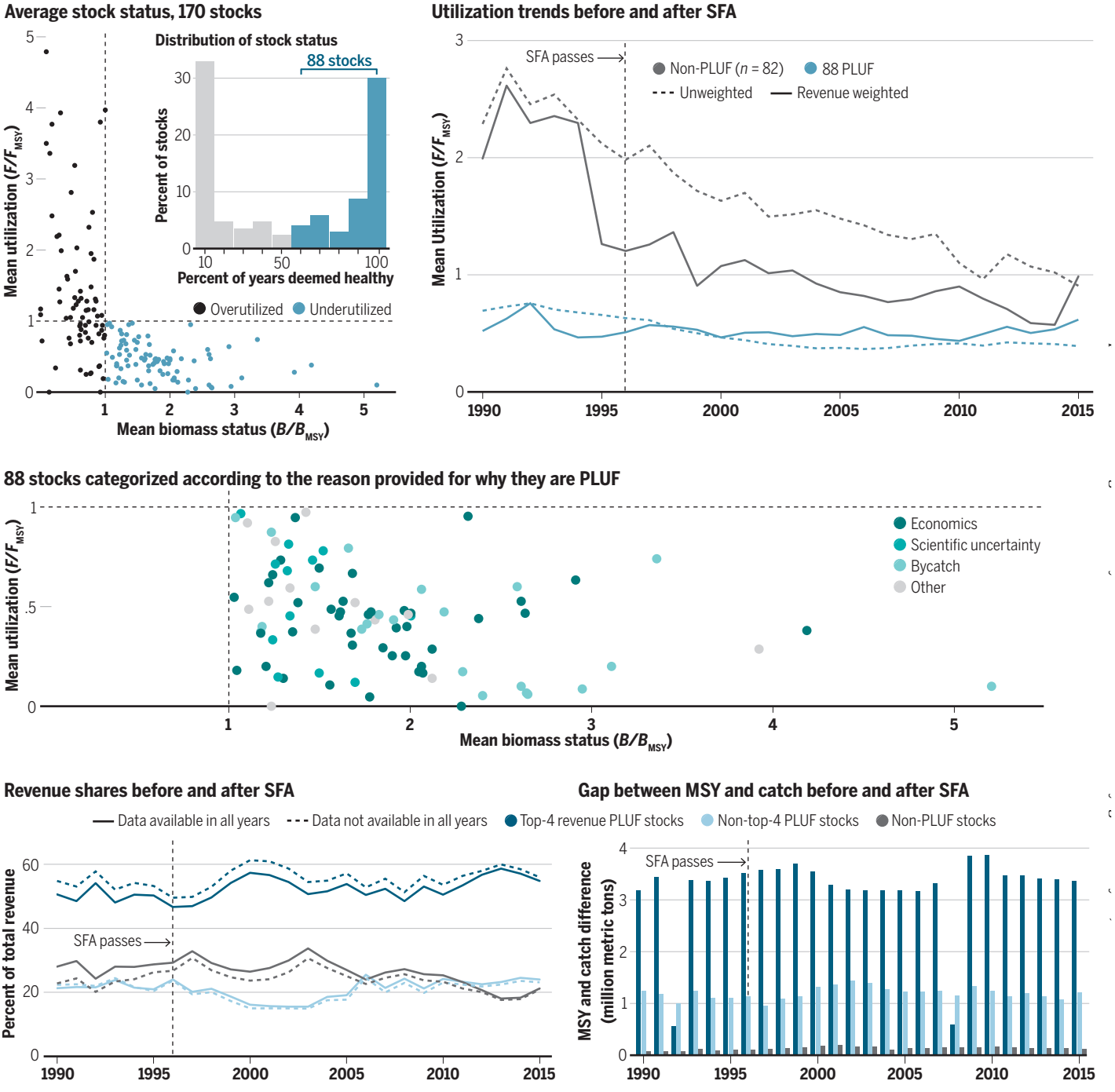
Of the four categories, "bycatch" and "uncertainty" imply that fishery policies are constraining catch either directly or indirectly to prevent stocks from becoming overfished or to prevent overfishing; "economics" implies that policy is not the relevant constraint; and "other" includes some of each. However, MSA policies are only constraining 14 of 23 stocks categorized as "bycatch" and partially constraining 6 of 10 stocks categorized as "other."

ECONOMICALLY IMPORTANT STOCKS

From a market standpoint, not all fish are equally important (fig. S2). Just four stocks made up, on average, 78.3% of the revenue of all PLUF stocks and 52.2% of the revenue of the 147 stocks that had revenue data (see the bottom left panel of the figure). These four stocks all happen to be categorized as "other." In all, stocks categorized as "other" accounted for an average of 56.6% of revenue from those 147 stocks. "Uncertainty" made up 7.9%, "bycatch" 3.9%, and "economics" averaged just 4.8% of revenue despite comprising

Characterization and categorization of potentially less utilized and fished stocks

Panels draw upon data for 170 stocks, spanning 20 years from 1996 to 2015, for graphs without years on the x-axis. Stocks in the lower-right quadrant of the scatterplot of F/F_{msy} versus B/B_{msy} are on average healthy. The histogram reflects 88 stocks defined as potentially less utilized and fished (PLUF), owing to their annual status being deemed healthy in more than 50% of the 20 years. Trends are analyzed before and after the 1996 Sustainable Fisheries Act (SFA).



GRAPHIC: D. AN-PHAM BASED ON K. L. OREMUS *ET AL.*

the largest number of stocks (fig. S3a). The remaining 26.8% were non-PLUF stocks. This highlights a key takeaway from our analysis: Potential benefits from increasing utilization are not uniform across stocks. The four PLUF stocks that generated the greatest revenue were Eastern Bering Sea (EBS) walleye pollock, Atlantic sea scallop, Gulf of Mexico (GOM) brown shrimp, and GOM white shrimp. EBS walleye pollock catch is limited by an annual 2 million metric ton cap designed in 1984 to extract the optimum yield of multispecies groundfish in the Bering Sea and Aleutian Islands (*13*). Congress enshrined the cap in the 2004 Consolidated Appropriations Act, putting it outside the scope of the MSA. Atlantic sea scallop biomass increased rapidly because of area closures intended to address overfishing of both sea scallops and groundfish. However, a recent assessment documents concerns that MSY may be overestimated, in which case it might not be a PLUF stock at all (*14*). GOM brown and white shrimp are PLUF owing to a combination

of low profitability from competition with cheaper imports, bycatch considerations for red snapper, and scientific uncertainty (*15*). (See SM "One-page Summaries".)

The share of revenue for these top four revenue-generating PLUF stocks stayed fairly steady from 1990, 6 years before SFA was passed, to 2015, 19 years after its adoption (see the bottom left panel of the figure), suggesting that they were not disproportionately affected by the policy. The remaining PLUF stocks and non-PLUF stocks also had a fairly steady share of revenue from 1990-2015. To examine these trends further, we look at the gap between MSY and catch from 1990 to 2015 (see the bottom right panel of the figure). Although some call this "lost yield" or "forgone catch" (*11*), interpreting it as "forgone" implies a counterfactual in which every stock is fished precisely to MSY, which is unlikely to hold in reality for several reasons. These include trade-offs and spillovers between stocks, rising costs and diminishing economic returns from greater fishing of some stocks, and uncertainty in the MSY value itself (see SM section 3.5). Still, the gap is worth examining because the notion of forgone catch has become part of the debate as to whether policies such as the MSA are constraining catch more than necessary. Interestingly, we find that 60% of the gap between MSY and catch for all stocks in the dataset combined is due to one stock, EBS walleye pollock. In all, 72.1% of the gap comes from the top four revenue-generating stocks (including EBS walleye pollock), 4.4% from the remaining "other" stocks, 2.8% from "uncertainty," 15% from "bycatch," and 3% from "economics" (fig. S9). If policies were increasingly precautionary, as some critics have argued, this gap between MSY and catch would presumably increase over time. However, the total gap and total catch remains fairly steady from 1990 to 2015 (see the bottom right panel of the figure and fig. S11, respectively).

IMPLICATIONS

Previous literature and rhetoric (*5*–*11*) has focused on management as the primary driver of underfishing. By combining more up-to-date quantitative data on all US stocks with qualitative examinations of each stock, we find a more complicated picture. We observe that a plurality of PLUF US fish stocks are fished below MSY primarily for economic reasons. For these stocks, the key to achieving higher yields, if desired, may lie not in looser policy but in stimulating greater demand.



A boat leaves Gloucester, Massachusetts, to fish at George's Bank.

The MSA is more relevant to the 16% of stocks that are constrained by other stocks that are overfished or near OFLs under the MSA. It is possible that further research and development of technical measures to increase fishery selectivity could be made to avoid bycatch species. It has also been suggested that overfishing of certain stocks could be tolerated to allow more catch of others, which would require statutory and regulatory changes. Allowing such overfishing would risk long-term harms to these stocks, ecosystems, and reliant fishing communities.

For the one-fifth that are constrained by scientific uncertainty, it is conceivable that additional research investments could yield more precise estimates, thus allowing catch limits to be set closer to OFLs. But further research could also support the conclusion that these buffers are in fact necessary to prevent accidental overexploitation, e.g., owing to high variability in productivity.

Any future analysis on the "forgone catch" or revenue dimensions of underfishing should be informed by the finding that just a few stocks account for the majority of both. If the concern is lost revenue or catch, it would make the most sense to narrow the focus to this small number of high-value stocks, which are all in the "other" category and are constrained for specific reasons that may or may not prove tractable. For instance, the highest-value PLUF stock of all, EBS walleye pollock, is constrained for reasons beyond the scope of the MSA. Alternatively, policy-makers may conclude that catching less than the maximum sustainable amount of some stocks is a tolerable or even desirable outcome given the trade-offs that would be involved in exploiting them further. ■

REFERENCES AND NOTES

1. FAO, *The State of World Fisheries and Aquaculture* 2022 (2022).
2. Sustainable Fisheries Act: Amendments to the Magnuson Fishery Conservation and Management Act, Magnuson-Stevens Act, 16 U.S.C. § 1801 et seq. (1996).
3. Regulation 1380/2013, Common Fisheries Policy. European Parliament, Council of the European Union (2013).
4. Fisheries Act: An Act to amend the Fisheries Act and other Acts in consequence, S.C. c.14 (2019).
5. D. Trump, Remarks By President Trump In A Roundtable On Supporting America's Commercial Fishermen, Army Air National Guard Hangar A, Bangor, Maine (5 June 2020).
6. Congressional Research Service, Reauthorization Issues for the Magnuson Stevens Fishery Conservation and Management Act, CRS Report Prepared for Members and Committees of Congress (2014).
7. Serious Risk of Under-fishing, The National Federation of Fishermen's Organisations (9 October 2017).
8. V. A. Doğan, 2019 ist die Fischerei-Krise Überstanden (10 October 2017).
9. CNBC, How The U.S. Went From Overfishing To Underfishing (5 October 2022).
10. K. McQuaw, R. Hilborn, *Mar. Policy* **117**, 103931 (2020).
11. R. Hilborn *et al.*, *Fish Fish.* **22**, 911 (2021).
12. Magnuson-Stevens Act Provisions, 50 C.F.R. § 600 (2009).
13. NPFMC, Amendments to the Fishery Management Plan for Groundfish in the Bering Sea/ Aleutian Island Area (2004).
14. NEFSC, Assessment update for Atlantic Sea scallops for 2020, Fish stock assessment, NOAA (2020).
15. Fisheries of the Caribbean, Gulf of Mexico, and South Atlantic; Shrimp Fishery of the Gulf of Mexico; Amendment 17A. (2016).
16. K. L. Oremus *et al.*, Supplementary Materials and Replication folder for Oremus *et al.*, Zenodo (2023); https://doi.org/10.5281/zenodo.7868122.

ACKNOWLEDGMENTS

We thank National Oceanic and Atmospheric Administration staff and stock assessors for their time answering data and fishery questions. We also thank S. Jardine, C. Legault, J. Rising, and C. Szuwalski for helpful comments. K.L.O. and S.C. received funding from the Natural Resources Defense Council (NRDC) for this research. E.G.F. and J.J.A. receive funding from NRDC as part of a different research project. K.L.O. and E.G.F. contributed equally to this work. Supplementary data and code are available at Zenodo (*16*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf5595

10.1126/science.adf5595



Cryostats are used to cool quantum computers at the Leibniz Supercomputing Centre in Germany.

PHOTO: SVEN HOPPE/PICTURE ALLIANCE VIA GETTY IMAGES

BOOKS *et al.*

COMPUTER SCIENCE

Calculating the future

A quantum computing primer offers insights into the technology's most promising potential applications

By **Dov Greenbaum**[1,2] *and* **Mark Gerstein**[3]

In his latest book, *Quantum Supremacy*, theoretical physicist Michio Kaku expertly describes and rectifies common misconceptions about quantum computing—a technology regarded by experts as one that is likely to have profound societal implications.

The book has four distinct sections, the first focusing on foundational computing and physics concepts. Here, Kaku deftly navigates the relevant scientific landscape by delving into quantum mechanics and identifying four key postulates. This section also provides a lucid treatment of early computing, including Charles Babbage's digital computers and Alan Turing's seminal contributions to the notion of computability.

Kaku excels at developing understandable metaphors for the complexities of quantum mechanics and computing. Consider the distinction he draws between classical and quantum computing: "An ordinary digital computer…is like several accountants toiling away independently in an office…But a quantum computer is like a roomful of interacting accountants, each one simultaneously computing, and, importantly, communicating with each other via entanglement." In another metaphor, he likens electron orbits to a hotel in which different types of rooms fill up in a particular order.

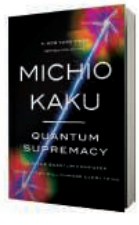
In the remaining sections, Kaku employs a formulaic process to explore potential applications of quantum computing that, at times, stray into the hyperbolic. Such examples include a tongue-in-cheek reference to how quantum computing could revolutionize cooking by enabling people to precisely calculate the outcomes of combining different ingredients.

Given the wide spectrum of proposed quantum computing applications covered, readers would have benefited from a more explicit description of the three broad categories into which they fall. The first includes instances wherein computing is not necessarily the bottleneck holding back advancement. This category includes applications such as asteroid detection, neural networks, and liquid biopsies. In these cases, a better understanding of the problem under consideration could arguably be more helpful than increased computational heft.

In the second category, increased computational power could potentially be useful in furthering a field; however, it is unclear whether quantum computing could appreciably accelerate a discovery because of the intrinsic structure of a problem's calculations. This would include the challenge posed by large-scale searches of genomic databases and calculations for various climate scenarios.

The third and most interesting set of applications relates to those challenges that remain intractable with classical computing but are thought to be potentially solvable by quantum computing. For example, when the application is related to reacting atoms and molecules, "only a quantum computer can truly simulate a quantum process," writes Kaku, summarizing Richard Feynman's 1981 observation. Likewise, many biochemical processes, such as photosynthesis and nitrogen fixation, can be best understood using quantum computers.

A number of potential problems presented by quantum computing are notably underrepresented in *Quantum Supremacy*. One such conundrum that is of particular popular and political interest relates to the ability of quantum computers to potentially crack current encryption technologies. This has spurred considerable interest in quantum computing in the defense industry and on Wall Street. Concerns surrounding a potential "Q-Day" scenario—when quantum computers can practically overcome existing cryptographic codes—have already spurred much development in postquantum cryptographic standards spearheaded by organizations such as the National Institute of Standards and Technology (*1*).



Quantum Supremacy: How the Quantum Computer Revolution Will Change Everything
Michio Kaku
Doubleday, 2023. 352 pp.

Another related application that could have benefited from greater discussion is the potential for quantum information transmission to provide secure transmission routes for information as a result of quantum entanglement. In 2022, Alain Aspect, John Clauser, and Anton Zeilinger were awarded the Nobel Prize in Physics for their work in this area.

Quantum Supremacy ends with a description of a hypothetical day in the year 2050. In this imagined future, quantum computing has cured cancer, made fusion power possible, and routinized travel to the Moon. But we should be cautious of expecting so much from a nascent technology. Nonetheless, the book is well written and accessible, offering readers a comprehensive overview of quantum computing, its underlying principles, and its potential. ■

The reviewers are at the [1]Zvi Meitar Institute for Legal Implications of Emerging Technologies and Harry Radzyner Law School, Reichman University, Herzliya, Israel; [2]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA; and [3]Program in Computational Biology and Bioinformatics, Integrated Graduate Program in Physical and Engineering Biology, and Department of Computer Science, Yale University, New Haven, CT, USA. Email: dov.greenbaum@aya.yale.edu

REFERENCES AND NOTES

1. R. Bavdekar *et al.*, in *2023 International Conference on Information Networking (ICOIN)* (IEEE, 2023), pp. 146–151.

10.1126/science.adh5148

MATERIALS SCIENCE

Making meaning from materials

Hands-on experiences can deepen our understanding of the substances that surround us

Handmade: A Scientist's Search for Meaning Through Making
Anna Ploszajski
Bloomsbury, 2023. 320 pp.



By **Dorothy Jones-Davis**

Makers—people who craft and build things—often look at a material and think "What can I make with this?" However, a maker's raw material has often already been "made," either by nature, nurture, or both, having emerged from processes that affect its atoms, molecules, and bonds. In *Handmade*, materials scientist and maker Anna Ploszajski considers how materials' microscopic properties inform their macroscopic possibilities, taking readers on an incredible journey through the origin, composition, use, and history of 10 everyday materials—from glass to plastic, wood to wool—that make up just about everything in our world.

Ploszajski begins the book with an incident that will be familiar to many scientists: the breaking of a piece of laboratory glassware. When she endeavors to repair it at her university's glass repair shop, she becomes immersed in the art and science of glassblowing. Ploszajski recounts handling glass through its transition from a rigid solid to a "deliciously viscous liquid," taking the reader through the chemical and physical properties that govern the structure of this material during this transition. We learn why medieval glass has uneven thickness (hint: it has to do with the process used to spin the glass) and why light passes (relatively) easily through glass (it has to do with its atomic structure).

The most important lesson in the book's first chapter comes from Ploszajski's realization that her hands-on experience with glassmaking offered her a greater understanding of the material. "Until I met the craftspeople behind it, glass was always looked through, never looked at," she writes. Through making, each material develops an identity beyond its atomic composition.

Yet even that identity is subject to interpretation. "Exactly the same material can range from being a priceless artefact to useless litter," muses Ploszajski in a chapter dedicated to paper. This brings forth the question: Can hands-on experience with a material change how one thinks about its value? Her answer is resoundingly yes.

Ploszajski introduces readers to various artisans, makers, and craftspeople and to the ways they have learned more about properties of the materials with which they work through making. "It might be that the skill of the craftsperson is to notice and tap into…inherent material personalities, to make the best of them when they transform materials into objects," she writes.



Blacksmith Andrea Kosec manipulates hot iron rods in Hajnáčka, Slovakia.

In chapter 10, we meet Andrew Ziminski, a stonemason and conservationist who is in the midst of repairing the spire of a 14th-century stone monument. Through his work, we learn how porous stone is subject to degradation by natural forces and materials, such as acidic rain and the molecular and physical state changes that occur in water as it freezes. We learn how the art and science of stone masonry, when done well, can repair this damage, restoring both stone's beauty and its structural soundness, and, when done poorly, can cause irreparable harm.

Again and again, Ploszajski's own experiences inform her narrative. In a chapter on brass, for example, readers learn about the material through the lens of the author's prized trumpet. Here, she poses a philosophical question that reiterates how a material's value is related to its use: "Would my trumpet still maintain its meaning if it weren't made from brass, but instead of silver or plastic, or even of nothing at all?"

Meanwhile, in a chapter on clay, readers see how a material and the act of making can be both functional and transformational. Clay and its pliable structure provided a mechanism for catharsis during a difficult time in Ploszajski's career. Throughout the book, her curiosity-driven approach to everyday objects encourages the reader to navigate the world similarly and to make sense of the material world around them through making.

Laudably, the book also seeks to offer readers an accurate and unbiased historical description of each material's origin, evolution, and usage throughout time, describing the contributions of international trade, war, and human choice. In chapter 2 ("Plastic"), for example, Ploszajski details the role that World War II played in increasing demand and innovation in the synthetic rubber industry, sometimes at the expense of human lives.

Similarly, the book also calls out injustices and inaccuracies surrounding the materials under consideration and celebrates commonly overlooked contributors to the histories of these materials. In chapter 3 ("Steel"), Ploszajski recounts the story of the "Women of Steel"—the hundreds of women who worked during World Wars I and II making tools and munitions in the steelworks of Sheffield, England.

Ultimately, Ploszajski succeeds in demonstrating that the act of making something can impart an appreciation for and understanding of the properties of the material from which it is made. "Just give it a go," she urges the reader. "There are tales to be told about these substances… I hope that this book will inspire you to tell yours." ■

The reviewer is at KID Museum, Bethesda, MD, USA, and is a senior adviser at Nation of Makers, Silver Spring, MD, USA. Email: dorothy.jonesdavis@gmail.com

10.1126/science.adh2248



Salt crystals accumulate on the shore of the Persian Gulf, where discharge from desalination plants has increased salinity.

PHOTO: TALEB SHADI

Edited by **Jennifer Sills**

Brine discharge disrupts Persian Gulf ecosystems

The Persian Gulf is home to a diverse marine ecosystem with a variety of species that rely on the delicate balance of saltwater and freshwater to survive (*1*). However, the growing demand for freshwater in the region has led to the construction of seawater desalination plants, which discharge brine (highly concentrated saltwater) into the Gulf. This brine discharge increases salinity levels and disrupts the natural balance of the Gulf's waters, threatening the marine ecosystem's sustainability (*2*).

Because the Gulf has a low rate of water refreshment and a high rate of evaporation (*3*), increases to salinity are difficult to reverse. High salinity levels can cause stress and death in many Gulf species of fish and other sea creatures. Increased salinity, and the resulting decrease in oxygen levels (*4*), can also threaten plankton (*5*), which form the base of the food chain. In coral reefs, high salinity levels can cause coral bleaching and death, which can have cascading effects throughout the ecosystem (*6*). The substantial harm to marine life could have implications for human wellbeing, including food security and livelihoods in coastal communities (*7*).

Given the harmful effects of increased salinity, it is essential to find solutions that minimize brine discharge from seawater desalination plants in the Persian Gulf. Government agencies, desalination plant operators, environmental organizations, and local communities should monitor salinity levels regularly and adjust desalination plant operations accordingly. Instead of dumping brine back into the Gulf, plants could use alternative storage options such as deep well injection or evaporation ponds, or they could arrange for the brine to be used for industrial purposes such as mining (*8*). In addition to monitoring and repurposing current levels of brine, Iran's long-term strategy should include the development of desalination technology that produces less waste by using more efficient membranes or other filtration methods (*9*). The region should also maximize the use of sources of freshwater that do not rely on seawater desalination, such as rainwater harvesting (*10*) and wastewater recycling (*11*), which have been used effectively in other Middle Eastern countries (*12*).

Reza Sheibani[1]*, Marjan Sheibani[2], Ahmad Parandoust[3]
[1]Amirkabir University of Technology-Mahshahr Campus, Nahiyeh san'ati, Mahshahr, Khouzestan, Iran. [2]Department of Chemistry, Omidiyeh Branch, Islamic Azad University, Omidiyeh, Iran. [3]Farabi Educational Institute, Tehran, Iran.
*Corresponding author.
Email: rsheibani2003@gmail.com

REFERENCES AND NOTES

1. G. O. Vaughan, N. Al-Mansoori, J. A. Burt, in *World Seas: An Environmental Evaluation: Volume II: the Indian Ocean to the* Pacific, C. Sheppard, Ed. (2019), pp. 1–23.
2. W. J. F. Le Quesne *et al.*, *Mar. Pollut. Bull.* **173**, 112940 (2021).
3. J. Kämpf, M. Sadrinasab, *Ocean Sci.* **2**, 27 (2006).
4. R. Lange *et al.*, *J. Exp. Mar. Biol. Ecol.* **9**, 217 (1972).
5. T. Al-Said *et al.*, *Environ. Monit. Assess.* **189**, 268 (2017).
6. H. A. Naser, in *Biodiversity: The Dynamic Balance of the Planet*, O. Grillo, Ed. (2014), chap. 12.
7. M. A. E. Forio, P. L. M. Goethals, *Sustainability* **12**, 5603 (2020).
8. A. Giwa, V. Dufour, F. Al-Marzooqi, M. Al-Kaabi, S. W. Hasan, *Desalintion* **407**, 1 (2017).
9. A. Subramani, J. Jacangelo, *Water Res.* **75**, 164 (2015).
10. United Nations Development Program, "Water governance in the Arab region: Managing scarcity and securing the future" (2013).
11. M. Abdel-Dayem *et al.*, "Water reuse in the Arab world, from principle to practice—voices from the field (English)" (World Bank Group, 2011).
12. M. Iftikhar Hussain *et al.*, *Agric. Water Manag.* **221**, 462 (2019).

10.1126/science.adh8179

Prioritize wild species abundance indicators

In addition to adopting the Kunming-Montreal Global Biodiversity Framework (GBF), the Parties to the Convention on Biological Diversity (CBD) agreed to an accompanying monitoring framework to assess progress (*1*). However, the metrics in the monitoring framework do not adequately align with the GBF's Goal A, one of the four overarching GBF goals (*2*), which states that "the abundance of native wild species [should be] increased to healthy and resilient levels." The emphasis on the abundance of wild species is appropriate, as it constitutes a fundamental aspect of biodiversity (*3*). However, a corresponding abundance metric is not among the 25 "headline indicators," on which the parties are required to report. The headline indicators, which include important metrics such as the extinction risk of species and the extent of natural ecosystems, are intended to capture the overall scope of and progress toward the goals and targets of GBF at both global and national levels (*1*). To meet the GBF's objectives, abundance should be added to the list of headline indicators.

Tracking abundance at the national level will require an increase in national and international support for monitoring efforts (*4*, *5*), an investment the parties to the CBD have committed to by agreeing to the GBF. However, given that countries are only required to report on the headline indicators, monitoring efforts will likely be restricted to tracking the metrics listed. As a result, data on abundance of wild species would be insufficient in some countries.

The Living Planet Database (*6*) could serve as the foundation for an abundance indicator. This still growing database currently includes more than 38,000 individual vertebrate population time series, providing a unique window into changes in species abundance since 1970 (*6*). It has been the backbone of some of the most important assessments of the successes and shortcomings of the CBD (*7*–*10*). In line with the CBD's commitment to headline indicators that are meaningful at the

country level, the abundance data have also been used to develop national population abundance indicators such as the China Biodiversity Observation Network (*11*) and community-based biodiversity indicators for East Africa (*12*). Such data could be incorporated into reports required by the GBF monitoring framework.

Producing national abundance indicators from the Living Planet Database poses challenges in data representation and analysis (*4*). However, this database provides a model for measuring changes in species abundance using the approach and methodology of the Living Planet Index. Countries could also deposit data in the database.

Parties to the convention have now given a mandate to an Ad Hoc Technical Expert Group to advise on the further implementation of the monitoring framework, including filling gaps (*1*). We urge this body to add a headline indicator on abundance. Emphasizing the need to assess trends in species abundance and providing a plan to do so effectively will increase the likelihood that countries prioritize and invest in abundance tracking, allowing each goal to be adequately measured.

Jonas Geldmann[1], Achilles B. Byaruhanga[2], Richard D. Gregory[3,4], Piero Visconti[5], Haigen Xu[6]
[1]Center for Macroecology, Evolution and Climate, Globe Institute, University of Copenhagen, Copenhagen, Denmark. [2]Nature Uganda, Katalima Crescent Lower Naguru, Kampala, Uganda. [3]RSPB Centre for Conservation Science, The Lodge, Sandy, Bedfordshire SG19 2DL, UK. [4]Centre for Biodiversity and Environment Research, Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK. [5]Biodiversity, Ecology and Conservation Group, International Institute for Applied System Analyses, 2361 Laxenburg, Austria. [6]Nanjing Institute of Environmental Sciences, Ministry of Ecology and Environment of China, Nanjing 210042, China.
*Corresponding author.
Email: jgeldmann@sund.ku.dk

REFERENCES AND NOTES

1. Convention on Biological Diversity, "Decision CBD/COP/15/L.26: Monitoring framework for the Kunming-Montreal global biodiversity framework" (2022).
2. Convention on Biological Diversity, "Decision CBD/COP/15/L.25: Kunming-Montreal Global biodiversity framework" (2022), p. 8.
3. H. M. Pereira *et al.*, *Science* **339**, 277 (2013).
4. L. McRae, S. Deinet, R. Freeman, *PLOS ONE* **12**, e0169156 (2017).
5. V. Proença *et al.*, *Biol. Conserv.* **213**, 256 (2017).
6. "Living Planet Report 2022–Building a nature-positive society" (World Wide Fund for Nature, Zoological Society of London, 2022).
7. D. P. Tittensor *et al.*, *Science* **346**, 241 (2014).
8. "The IPBES Global Assessment on Biodiversity and Ecosystem Services" (Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, Bonn, Germany, 2019).
9. Secretariat of the Convention on Biological Diversity, "Global Biodiversity Outlook 5" (Convention on Biological Diversity, 2020).
10. L. McRae *et al.*, *One Earth* **5**, 422 (2022).
11. J. Yi *et al.*, *Eco-Environ. Health* **1**, 201 (2022).
12. S. R. Wotton *et al.*, *Oryx* **54**, 62 (2020)

10.1126/science.adh4409

Clinical trials can adapt for refugees

Since Russia invaded Ukraine in February 2022, 8.2 million Ukrainian refugees—mostly women, older men, and children—have fled to neighboring countries (*1*–*3*). These refugees included cancer patients (*4*), some of whom were enrolled in international clinical trials (*5*, *6*). Restarting clinical trials with new participants can be expensive and time consuming, and delays to progress affect all cancer patients. To minimize trial disruptions, cancer clinical trial oncologists, researchers, regulatory authorities, and industry partners should bring the trial to the patient by using satellite (decentralized) trial sites with patient navigators and remote technology.

Many countries that host refugees, such as the United States, Germany, Romania, and South Africa, have existing clinical trial infrastructure (*7*). Ukrainian trial participants were able to access care at previously established trial sites in neighboring countries (*8*). Other countries hosting displaced trial participants should also facilitate transfers to local trial sites. Trained patient navigators, who understand the clinical trial process and digital technologies, can help guide, educate, and assist displaced patients through the decentralized clinical trial landscape. In the case of a trial site with no or low enrollment of citizens, researchers should collaborate to determine if the eligibility criteria can include refugees.

Decentralized digital solutions, such as remote access to healthcare providers, can enable seamless and sustainable information sharing between trial sites. Online tools can translate electronic consent forms, allow patients to report outcomes and adverse events remotely, and automate invoice generation. Contract research organizations can assess the cancer medicine supply in real time. Trial researchers in home and host countries can share their assessments of the patients' and researchers' needs.

Many refugee populations are hosted by countries with inadequate clinical trial infrastructure or in locations without easy access to trial sites. In some cases, loaning WiFi-enabled tablets to trial participants could help reduce the digital divide. Participants without easy access to care could also benefit from local collection of blood or biospecimens, which could then be transported to central sites for biomarker monitoring.

In addition to protecting the integrity of ongoing trials, academic decentralized clinical trials could improve cancer care equity and consistency in care delivery. Adding flexibility to cancer care (*9*, *10*) will improve diversity and inclusion in trial enrollment and mitigate the impact of war and other disasters on trial completion. Changes to trial infrastructure could also serve as a model for continuity of care beyond clinical trials, potentially improving cancer treatment for all underserved citizens and refugees.

Aditi Hazra[1]* and Igor Bondarenko[2]
[1]Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [2]Oncology and Medical Radiology Department, Dnipropetrovsk Medical Academy, Dnipro State Medical University, Dnipro, Ukraine.
*Corresponding author.
Email: ahazra@bwh.harvard.edu

REFERENCES AND NOTES

1. Ukraine Refugee Situation, United Nations High Commissioner for Refugees Operational Data Portal (2023); https://data.unhcr.org/en/situations/ukraine.
2. I. Bondarenko *et al.*, *Eur. J. Cancer* **183**, 95 (2023).
3. M. Babicki, A. Mastalerz-Migas, *BMJ* **377**, o1440 (2022).
4. A. Pandey, *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2215424120 (2023).
5. S. M. Wren, H. Wild, *World J. Surg.* **46**, 2487 (2022).
6. E. Dolgin, *Cancer Discov.* **12**, 1178 (2022).
7. F. Rubagumya, *JAMA Netw. Open* **5**, e2227252 (2022).
8. C. Thallinger, C. C. Zielinski, *Lancet Oncol.* **23**, e440 (2022).
9. C. Kurihara *et al.*, *Front. Med.* **9**, 966220 (2022).
10. J. Moore *et al.*, *Orphanet. J. Rare Dis.* **17**, 240 (2022).

COMPETING INTERESTS

A.H. and I.B. are members of the Ukraine Clinical Research Support Initiative.

10.1126/science.adh1190

TECHNICAL COMMENT ABSTRACTS

Comment on "Exceptional preservation of organs in Devonian placoderms from the Gogo lagerstätte"

Bjarke Jensen *et al.*

Trinajstic *et al.*, (*Science*, 16 September 2022, p. 1311–1314) describe exceptionally well-preserved organs in fossilized Devonian placoderms to infer the early evolution of the vertebrate heart. We argue that the report has numerous shortcomings and examples of mixed specimen codes. Further, we question whether there indeed is any evidence for a mineralized chambered heart in these placoderms.

Full text: dx.doi.org/10.1126/science.adg2748

Response to Comment on "Exceptional preservation of organs in Devonian placoderms from the Gogo lagerstätte"

Kate Trinajstic *et al.*

Jensen *et al*. question evidence presented of a chambered heart within placoderms, citing its small size and apparently ventral atrium. However, they fail to note the belly-up orientation of the placoderm within one nodule, and the variability of heart morphology within extant taxa. Thus, we remain confident in our interpretation of the mineralized organ as the heart.

Full text: dx.doi.org/10.1126/science.adg3748

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**



Artist's depiction of a herd of *Megacerops*, an elephant-sized brontothere herbivore that lived in the early Holocene

PHOTO: COREY FORD/STOCKTREK IMAGES/SCIENCE SOURCE

PALEONTOLOGY

Growing bigger

After the extinction of large dinosaurs at the end of the Cretaceous, many mammals underwent a rapid increase in size. Several hypotheses for this change have been put forward, with much debate about the drivers. Sanisidro *et al.* looked at the record of body size in brontotheres, large herbivorous mammals from the Eocene that experienced orders of magnitude changes in size. The authors saw no evidence for directional selection, but instead found a pattern of increased survival of larger species caused by reduced competition with other herbivores in their niches. Thus, over time, the increase in body size in these mammals was caused by the macroevolutionary process of species sorting. —SNV

Science, ade1833, this issue p. 616

ORGANIC CHEMISTRY

Carbon–hydrogen bonds get squared

The conventional synthetic route to four-membered carbon rings involves the reaction of two unsaturated compounds with double or triple carbon–carbon bonds. However, the relative orientation of these two partners is often difficult to control. Yang *et al*. present an alternative approach to benzocyclobutenes using oxidative palladium catalysis. One reaction partner is a carboxylic acid that undergoes carbon–hydrogen oxidation at two adjacent alkyl carbon centers; the other is an arene with adjacent bromide and iodide substituents that react at different rates to achieve the desired regiocontrol. —JSY

Science, adg5282, this issue p. 639

AGING

Hungry to live longer

Caloric restriction increases the life span of experimental animals. Weaver *et al*. show that hunger alone can also increase life span in fruit flies. Hunger caused either by depriving flies of branched-chain amino acids (in particular isoleucine) or by optogenetically stimulating brain neurons associated with motivation to feed extended their life span. Prolonged hunger in flies deprived of amino acids seemed to reset a set point for the amount of protein the flies consumed and was associated with changes in histone acetylation in their brains. —LBR

Science, ade1662, this issue p. 625

VIROLOGY

Manipulation of microenvironment

Human cytomegalovirus (HCMV) infection of a single cell in culture induces a proviral state in cells in contact with the single infected cell. The proviral state is not only beneficial to HCMV but can also be exploited by other viruses. Song *et al.* distinguished cells in culture on the basis of their proximity to the infected cell. Elucidation of the proteome of cells at similar distances from the infected cell revealed signatures that explain the susceptibility of proximal cells and the resistance of distal cells to infection. This study underscores the spatiotemporal complexity driving virus infection dynamics in cell culture and has profound implications for virus-induced intercellular communication in vivo. —CEC

Sci. Adv. (2023) 10.1126/sciadv.adg3433

ZEOLITE STRUCTURE

Imaging zeolite variability

Zeolite structures are prone to structural and compositional inhomogeneities that can cause batch-to-batch variations in applications. However, imaging these variations is difficult in transmission electron microscopy (TEM) because zeolites are prone to electron beam damage at the doses needed for atomic resolution. Zhang *et al*. found that electron ptychography based on low-dose four-dimensional scanning TEM data could achieve subangstrom resolution. The authors resolved individual oxygen atom columns in various zeolites with specimen thicknesses of up to 40 nanometers and mapped the distribution of oxygen vacancies throughout a zeolite. Complex intergrown structures between different zeolite phases were also imaged. —PDS

Science, adg3183, this issue p. 633

CANCER IMMUNOLOGY

A STANd against tumors

Effective antitumor T cell responses are often hampered by the constraints of aberrant tumor-associated vasculature. Wang-Bishop *et al*. present an approach harnessing activation of the stimulator of interferon genes (STING) pathway to promote vascular normalization and enhance antitumor responses. The authors

evaluated STING-activating nanoparticles (STANs) intravenously in multiple tumor models. STANs were able to normalize tumor vasculature and integrity, decrease tumor hypoxia, and promote the expression of T cell adhesion molecules, which enhanced antitumor T cell infiltration and improved the efficacy of immune checkpoint inhibitors and adoptive cell therapy. These findings suggest that STANs are a potential therapeutic approach that can target tumor vasculature and enhance the effects of immunotherapy. —CNF

Sci. Immunol. (2023)
10.1126/sciimmunol.add1153

ANCIENT GENOMES

Resurrecting Paleolithic natural products

With advances in the technologies used to extract and analyze ancient DNA, we are learning more about the microorganisms that lived around and within our distant human ancestors and hominin cousins. However, little is yet known about the kinds of molecules these organisms might produce or their roles in health and disease. Klapper *et al*. extracted metagenome-assembled genomes of bacteria in dental calculus from ancient human and Neandertal remains. Natural product biosynthetic gene clusters in these samples were distinctive, and heterologous expression of these clusters yielded 5-alkylfuran-3-carboxylic acid products that the authors refer to as paleofurans. Such methods for resurrecting ancient microbial natural products may provide insight into the nutrition and health of early hominins—and potentially bioactive compounds lost to time. —MAF

Science, adf5300, this issue p. 619

BLACK HOLE PHYSICS

Polarization of a tidal disruption event

If a star passes too close to a supermassive black hole, it gets ripped apart by tidal forces. As the debris falls toward the black hole, it heats up and emits light, producing a transient astronomical source known as a tidal disruption event (TDE). Liodakis *et al*. measured the optical polarization of a TDE, finding that it varied during the event, peaking at 25 ± 4% linear polarization. By comparing this behavior with models, the authors posit that TDE optical emission occurs when the stream of debris extends around a full orbit of the black hole, generating shocks where it collides with itself. —KTS

Science, abj9570, this issue p. 656



Dental calculus from hominin remains, such as this mandible, contains ancient microbial DNA encoding for previously unknown natural products.



IN OTHER JOURNALS

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

IMMUNOPSYCHIATRY

γδ17 T cells stress the system

Interleukin-17–producing γδ (γδ17) T cells, which reside in the meninges, play important roles in anxiety-like behavior, synaptic plasticity, and memory function. Zhu *et al.* report that mice subjected to chronic social-defeat stress and human patients with major depressive disorder show a reduction in specific *Lactobacillus* species in their gut microbiota. In mice, a dearth of these bacteria, which normally degrade fungal-derived β-glucan polysaccharides in the gut, was associated with an increase in colonic and meningeal γδ17 T cells expressing the β-glucan polysaccharide receptor dectin-1. Dectin-1–mediated expansion and differentiation of γδ17 T cells were associated with social avoidance by mice exposed to chronic social-defeat stress. Moreover, inhibition of γδ T cells led to a reduction in stress-susceptible behavior in these animals. —STS

Nat. Immunol. (2023)
10.1038/s41590-023-01447-8

IMMUNITY

Benefits of a grown-up microbiota

The intestinal microbiota modulates immune functions. However, whether the changes occurring in the composition of the microbiota during development affect the adult immune system remains to be verified. Lubin *et al.* created adult mice bearing preweaning microbiota and showed that the animals had an abnormal immune system with reduced numbers of

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

ECOLOGY

Land abandonment and biodiversity

Land abandonment, the end of any human activities, is increasing in rural areas. It is unknown exactly how much and where land is abandoned, how it can best be managed or left to recover to benefit biodiversity and conservation, and how reuse can be implemented effectively. In a Perspective, Daskalova and Kamp discuss the differing outcomes of land abandonment on recovery of biodiversity, whether human intervention is needed, and, if so, in what form. The authors also discuss the need for balance to ensure that rural populations can repurpose land when needed and how working with local or Indigenous people brings historical knowledge of ecosystems and conservation goals. Understanding the complex interactions that lead to land abandonment and subsequent impacts on biodiversity and conservation could improve interactions between people and nature. —GKA

Science, adf1099, this issue p. 581

FERROELECTRICS

Organic actuation

Certain fluoropolymers have ferroelectric properties and attractive mechanical properties for a wide range of applications. Qian *et al*. review the history of these polymers and discuss recent progress, focusing on their potential use in electromechanical, electrocaloric, and dielectric applications. Fluoropolymers are relatively flexible, making these materials attractive for a wide range of applications, including wearable devices. Many challenges remain for improving the properties of these materials for commercial applications. —BG

Science, adg0902, this issue p. 596

CANCER

Tracking pancreatic cancer evolution

The development of pancreatic cancer involves a complex interplay between pancreatic lesions and the surrounding tissue microenvironment. Burdziak *et al*. performed an in-depth analysis of the evolution of pancreatic tumors. Using several model systems, including newly developed computational methods, tumor progression stages leading to pancreatic adenocarcinoma were predicted based upon communication gene modules. Tissue remodeling was traced from a healthy pancreas to an inflamed state, the premalignant stage, and then to full-blown malignancy and distant metastasis. Epigenetic plasticity in early progenitor–like epithelial cells primed pancreatic cells for neoplastic transformation. Plasticity was associated with the remodeling of accessible chromatin close to cell-cell signaling genes and an interleukin-33–mediated inflammatory feedback loop between epithelial and immune cells. —PNK

Science, add5327, this issue p. 597

HUMAN GENETICS

Ancient Mexican genetic structure endures

Ancient DNA has revealed much about the movement and genetic legacy of human populations. However, DNA is notoriously sensitive to degradation in hot climates, resulting in a greatly imbalanced study of regions and ancestries because of climatological restrictions. Villa-Islas *et al*. analyzed ancient DNA from pre-Hispanic individuals in northern and central Mexico, revealing contributions from an unknown "ghost" population (see the Perspective by Llamas and Roca-Rada). They found that some populations endured in regions of central Mexico despite changing climate beginning in the 10th century. Much of the ancient population's genetic structure has withstood population declines and is represented in present-day Indigenous populations. —CNS

Science, add6142, this issue p. 598; see also adh7902, p. 578

RADIO ASTRONOMY

FRBs in a binary star system

Fast radio bursts (FRBs) are brief flashes of radio waves from distant galaxies that are probably emitted by neutron stars. Some FRBs repeat, which enables follow-up observations to constrain their local environments. Anna-Thomas *et al*. have observed more than a hundred bursts from a single FRB source, of which 13 were bright enough to measure the polarization (see the Perspective by Paragi). The fluctuating polarization properties indicate that the magnetic field close to the FRB source was highly variable, twice reversing direction. The authors contend that the most likely explanation is that the FRB source orbits in a binary system, with the radio waves passing through the magnetized stellar wind of the companion star. —KTS

Science, abo6526, this issue p. 599; see also adh8099, p. 580

PALEOECOLOGY

Early human habitats

Early humans and their hominin relatives had to adapt to new environments to spread out of Africa. Zeller *et al*. explored the movements of hominins across and preferences for different biomes by comparing six *Homo* species distributions from the fossil record against simulated climate and vegetation over the past 3 million years. They found that some later species inhabited a broader range of biomes as they spread to colder and more forested areas, especially *H. sapiens*, which settled in more extreme habitats (deserts and tundra). In addition to adapting to changing environmental conditions over time, models suggest that *Homo* species may have preferentially selected areas with more diverse habitats. —BEL

Science, abq1288, this issue p. 604

CATALYSIS

Retaining reactive oxide overlayers

Metal nanoparticles can interact so strongly with reducible oxide supports that the particle becomes coated with the metal oxide during the initial reduction steps. Monai *et al*. used scanning transmission electron microscopy and infrared spectroscopy to explore how titanium oxide overlayers of nickel nanoparticles change during carbon monoxide and carbon dioxide hydrogenation reactions. The thicker, less crystalline titania overlayers that formed under more strongly reducing conditions were partially preserved under hydrogenation conditions. Restructuring of this overlayer created interfacial sites that favored carbon–carbon coupling reactions by providing a reservoir of surface carbon species. —PDS

Science, adf6984, this issue p. 644

ELECTROCHEMISTRY

Doping cobalt avoids iridium

Water electrolysis is a potentially sustainable means of producing hydrogen. Unfortunately, only rare and expensive iridium is a sufficiently stable and active oxidation catalyst in the most efficient acidic environment. Chong *et al*. now report that doping an Earth-abundant cobalt oxide catalyst with lanthanum and manganese ions promotes activity and stability in an acidic proton-exchange membrane water electrolyzer. Simulations

suggest that the lanthanum stabilizes the surface of the catalyst and the manganese enhances conductivity in the bulk. —JSY

Science, ade1499, this issue p. 609

THERMAL REGULATION

Hold your breath

Most fishes are fully ectothermic, meaning that their body temperatures are heavily regulated by their environment. This physiological feature can present a challenge for large, predatory fishes that must maintain a certain body temperature to fully function but must also venture into different thermal environments to find prey. Using state-of-the-art remote loggers, Royer *et al*. found that scalloped hammerhead sharks were able to maintain warmer body temperatures during deep (more than 800 meters) dives by closing their gill slits and thus preventing heat transfer (see the Perspective by Meekan and Gleiss). Functionally, these sharks hold their breath during dives to facilitate access to prey in deep, cold waters. —SNV

Science, add4445, this issue p. 651; see also adg8452, p. 583

HIV

ARTsy antibodies

Posttreatment controllers (PTCs) are able to durably suppress HIV-1 after patients stop taking antiretroviral therapy (ART). However, the features that distinguish PTCs from noncontrollers (NCs) remain unclear. Esmaeilzadeh *et al*. evaluated plasma samples from six patients taking PTCs and six taking NCs who started ART early after infection, measuring viral load, viral diversity, and titers of HIV-1–neutralizing antibodies. The authors found that autologous neutralizing antibodies matured during ART and were an important contributor to viral suppression after ART discontinuation. PTCs could be distinguished from NCs by a stronger neutralizing antibody response and a less diverse proviral reservoir. These features may be used to identify PTCs, although the authors stress the need to look at a more diverse pool of individuals living with HIV-1. —CSM

Sci. Transl. Med. (2023) 10.1126/scitranslmed.abq4490

CORONAVIRUS

The painful consequences of long COVID

After the resolution of infection with the COVID-19–causing virus severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), some individuals experience neurological issues such as generalized pain. Serafini *et al*. investigated the basis for pain associated with long COVID and other neuropathies. Mechanical hypersensitivity in SARS-CoV-2–infected hamsters was associated with a durable gene expression signature in sensory neurons that partly resembled those of mouse models of neuropathic pain. Some of these genes, including the one encoding the RNA-binding protein ILF3, were validated as potential therapeutic targets in mice. —LKF

Sci. Signal. (2023) 10.1126/scisignal.ade4984

evaluated STING-activating nanoparticles (STANs) intravenously in multiple tumor models. STANs were able to normalize tumor vasculature and integrity, decrease tumor hypoxia, and promote the expression of T cell adhesion molecules, which enhanced antitumor T cell infiltration and improved the efficacy of immune checkpoint inhibitors and adoptive cell therapy. These findings suggest that STANs are a potential therapeutic approach that can target tumor vasculature and enhance the effects of immunotherapy. —CNF

Sci. Immunol. (2023)
10.1126/sciimmunol.add1153

ANCIENT GENOMES

Resurrecting Paleolithic natural products

With advances in the technologies used to extract and analyze ancient DNA, we are learning more about the microorganisms that lived around and within our distant human ancestors and hominin cousins. However, little is yet known about the kinds of molecules these organisms might produce or their roles in health and disease. Klapper *et al*. extracted metagenome-assembled genomes of bacteria in dental calculus from ancient human and Neandertal remains. Natural product biosynthetic gene clusters in these samples were distinctive, and heterologous expression of these clusters yielded 5-alkylfuran-3-carboxylic acid products that the authors refer to as paleofurans. Such methods for resurrecting ancient microbial natural products may provide insight into the nutrition and health of early hominins—and potentially bioactive compounds lost to time. —MAF

Science, adf5300, this issue p. 619

BLACK HOLE PHYSICS

Polarization of a tidal disruption event

If a star passes too close to a supermassive black hole, it gets ripped apart by tidal forces. As the debris falls toward the black hole, it heats up and emits light, producing a transient astronomical source known as a tidal disruption event (TDE). Liodakis *et al*. measured the optical polarization of a TDE, finding that it varied during the event, peaking at 25 ± 4% linear polarization. By comparing this behavior with models, the authors posit that TDE optical emission occurs when the stream of debris extends around a full orbit of the black hole, generating shocks where it collides with itself. —KTS

Science, abj9570, this issue p. 656



Dental calculus from hominin remains, such as this mandible, contains ancient microbial DNA encoding for previously unknown natural products.



IMMUNOPSYCHIATRY

γδ17 T cells stress the system

Interleukin-17–producing γδ (γδ17) T cells, which reside in the meninges, play important roles in anxiety-like behavior, synaptic plasticity, and memory function. Zhu *et al.* report that mice subjected to chronic social-defeat stress and human patients with major depressive disorder show a reduction in specific *Lactobacillus* species in their gut microbiota. In mice, a dearth of these bacteria, which normally degrade fungal-derived β-glucan polysaccharides in the gut, was associated with an increase in colonic and meningeal γδ17 T cells expressing the β-glucan polysaccharide receptor dectin-1. Dectin-1–mediated expansion and differentiation of γδ17 T cells were associated with social avoidance by mice exposed to chronic social-defeat stress. Moreover, inhibition of γδ T cells led to a reduction in stress-susceptible behavior in these animals. —STS

Nat. Immunol. (2023)
10.1038/s41590-023-01447-8

IMMUNITY

Benefits of a grown-up microbiota

The intestinal microbiota modulates immune functions. However, whether the changes occurring in the composition of the microbiota during development affect the adult immune system remains to be verified. Lubin *et al.* created adult mice bearing preweaning microbiota and showed that the animals had an abnormal immune system with reduced numbers of



Seagrasses, in combination with lucinid bivalves and their microbial symbionts, help to detoxify sulfides in marine sediments.

PHOTO: _548901005677/GETTY IMAGES

COASTAL ECOLOGY

Facilitating a foundation species

Seagrasses act as habitat for many coastal species but are threatened by global climate change. Studies have shown a positive relationship between seagrasses and lucinid bivalves, which shelter among seagrass roots and host bacteria that detoxify sulfides accumulating in the sediment. In a field survey, de Fouw *et al.* determined whether this facilitative relationship occurs in seagrasses found in western European seas. Lucinid bivalves only occurred with seagrasses at sites that remain above 1°C all year, which comprise about half of the 31 study sites. Models supported positive feedback between bivalve (*Loripes orbiculatus*) density and seagrass biomass, but only for one of the seagrass species, *Zostera noltii*. —BEL

Ecography (2023)
10.1111/ecog.06636

PHOTOCATALYSIS

Spotlight on the alkoxy radical

Substantial recent research has sought to channel light energy into the cleavage of strong carbon–hydrogen bonds. In one implementation, photoexcitation of a cerium catalyst produced new carbon–carbon and carbon–nitrogen bonds from methane, ethane, and propane [*Science* (2018) 10.1126/science.aat9750]. Those authors attributed the carbon–hydrogen cleavage to alkoxy radicals from an alcohol co-catalyst. A later study suggested that chlorine radicals might be responsible instead [*Science* (2021) 10.1126/science.abd8408]. An *et al.* now present thorough spectroscopic, kinetic, and theoretical evidence for the original alkoxy hypothesis, suggesting that the uncertainty hinged on variation of the cerium/alcohol ratio. —JSY

J. Am. Chem. Soc. (2023)
10.1021/jacs.2c10126

PHYSICS

What it takes to go flat

Solids in which interactions among electrons dominate over their kinetic energy often host exotic many-body states. A recent prominent example is magic angle twisted bilayer graphene, in which the angle between the two graphene layers controls the nature of the system. At close to one degree, the bilayer acquires a nearly flat energy band and can exhibit superconductivity. Sheffer *et al.* studied theoretically the conditions under which related systems featuring Dirac cones can have zero Dirac velocity and perfectly flat bands. The researchers found that symmetries play a crucial role, providing the necessary conditions for the system to reach an interaction-dominated state by tuning a small number of parameters. —JS

Phys. Rev. X (2023)
10.1103/PhysRevX.13.021012

regulatory T cells and immunoglobulin A. Importantly, the animals also showed increased susceptibility to *Salmonella* infections, indicating that changes in the microbiota during development are critical for the establishment of an effective immune system. —MMa

Cell Host Microbe. (2023)
10.1016/j.chom.2023.03.006

HYBRIDIZATION

What underlies successful hybridization?

Interspecific hybridization can be an important source of genetic variation but can result in infertile offspring. Why and when hybridization occurs, and whether offspring have improved or reduced fitness, has long been an area of interest, particularly in plants. Brown *et al.* took advantage of extensively studied flowering plants in the United Kingdom to identify genetic and nongenetic traits associated with hybridization. They found that hybridization was reduced with greater phylogenetic distance between parental plants, and that having different ploidy was not always a barrier to hybridization. Increased range overlap was associated with increased hybridization, though to a smaller degree than expected. —CNS

Proc. Natl. Acad. Sci. U. S. A. (2023)
10.1073/pnas.2220261120

TRANSCRIPTOMICS

Community coordination in biofilms

Bacterial communities, even those composed of a single species, can be heterogeneous in terms of their metabolism, especially within biofilms, in which there can be nutrient, oxygen, light, or pH gradients. Wang *et al*. developed RAINBOW-seq, a transcriptomics method for probing bacterial biofilm communities that uses temporal labeling with a growth-dependent fluorescent dye. Spatial encoding occurs because the interior of the biofilm grows at a much slower rate. Studying *Escherichia coli* in minimal medium, the authors showed that transmembrane transporters are up-regulated throughout biofilms, and, based on the known transporter substrate profiles, propose that resource sharing and nutrient cycling might be involved in cell survival in the interior of the biofilm. —MAF

Nat. Chem. Biol. (2023)
10.1038/s41589-023-01282-w

REVIEW SUMMARY

FERROELECTRICS

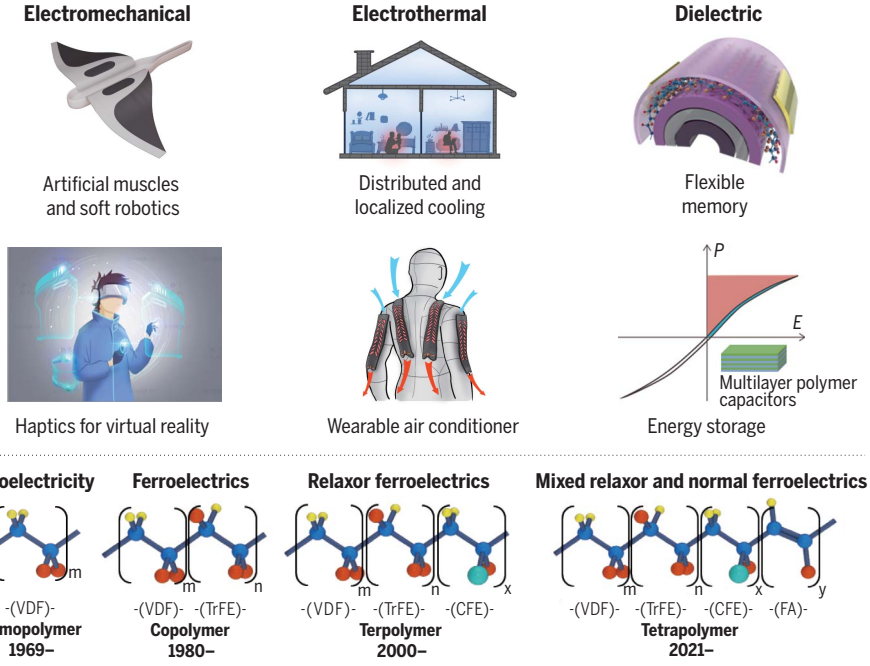
Fluoropolymer ferroelectrics: Multifunctional platform for polar-structured energy conversion

Xiaoshi Qian†, Xin Chen†, Lei Zhu, Q. M. Zhang*

BACKGROUND: Polymeric ferroelectrics are distinguished by their high pliability, easy fabrication into complicated shapes, mechanical robustness, and polar active nature. Ferroelectricity in polymers was discovered around the 1970s in poly(vinylidene fluoride), which has served as a platform for efficient cross-coupling between electrical, mechanical, and thermal energies. Such ferroelectric soft materials and their polar active derivatives undergo a change in electrical polarization in response to general forces (mechanical stresses or temperature changes) and vice versa, enabling a series of physical effects, including piezoelectric and electrostriction, electrocaloric and pyroelectric, and a variety of dielectric and ferroelectric effects. These multifunctional polymeric materials are suitable for many different applications in portable, miniaturized, and wearable electroactive devices applied at human–machine interfaces because of their easy processability into thin, light, tough, and pliable films and fibers.

ADVANCES: Polymer ferroelectrics have exhibited marked improvements in electromechanical coupling efficiency, electrostrictive strain, electrocaloric heat-pumping capability, and lifetime, which have substantially boosted the development of practical applications based on these polar soft materials owing to the facile application of defects in tuning and controlling the polarization processes at the monomeric, macromolecular, and morphological structure levels. For the first time, the piezoelectric and electromechanical coupling factors of fluorinated alkyne (FA)–modified relaxor ferroelectric tetrapolymers have surpassed those of lead zirconate titanate (PZT) piezoceramics, the presently most widely used piezoceramics in the world. Coupled with the progressive 4% electrostrictive strain under a low electric field of 50 MV/m, this advancement represents a step forward in developing efficient wearable sensory and haptic devices and soft robots. Additionally, advances in ferroelectric-based electrocaloric polymers have led to large electrocaloric cooling of >7.5 K under ultralow electric fields without fatigue. These ferroelectric polymers can offer customized, energy-efficient solutions to curb the CO_2 emissions of current commercial heat pumps, air conditioners (ACs), and refrigerators, which are responsible for 60% of building emissions. In addition, the most recently reported electroactive fabrics demonstrate versatile strategies for integrating these polymer ferroelectrics at human–machine interfaces in this envisioned low-carbon society.

OUTLOOK: Understanding and then tailor-making the structures and polarization responses of polymeric ferroelectrics to obtain respective functionalities are critical for the development of these polymeric systems. Given their rich underlying chemistry, FA-modified relaxor ferroelectric polymers are likely still in their infancy. Defect modifications on the molecular scale provide a plethora of methods to manipulate the polar structures and field-induced phase transitions on demand. Considering the vast pool of monomers and nanoscale extrinsic inclusions that can be selected, defect modification in polymer ferroelectrics remains largely unexplored and holds great possibility for contributing to green, smart, and meta lifestyles. Further identifying and understanding the various polarization mechanisms and processes for each functionality at multiple scales will be accomplished by utilizing the current advanced, in situ characterization and simulation tools at our disposal. For different cross-couplings and correlated applications, materials should be fine-tuned to exhibit their respective collection of optimized properties. Several mutual challenges should be addressed, including realizing low-field operation, a long lifetime, viable strategies for integration and mass production, and so on. Considering the commercially available processes for polymeric films, multilayer capacitors, fibers, and fabrics, these flexible ferroelectrics are expected to play a key role in haptic, sensory, and robotic applications in the metaverse, serve as a solid-state refrigerant for flat-panel and/or wearable ACs, and provide a broad range of localized, bodily sensations and tactile effects currently unavailable on the market. ■



Multiphysical cross-couplings and enabled applications offered by polymeric ferroelectrics. Advances in ferroelectric polymers—from homopolymers to tetrapolymers—have promoted applications that harness the power of efficient electromechanical, electrothermal, and a variety of dielectric interactions. The synergy of these physical effects has stimulated the fast development of flexural active systems, wearable refrigeration, flexible memories, and high-efficiency and compact energy storage over the past four decades. VDF, vinylidene fluoride; TrFE, trifluoroethylene; CFE, chlorofluoroethylene.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: qxz1@psu.edu
†These authors contributed equally to this work.
Cite this article as X. Qian *et al.*, *Science* **380**, eadg0902 (2023). DOI: 10.1126/science.adg0902

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adg0902

REVIEW

FERROELECTRICS

Fluoropolymer ferroelectrics: Multifunctional platform for polar-structured energy conversion

Xiaoshi Qian[1]†, Xin Chen[2]†, Lei Zhu[3], Q. M. Zhang[2,4]*

Ferroelectric materials are currently some of the most widely applied material systems and are constantly generating improved functions with higher efficiencies. Advancements in poly(vinylidene fluoride) (PVDF)–based polymer ferroelectrics provide flexural, coupling-efficient, and multifunctional material platforms for applications that demand portable, lightweight, wearable, and durable features. We highlight the recent advances in fluoropolymer ferroelectrics, their energetic cross-coupling effects, and emerging technologies, including wearable, highly efficient electromechanical actuators and sensors, electrocaloric refrigeration, and dielectric devices. These developments reveal that the molecular and nanostructure manipulations of the polarization-field interactions, through facile defect biasing, could introduce enhancements in the physical effects that would enable the realization of multisensory and multifunctional wearables for the emerging immersive virtual world and smart systems for a sustainable future.

Although functional inorganic materials such as semiconductors, piezoelectrics, and ferroelectrics often have higher performance metrics than polymers, arising from crystalline order, polymers are attractive alternatives because of the ease of fabrication, mechanical robustness, lower weight and cost, and acoustic impedance matching to tissues and water. Ferroelectrics are insulators that exhibit additional internal polarization ordering, called spontaneous polarization P_s, whose direction can be reversed by applying an electric field (Fig. 1A). Piezoelectricity is a linear electromechanical effect in which an applied electric field E can induce changes in the mechanical stress X in a material and vice versa (Fig. 1B). Piezoelectrics are used in actuators, transducers, sensors, and solid-state motors (*1*). While not all piezoelectric materials are ferroelectric, ferroelectrics exhibit the strongest piezoelectricity among the known piezoelectrics. Ferroelectrics also find applications in high-permittivity capacitors, nonvolatile memories, and electro-optical devices (*1*, *2*).

These effects were first discovered in inorganic materials and later in polymers. Piezoelectricity in mechanically stretched poly(vinylidene fluoride) (PVDF) was reported by Kawai in 1969 (*3*), and ferroelectricity in the 1970s (*4*). The development of P(VDF-TrFE) in the early 1980s, in which VDF is vinylidene fluoride and TrFE is trifluoroethylene (*5*–*8*), enabled stronger piezoelectricity and ferroelectric responses without the need for mechanical drawing (*9*). These polymer ferroelectrics have found application in commercial products such as wearable sensors, deformable and flexible transducers for medical imaging, underwater navigation, soft robots, and actuators (*9*).

Polarization in normal ferroelectrics is present as macroscopic domains. Breaking macroscopic polar domains into nanoscale polar regions through defects and dopants transforms a normal ferroelectric into a relaxor ferroelectric that exhibits a large and reversible (hysteresis-free) electric field–induced polarization change over a broad temperature range (Fig. 1C) (*10*). In all ferroelectrics, the direct coupling of the polarization in the material with external stimuli such as mechanical stress X, temperature T, magnetic field, and optical signals is what leads to a myriad of functional properties and applications. Thus, the discovery and development of relaxor ferroelectricity in the late 1990s and early 2000s in defect-modified P(VDF-TrFE) polymers such as P(VDF-TrFE-CFE) terpolymers (where CFE is chlorofluoroethylene) created a design paradigm for polymer ferroelectrics with useful properties such as giant electrostriction and a giant electrocaloric effect (ECE) owing to the electric field–induced large and reversible polarization changes between the nonpolar molecular conformations and polar conformation (*11*, *12*). Electrostriction is a nonlinear electromechanical effect in which the mechanical strain x and stress X are proportional to the square of the electric field (and polarization) (*13*). Giant electrostriction expands the applications of ferroelectrics to artificial muscles, better-performing soft robots, and many other electromechanical applications. The ECE is the electric field–induced temperature change in a material (Fig. 1D) (*11*, *12*, *14*, *15*). The ECE is attractive for solid-state refrigeration that emits zero greenhouse gas and is compressor free.

The easy fabrication, mechanical robustness, and acoustic impedance matching to tissues and water of these ferroelectric polymers, coupled with advanced manufacturing, have led to developments and potential applications in fiber wearables, flat-panel and flexible air conditioners (ACs), biomedical devices, artificial muscles, soft robots, emissive energy sensing, and imaging. Several products in wearables, biomedical devices, and soft robots have been commercialized, and others are still under research development (*16*–*26*).

Cross-coupling phenomena

The coupling between polymers' electric and mechanical energies in the form of the piezoelectric effect (*3*–*9*, *16*–*18*, *20*, *27*) is their most widely used feature. The piezoelectric effect is a linear electromechanical effect (*1*, *28*), where a mechanical strain x and stress X generate a charge or voltage output and vice versa

$$D = dX \text{ and } x = dE \quad (1)$$

where E is the electric field, D ($\approx P$ in ferroelectrics) is the surface charge density (electric displacement), and d is the piezoelectric coefficient. The electromechanical coupling factor k is another key parameter, and k^2 (or gdY, where g is the piezoelectric voltage coefficient and Y is the elastic modulus) measures the energy conversion efficiency between electric and mechanical forms (*1*, *2*, *27*).

Electrostriction is another nonlinear electromechanical effect that exists in all insulators, with

$$x = QP^2 \quad (2)$$

where Q is the electrostrictive coefficient. It has gained increased interest after the discovery of giant electrostriction in polymer relaxor ferroelectrics (*11*, *13*, *29*). In ferroelectrics, the piezoelectric effect is considered to arise from the electrostriction under a remanent polarization bias (*13*, *27*).

The cross-coupling between polarization and temperature in insulating dielectrics that have polar ordering generates the pyroelectric effect (*1*, *15*). In simple terms, the pyroelectric effect refers to the polarization change caused by a temperature variation, $\left(\frac{\partial P}{\partial T}\right)_E$, whereas the reverse process, a reversible entropy (S) or temperature (T) change generated by an electric field change, $\left(\frac{\partial S}{\partial E}\right)_T$, for example, the ECE, exists in all insulators. The ECE is usually characterized by an electric field–induced isothermal

[1]State Key Laboratory of Mechanical System and Vibration, Interdisciplinary Research Centre, and MOE Key Laboratory for Power Machinery and Engineering, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China. [2]Materials Research Institute and Department of Materials Science and Engineering, The Pennsylvania State University, University Park, PA 16802, USA. [3]Department of Macromolecular Science and Engineering, Case Western Reserve University, Cleveland, OH 44106, USA. [4]School of Electrical Engineering and Computer Science, The Pennsylvania State University, University Park, PA 16802, USA.
*Corresponding author. Email: qxz1@psu.edu
†These authors contributed equally to this work.

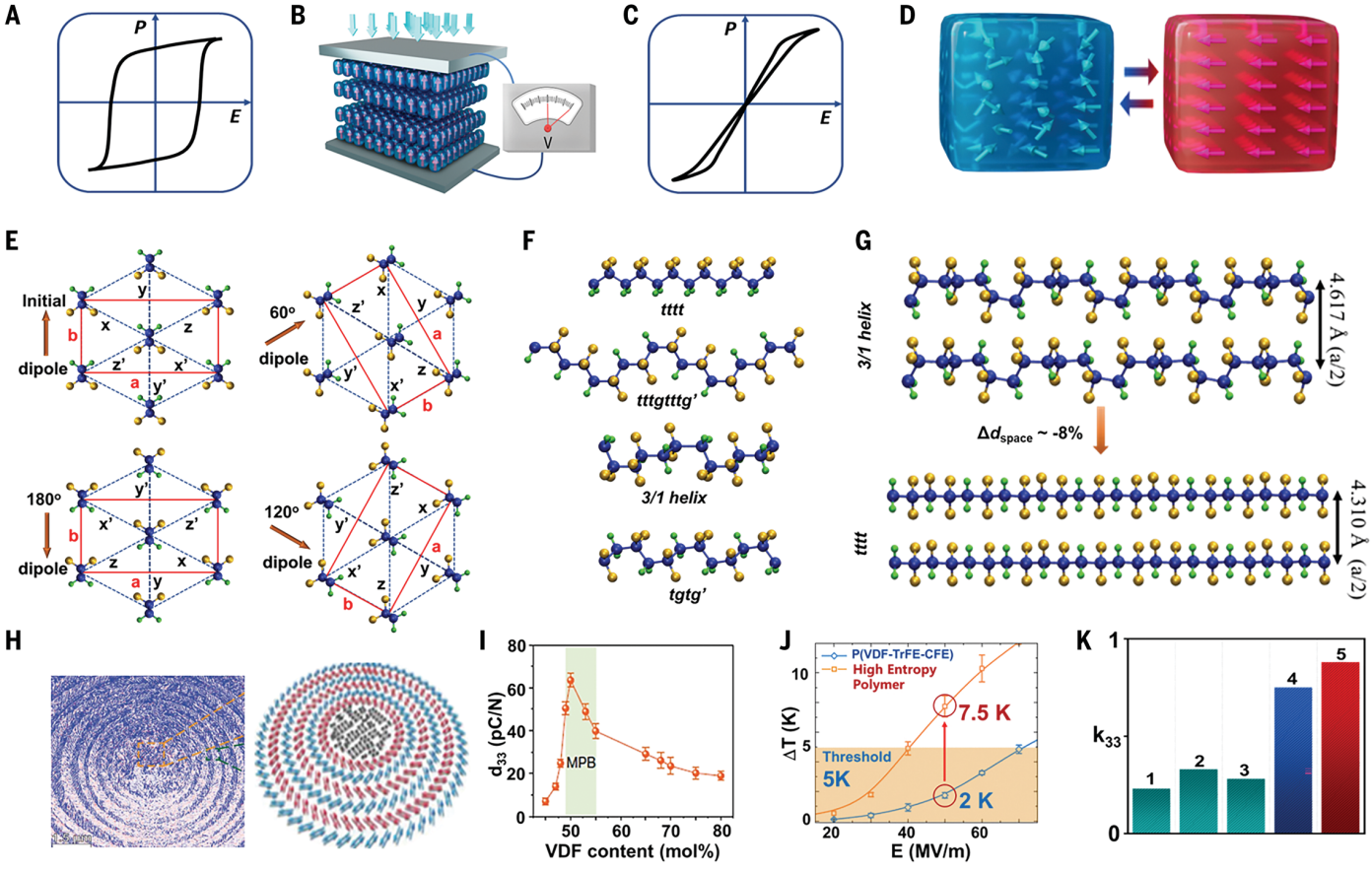


Fig. 1. Underlying mechanism in polymer ferroelectrics. (**A**) Typical *P-E* loop of normal ferroelectric polymers. (**B**) Schematic of piezoelectricity. (**C**) Typical *P-E* loop of a relaxor ferroelectric polymer. (**D**) Schematic of electrothermal coupling. (**E**) Schematics of dipole rotation in the pseudo-hexagonal crystal structure and orthorhombic unit cell (red line) of β-phase PVDF (*27*). (**F**) Basic conformations, different interchain and intrachain combinations of which form various phases in PVDF polymers. *t*, trans; *g*/*g*′, gauche. (**G**) Schematic of the geometry changes in a conformational change from 3/1 helix to *tttt* (*27*). In (E) to (G), yellow atoms are fluorine, blue atoms are carbon, and green atoms are hydrogen (*27*). (**H**) Toroidal topological texture self-organized in a ferroelectric P(VDF-TrFE) polymer (*47*). (**I**) Enhanced piezoelectric effect at the MPB of P(VDF-TrFE) copolymers (*48*). (**J** and **K**) Giant enhancement of the ECE (J) and EM coupling efficiency (K) in C=C modulated tetrapolymers (*22*, *27*). Numbers one to five in (K) refer to k_{33} in the following ferroelectric materials: 1, P(VDF-TrFE); 2, P(VDF-TrFE) single crystal; 3, P(VDF-TrFE) at the MPB; 4, PZT; and 5, P(VDF-TrFE-CFE-FA) (*27*). [(E) to (G) and (K) adapted with permission from (*27*). (H), (I), and (J) adapted with permission from (*47*), (*48*) copyright © 2018 Springer Nature, and (*22*) copyright © 2021 Springer Nature, respectively.]

entropy change ΔS and an adiabatic temperature change ΔT. From Landau-Devonshire (L-D) phenomenological theory

$$\Delta S = -{}^{1}/_{2}\beta P^{2} \qquad (3)$$

in ferroelectrics, where β is a coefficient in L-D theory that characterizes the electrothermal coupling (*15*). In many cases, the two are related as $T\Delta S = C_E \Delta T$, where C_E is the specific heat of the EC material at a constant E (*1*, *15*).

The ECE in polymers is a more recent discovery than the piezo- and pyroelectric effects. Electrocaloric (EC) refrigeration is a highly promising solid-state alternative for cooling (*30*–*32*) that emits zero greenhouse gases and is compressor free, scalable, and compact (*32*). Although the ECE was observed in Rochelle salt in 1930 (*33*, *34*), ECE studies in the past century were performed on inorganic ferroelectrics, where the small ECE observed was not sufficient for practical use (*35*–*37*). Only recently has EC polymer research discovered that facile defect modification of PVDF-based relaxor ferroelectric polymers can lead to a giant room-temperature ECE at ultralow electric fields (*14*, *22*, *38*, *39*).

Polarization responses and coupling to external stimuli

In ferroelectrics, the coupling (interaction) between the polarization and various external stimuli such as mechanical stress X, temperature T, magnetic field B, and light signals is what generates various cross-coupling phenomena (ferroelectrics are multifunctional). Polarization changes in ferroelectrics originate from different sources and processes that generate different cross-coupling phenomena. Although exhibiting a similar polarization change, the domain switching of an antiparallel polar arrangement, which generates very little electromechanical (EM) coupling, is fundamentally distinct from the field-induced polar reorientation in a three-dimensional (3D) polar structure, which produces strong EM coupling (*40*). In addition, nonvolatile memory demands an irreversible P_s arising from the existence of macroscopic polarization domains, whereas giant EC cooling materials were discovered among relaxor polymers, in which the polarization responses arise from highly polar-disordered to ordered phases that exhibit high reversibility and efficiency (Fig. 1D) (*22*).

In polymer ferroelectrics, the polarization processes occur at molecular, nanoscopic, mesoscopic, and macroscopic scales. In the ferroelectric phase (β-phase), chains of the all-trans (*tttt*) conformation form a hexagonal crystal structure (Fig. 1E). The polarization change under electric fields occurs through successive

60° dipole rotations (*9*). Here, a 180° (y' direction to y direction) dipole rotation generates a large polarization change but zero strain. The dipole rotations from the x (60°) and z' (120°) orientations to the y direction generate both a polarization change and mechanical strain. Bernholc and colleagues carried out a density functional theory calculation of the EM effect realized through these polarization rotations and showed an electrostrain of −1.1% and a small electrostrictive coefficient Q_{33} of −2.9 m^4/C^2, which is nearly the same as the experimentally observed Q_{33} in P(VDF-TrFE) polymers (*27*).

Adding small amounts of defects such as monomers (e.g., CFE), which are bulkier in size than the VDF and TrFE monomers in VDF-TrFE polymer chains, converts the all-trans chain conformation of the β-phase to a chain morphology of mixed conformations of all four types (Fig. 1F), the proportions of which depend on the CFE content. Including ~6 mol % CFE in P(VDF-TrFE) chains eliminates the all-trans conformation and converts the polymer into a relaxor that displays a reversible and large polarization change at room temperature. Moreover, the relaxor polymer generates an electroactuation strain of −7% and a $-Q_{33}$ of >10 m^4/C^2, indicating that the polarization changes in the relaxor terpolymer are more efficient in EM coupling. This can be understood from the large difference in the polymer chain dimensions between the highly polar all-trans bonds and the three other bonds, for example, trans-gauche–trans-gauche′ (*tgtg′*), t_3gt_3g', and 3/1 helix (*tgtgtg* or *tg′tg′tg′*) (Fig. 1F). Switching from these three nonpolar or nearly nonpolar bonds to the *tttt* bonds generates larger electroactuation and a higher $|Q_{33}|$. For example, the conformational change from 3/1 helix to *tttt* can generate an interchain strain (d_{space}) of ~8% (Fig. 1G).

By replacing a small amount (~2 mol %) of bulky CFE with fluorinated alkyne (FA), which has a smaller monomer size than VDF and TrFE, the relaxor P(VDF-TrFE-CFE-FA) tetrapolymer displayed a Q_{33} of −40 m^4/C^2, generating large electroactuation at ultralow electric fields (<50 MV/m) and ultrahigh piezoelectric coefficient and EM coupling at a DC bias as low as 20 MV/m. The tetrapolymer exhibited a diffused critical endpoint transition region at which the energy barriers for switching from nonpolar to polar molecular conformations become small owing to the presence of a small number of polar seeds. Thus, a small change in the electric field induces large electroactuation (*41*). A similar enhancement was also observed in the ECE. A giant ECE at an ultralow electric field was observed in P(VDF-TrFE-CFE-FA) relaxor tetrapolymers with only 0.6 mol % FAs. Moreover, the β coefficient in Eq. 3 of the tetrapolymer is four times that of P(VDF-TrFE-CFE) (*22*). The results for these relaxor polymers reveal the critical importance in facile application of molecular dissimilarities and defects to control different polarization processes and tailor barriers for desired polarization switching to generate the desired material performance at high efficiency.

Among polymer ferroelectrics, including nylon-based ferroelectrics and others (*9*, *42*, *43*), PVDF-based polymers exhibit the best ferroelectric, EM coupling, and electrothermal coupling properties. Moreover, PVDF-based ferroelectric relaxor polymers display the highest reversible room-temperature polarization changes, making them attractive for energy storage and other dielectric applications.

PVDF-based ferroelectrics

PVDF and P(VDF-TrFE) copolymers

PVDF is the first known polymer ferroelectric. Copolymerizing VDF with TrFE stabilizes the ferroelectric β-phase without drawing, and P(VDF-TrFE) shows a ferroelectric-paraelectric (F-P) transition at >18 mol % TrFE (*9*, *44*). VDF has a dipole moment of 3.0 D in the β-phase. TrFE has a dipole moment of approximately half that of VDF in the β-phase. Therefore, increasing the amount of TrFE in the copolymers will weaken the ferroelectricity, as reflected by the decrease in the F-P transition temperature with the TrFE content; and at a TrFE content of >45 mol %, the F-P transition becomes continuous (*9*). Additionally, P(VDF-TrFE) is a semicrystalline polymer, and the ferroelectricity is from the crystalline phase. Whereas the crystallinity of PVDF is ~50%, P(VDF-TrFE) at compositions near 70/30 mol % can reach a crystallinity of >90%. The copolymers near this composition exhibit the best ferroelectric and EM properties (*9*). In addition to the crystallinity, the crystalline orientation (and mechanical drawing) also has a big influence on ferroelectric and other functional properties (*9*, *27*, *45*). Moreover, as early as 1998, Bune *et al.* reported 2D ferroelectrics for Langmuir-Blodgett P(VDF-TrFE) films two molecular layers thick (*46*).

All of these properties make P(VDF-TrFE) a useful material system for developing an understanding of ferroelectricity in polymers. For example, polar topological textures have become an emerging research field for exotic phenomena and potential application in reconfigurable electronic devices. In 2021, Guo *et al.* reported a toroidal topological texture self-organized P(VDF-TrFE) copolymer that exhibits a concentric topology with anti-coupled chiral domains (Fig. 1I) in spin-coated thin films with polymer chains vertically aligned and a permanent polarization lying along an in-plane direction of the film (*47*). The interplay among the elastic, electric, and gradient energies in the highly strained polymer films results in continuous rotation and toroidal assembly of the polarization perpendicular to polymer chains, whereas relaxor behavior is observed along polymer chains. Moreover, the toroidal polar topology of the films also exhibits the ability to microscopically manipulate terahertz waves, which might find application in terahertz raster scanning and spatial light modulators (*47*).

The piezo d_{ij} coefficients and EM coupling efficiency of PVDF and P(VDF-TrFE) are approximately one order of magnitude lower than those of their ceramic counterparts (Table 1). In ferroelectrics, the morphotropic phase boundary (MPB) is a transition region in the phase diagram bridging two nearly energetically degenerate phases with distinct symmetries, thus lowering the polarization rotation barriers and consequently enhancing the piezoelectric response. In 2018, Liu *et al.* reported the discovery of an MPB in P(VDF-TrFE) copolymers, and the P(VDF-TrFE) copolymers at MPB compositions had a piezoelectric d_{33} of −63.5 pm/V, double that of compositions away from the MPB, albeit still far below the d_{33} of 650 pm/V for the inorganic counterpart (Fig. 1I) (*48*). In contrast to their inorganic counterparts, the d_{33} and Q_{33} of PVDF-based ferroelectric polymers exhibit negative signs because of the chain conformations and crystal structures of polar and nonpolar phases (*27*, *49*).

In contrast to the piezoelectric effect, P(VDF-TrFE) copolymers, which have intrinsic "polar disorder" in the paraelectric phase, should generate a large ECE near the F-P transition. A paraelectric is a dielectric that lacks polar ordering and transitions to a ferroelectric at certain temperatures or under mechanical stresses. Using Eq. 3, $\Delta S = -\frac{1}{2}\beta P^2$, it leads to ΔS = 96 J/(kg·K) and ΔT = 26 K at the F-P transition of 100°C of P(VDF-TrFE) copolymers, orders of magnitude higher than observed in ceramics (*15*, *50*). Experimental studies of the ECE in P(VDF-TrFE) near the F-P transition confirmed this in the late 2000s (*14*, *51*).

Relaxor ferroelectrics and their nanocomposites

Relaxor ferroelectrics were discovered in the late 1990s, creating a new polymer ferroelectric design paradigm and offering promising properties in several technological areas (*11*, *12*, *52*, *53*). Incorporating molecular defects such as a small molar percentage of monomers, for example, CFE and chlorotrifluoroethylene (CTFE), which are bulkier than VDF and TrFE in P(VDF-TrFE) normal ferroelectrics, effectively modulated the crystalline and domain structures and eliminated undesired ferroelectric hysteresis. The broad dielectric transition peak of the resulting relaxor ferroelectric polymer moved to room temperature, with a high dielectric constant $K > 50$, which is the highest among all known polymers at room temperature, while maintaining a large and reversible polarization change (*11*, *12*).

Table 1. Advances in EM performance. Summary of piezoelectric effects and EM couplings for PVDF-based polymers and comparison with ceramic PZT (*2*, *9*, *27*).

Material	DC bias (MV/m)	d_{33} (pm/V)	k_{33}	K ($\varepsilon/\varepsilon_0$)	g_{33} (V·m/N)	Y (GPa)	Acoustic impedance [Gg/(m^{2}s)]
PVDF	0	−31.5	0.11	11	−0.32	3	3
P(VDF-TrFE)	0	−33.5	0.3	8	−0.47	3.3	4
PZT piezoceramic	0	650	0.75	3800	0.019	50	33.7
P(VDF-TrFE-CFE-FA) 63/29.7/5.4/1.9	40	−858	0.74	34	−2.85	0.22	–
P(VDF-TrFE-CFE-FA) s-63/29.7/5.4/1.9*	40	−1050	0.88	36	−3.29	0.22	–
P(VDF-TrFE-CFE-FA) 63.6/30/4.4/2	20	−1177	0.71	64	−2.08	0.21	–

*s-63/29.7/5.4/1.9 is a uniaxially stretched film

In most EM applications, piezoelectric materials with large electroactuation strains (large shape changes) are highly desired. For ferroelectric polymers, the small piezoelectric coefficients, even under a high electric field of 150 MV/m, limit the electroactuation strain to <−1%. The PVDF-based relaxor polymers were the first among electroactive polymers that generated ultrahigh electroactuation of >−7% at room temperature (and $\Delta x/\Delta E$ > 400 pm/V), much higher than that of the normal ferroelectric P(VDF-TrFE) (*12*, *29*).

A giant ECE was discovered in PVDF-based relaxor polymers at room temperature, ΔT = 20 K [and ΔS = 100 J/(kg·K)], in electron beam–irradiated P(VDF-TrFE) copolymers and ΔT = 16 K [and ΔS = 80 J/(kg·K)] in P(VDF-TrFE-CFE) terpolymers at an electric field of 150 MV/m (*38*, *39*). The relaxor P(VDF-TrFE-CFE) and P(VDF-TrFE-CTFE) terpolymers have been scaled up and are now commercially available from Arkema Piezotech. These EC polymers enabled EC device studies, which demonstrated the potential of EC cooling, especially for wearable, localized, and distributed cooling (*23*, *24*, *32*, *54*–*56*). The large ECE and wide temperature window rendered the terpolymer an effective base matrix for developing various types of nanocomposites in seeking a larger ECE, better field efficiency, and better overall thermal and mechanical parameters than the neat polymers (*19*, *57*, *58*). At higher electric fields, >200 MV/m (Table 2), a much larger ECE can be generated in relaxor ferroelectric polymer nanocomposites (*58*, *59*).

Polymer relaxors also provide another avenue for studying relaxor phenomena. Despite more than five decades of intensive research, polymer relaxors remain one of the least understood material families among ferroelectric materials. Liu *et al.* showed that the relaxor behavior of ferroelectric polymers originates from conformational disorder, differing completely from classic perovskite relaxors, which are typically characterized by chemical disorder and the presence of nanoscale polar domains (*10*, *60*).

Table 2. Advances in the EC effect. Summary of the EC performance of PVDF-based polymers and their inorganic nanocomposites. BNNs, boron nitride nanosheets; BST, barium strontium titanate; nfs, nanofibers; BFBZT, $BiFeO_3$ nanoparticles in $Ba(Zr_{0.21}Ti_{0.79})O_3$.

Material	T (K)	ΔE (MV/m)	ΔS (J/kg·K)	ΔT (K)	Reference
Terpolymer P(VDF-TrFE-CFE) 59.2/33.6/7.2 mol %	303	150	80	16	(*39*)
Irradiated P(VDF-TrFE)	323	50/180	28/160	6.2/35	(*51*)
P(VDF-TrFE-CFE)/BNNSs/BST1/2/3	273~333	200	170	35	(*58*)
P(VDF-TrFE-CFE)/BST_nfs	303	200	105	44.3	(*59*)
P(VDF-TrFE-CFE)/BST_nfs	300	70	49	11.3	(*124*)
P(VDF-TrFE-CFE)/BFBZT	303	75	78	13.8	(*19*)
P(VDF-TrFE-CFE)–ceramic scaffold	300	60	26.8	6	(*125*)
PBZ-nfs/PVDF	243	150	52.7	14	(*126*)
P(VDF-TrFE-CFE)/BTO	308	125	2.6	–	(*127*)
P(VDF-TrFE-CFE-DB)-BST-15	RT	100	6.5	–	(*128*)
P(VDF-TrFE-CTFE-FA)	313	80	12.5	2.5	(*65*)
P(VDF-TrFE-CFE-FA)	303	50/80	38/70	7.5/14	(*22*)

FA defect–modified relaxor ferroelectrics

Although the giant ECE and giant electroactuation in relaxor polymers such as P(VDF-TrFE-CFE) terpolymers have generated great interest, a key limitation for applications has been electric breakdown, for example, the voltage is so high that the device arcs and fails. The large ECE (Table 2) and large electrostrain reported are measured near the dielectric breakdown of small samples. For reliable device operation, the applied voltage (and electric field) should be far below the electric breakdown for the materials used in the device. For example, for P(VDF-TrFE-CFE) terpolymers, this limiting field is ~60 MV/m. The EC-induced ΔT under this field for the best terpolymers is 3 K (*23*, *24*, *39*). For caloric refrigeration to provide meaningful cooling, an EC ΔT > 5 K at the limiting field is required (*22*–*24*, *61*–*63*).

A giant ECE under low electric fields was recently reported in a high-entropy EC polymer. By replacing a small amount (0.6 mol %) of bulky CFE in P(VDF-TrFE-CFE) with double bonds (FA), the resulting polymer exhibited a large ECE (ΔT > 7 K) under 50 MV/m (Fig. 1J) (*22*). Moreover, no sign of fatigue appeared after 1 million electric cycles (*22*). The simulation results showed that the solid-state refrigerator operating the high-entropy polymer through an active regeneration cycle exhibited a high cooling power density greater than 10 W/cm^3 (and 5 kW/kg) at zero temperature

span, T_{span}, or a T_{span} of >50 K at no load (*22*, *56*). The simulation results suggested that the EC-based air conditioning/heat pump (AC/HP) system has the potential to reach a similar refrigeration power/refrigerant charge ratio to that of a vapor compression cycle–based AC/HP system, which requires between ~1 and ~1.6 kg of refrigerant charge to offer between ~5 and ~6 kW of heating and cooling capacity (*64*).

In addition to the FA-modified P(VDF-TrFE-CFE) terpolymers, Le Goupil *et al.* reported ECE enhancement for FA-modified P(VDF-TrFE-CTFE), from ΔT = 1.3 K for the pristine terpolymer to ΔT > 2 K with 5 mol % FAs, a 60% enhancement (*65*). The CTFE-based relaxor terpolymer and tetrapolymer generate much smaller ECEs than their CFE-based counterparts, indicating the critical importance of properly designed defects in influencing and controlling the polarization responses of PVDF-based ferroelectric polymers.

Ultrahigh piezoelectricity and EM coupling efficiency were observed in FA-modified P(VDF-TrFE-CFE) at ~2 mol % FAs (*27*). Under a low DC bias of 40 MV/m, the mechanically stretched tetrapolymer exhibits an EM coupling factor k_{33} (an indicator of the efficiency of EM energy conversion) of 88% and a piezoelectric coefficient d_{33} of −1050 pm/V (Table 1 and Fig. 1K). For the tetrapolymer P(VDF-TrFE-CFE-FA) (63.6/30/4.4/2 mol %), even at a 20 MV/m DC bias, d_{33} = −1177 pm/V and a coupling factor of 71% can still be obtained. These values are higher than those of the most widely used piezoelectric material, lead zirconate titanate (PZT). Considering that these tetrapolymer films can be easily made into films less than 3 μm thick, the DC bias voltage for 20 MV/m is <60 V. Most ferroelectric devices are in the multilayer (ML) film form, which decouples the dielectric layer thickness from the device thickness (*63*, *66*). Moreover, ~2 mol % FAs in the P(VDF-TrFE-CFE) 68/32/7.3 mol % terpolymer enhanced the electroactuation by more than three times at 60 MV/m compared with the neat terpolymer (*27*).

Electromechanical applications of PVDF-based polymers

PVDF-based ferroelectric polymers have been targeted for wearable and implantable mechanical energy harvesters and green energy harvesters since the early 1980s. Examples include an implantable physiological power supply driven by the spontaneous breathing of a dog (*67*) and a PVDF stave inserted beneath the feet for generating electrical energy during walking (*68*). The low acoustic impedance of ferroelectric polymers is well matched to water and tissues and enables efficient power transmission between acoustic and ultrasound transducers and tissues as well as water, which is desirable for sensing and imaging (*16*, *69*). For electroactuation, relaxor polymers with an elastic modulus of >0.2 GPa and an actuation strain of >5% at low electric fields (*27*) perform better than ferroelectric ceramics and are better suited for applications that require a large strain and a high force level in free-standing films at low electric fields. Such applications include artificial muscles, soft robots, and haptic feedback.

Energy harvesters and charge generators

Extensive studies have been performed with PVDF-based polymers for energy harvesting (*70*–*75*). For a piezoelectric energy harvester, the output energy is proportional to the square of the EM coupling factor. To improve the energy harvesting performance, polymer composites with piezoelectric ceramic nanofillers have been shown to be effective. The selection of nanofillers should focus on high-d but low-K ceramics. For example, a PVDF composite with 5 vol % oriented $BaTi_2O_5$ nanorods, which has a K < 100 and a d_{33} comparable to that of $PbTiO_3$, exhibits twice the peak output power density (0.82 μW/cm^2) than that of neat PVDF (*76*).

Synergy of piezoelectric and triboelectric effects is another effective way to reinforce energy harvesting. The remnant polarization of ferroelectric polymers can effectively strengthen the triboelectric effect, and Yousry *et al.* reported, together with the piezoelectricity of ferroelectrics, an effective piezoelectric charge coefficient $\Delta Q/\Delta F$ = −1065 pC/N and a peak output power density of 0.21 μW/cm^2 at an acceleration of 3G (where G = 9.8 m/s^2, the gravitational acceleration) (*77*).

Polymer fiber wearables

For wearable piezoelectric polymer devices (*74*, *78*, *79*), PVDF-based polymers can be fabricated into fibers and then into clothing. Electrospun PVDF fibers (*18*, *80*, *81*) can be easily incorporated into composite materials to further enhance the performance. The surface of nanofillers was proposed to generate electrostatic interactions with PVDF chains that facilitate the formation of the β-phase, thus enhancing the piezoelectric performance. For example, adding both graphene nanosheets and barium titanate ($BaTiO_3$, or BTO) nanoparticles resulted in a >90% content of the β-phase of PVDF fibers, which can work as a high-performance flexible piezoelectric nanogenerator (PENG) (Fig. 2A). The device exhibited an output voltage of ~11 V with a maximum electric power of 4.1 μW under a loading frequency of 2 Hz and a strain of 4 mm (*82*).

A bioinspired fabric containing P(VDF-TrFE)/$BaTiO_3$ nanocomposite fibers as acoustic sensors exhibits good performance in sound detection. The nanocomposite fiber exhibits a d_{31} of ~45 pC/N (pm/V), compared with a value of 20 pC/N for the pure P(VDF-TrFE) fiber (*20*). By mimicking the natural tympanic system (Fig. 2B), a membrane with a combination of high- and low-modulus yarns enables recording of short impulses of acoustic vibrations of only 10^{-7} atmosphere pressure waves. The P(VDF-TrFE) composite fibers on the membrane produce an electrical output that is two orders of magnitude higher than that of the standalone fibers. The resulting minimum sound detection capability is 0.002 Pa (40 dB, the sound pressure level in a quiet library). A sensitivity of 19.6 mV was measured for the fiber-on-membrane, which is comparable to that of the off-the-shelf condenser and dynamic microphones, for which the values are typically 5 mV (*20*).

An emerging area in which polymer ferroelectrics can play a role is face masks (Fig. 2C), which are key personal protective equipment in protecting people from viral infection. The heart of state-of-the-art face masks (e.g., the N95 mask) is the electrostatic polypropylene (PP) polymer fiber air filter, which, through surface charges (electrostatic forces), traps (filters) viruses, prevents them from reaching the face, and increases the efficiency by several orders of magnitude for the same pressure drop compared with pure mechanical filters. However, the low thermal stability (<60°C) of PP electrets limits their shelf life and service life. In addition, the surface charge in PP electrets is several orders of magnitude lower than that in PVDF polymers. Indeed, PVDF polymer fiber meshes, through the piezoelectric effect (mechanical motion during breathing) and pyroelectric effect (temperature change during breathing), have been shown to be able to generate a much higher voltage (and charge) than PP electrets (*21*). Furthermore, PVDF fiber meshes can be easily cleaned (disinfected) by washing. In addition to PVDF fiber meshes, the piezoelectric poly(L-lactic acid) (PLLA) filter has also been shown to possess a high filtration efficiency for human breathing, also due to the piezoelectric charge naturally activated by respiration through the mask (*26*).

Ultrasonic imaging

In high-frequency medical and biological ultrasound imaging, P(VDF-TrFE) is extremely useful because of its low acoustic impedance, excellent flexibility, and ability to exhibit stable piezoelectric/dielectric properties in thin films (micrometer scale) (*16*). Commercial instrumentation of high-frequency ultrasound transducers has also been developed for image recognition, such as the Qualcomm fingerprint scanner. The scanner is constructed by film capacitor field-effect transistors, wherein P(VDF-TrFE) high-density arrays work as the gate dielectric to recognize the pressure with good spatial resolution. The scanner detects 3D details and fingerprint characteristics such as ridges and sweat pores, which cannot be

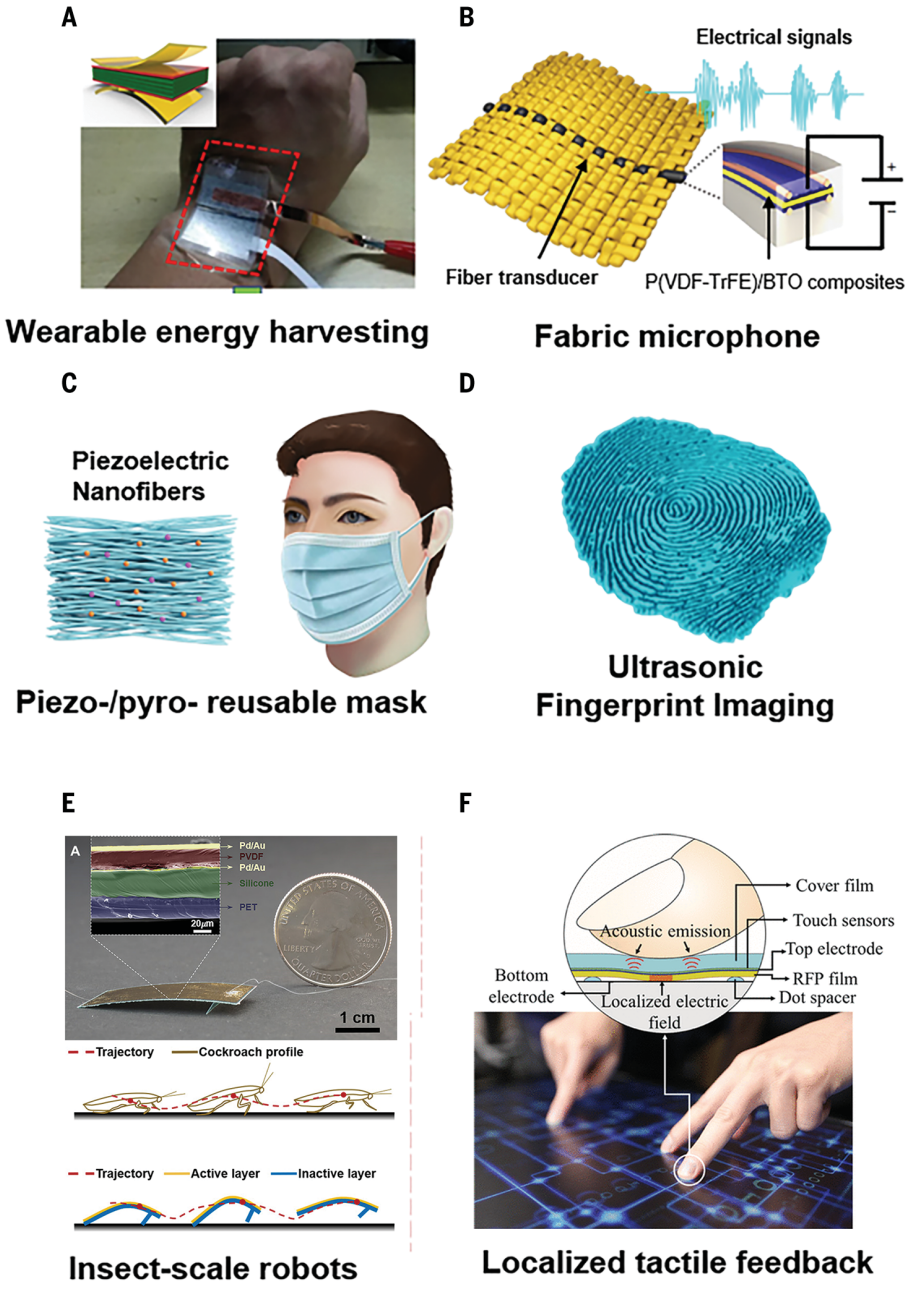


Fig. 2. Advances in EM applications using PVDF-based polymers. (**A**) Optical image of a wearable PVDF-based energy harvester. Schematic structure of a PENG and diagram of its flexibility. The inset shows a schematic of the device (red layer, aluminum foil; green layer, PVDF fibers; gold layer, PET) (*82*). (**B**) Schematic of PVDF/BTO composite fibers integrated with an acoustic fabric (*20*). (**C**) Schematic of a prototype piezoelectric polymer–based reusable mask (*26*). (**D**) Schematic of the finger pattern imagined by the Qualcomm finger scanner. (**E**) (Top) Optical photo showing an insect-scale robot. The inset scanning electron microscopy image shows the cross-sectional view of the prototype robot with different layers of materials. (Bottom) Comparison of the wavelike running paths showing the movement of the center of mass of a cockroach (*17*). (**F**) Schematic of PVDF relaxor–based localized tactile feedback devices (*25*). RFP, relaxor ferroelectric polymer. [(A), (B), (C), (D)/(E), and (F) adapted with permission from (*82*) copyright © 2018 Elsevier, (*20*) copyright © 2022 Springer Nature, (*26*) copyright © 2022 Wiley, (*17*), and (*25*) copyright © 2019 American Chemical Society, respectively.]

detected with capacitive touch-based or optical options and has a thin and compact device size (Fig. 2D) (*83*). Such ferroelectric polymer ultrasonic image recognition technology has broad impacts, including personal IDs, gate access for buildings, smart car ignition, and others by providing higher sensitivity and more-detailed biological information.

Soft robotics

The high ductility and easy fabrication in the desired and complicated form enable the use of PVDF-based polymers in soft robotics. Mobility and robustness are two important features for practical applications of robots. Soft robots made of polymeric materials may potentially achieve both attributes simultaneously. Wu *et al.* reported soft robots based on a curved PVDF unimorph piezoelectric structure, whose relative speed of 20 body lengths per second was one of the fastest reported for artificial insect-scale robots (*17*). The design (Fig. 2E) is inspired by several principles of animal locomotion. Such fast and ultrarobust insect-scale soft robots have potential for application in environmental exploration, structural inspection, information reconnaissance, and disaster relief. The results also showed that the improved piezoelectric performance of the tetrapolymers enables better mobility of soft robots.

Haptic actuators

Emerging virtual and augmented reality technologies are raising the demand for communication between humans and the virtual world, including optical, auditory, and tactile feedback. Haptic technologies enable humans to touch and interact with the contents of virtual environments (*84*). Haptic actuators generate vibrations at frequencies most sensitive for human touch, a few hundred hertz, with a high acceleration of a few *G*s. The large strain (and hence displacement) of the relaxor polymer thin-film actuators provides excellent matching to the required haptic actuators. Compared with traditional actuator technologies that rely on piezoceramics, relaxor ferroelectric polymers can generate >50 times larger strain without mechanical failure. A haptic device commercialized by KEMET with thin-film relaxor polymer actuators can generate a 0.2 mm out-of-plane displacement with a 210 V operation voltage (*85*).

Another design for localized tactile feedback was reported by Duong *et al.* The active layer of the P(VDF-TrFE-CTFE) terpolymer is sandwiched by an electrode and a spacer that separates the layer from touching another electrode (Fig. 2F). The flexible active material and large actuation enable localized tactile feedback. A 200-Hz AC voltage is applied on the electrode. Upon finger pressing, the active layer locally touches the other electrode, resulting in a localized electric field being applied on the terpolymer and triggering its electrostrictive actuation to generate localized vibrational tactile feedback (*25*).

Polymer-based EC refrigeration

As an EC material reversibly heats up and cools down under electric cycles, a refrigeration cycle, such as the Brayton cycle, can be

formed to pump heat from a heat load to a heat sink (*32*). By exploiting the large ECE in commercial relaxor terpolymers, several EC cooling modules have been developed.

Active EC regeneration

Gu *et al.* reported a series of EC devices (Fig. 3A) that use EC polymers and ceramics fabricated in multilayer capacitor structures (*56*, *86*, *87*). To accumulate the heating and cooling effects at the cold and hot ends of the device and further extend the temperature span of the device, active electrocaloric regeneration (AER) in a polymeric EC device was first demonstrated in 2013 (*56*). A regenerator with anisotropic thermal conductivity was designed and fabricated to minimize the heat loss between the hot and cold ends of the chip-sized device. The device exhibited a 6.6 K temperature span at room temperature and a frequency of 1 Hz, which is comparable to that of the all-solid EC device reported by Wang *et al.* in 2020 using multilayer ceramic capacitors (MLCCs) (*88*).

An AER device with counterrotating EC disks has also been developed, numerically and experimentally (Fig. 3B) (*86*, *89*). In this AER counterrotating disk device design, the inactive heat regeneration layers are eliminated, and, instead, the active EC layers stacked in the counterrotating disks serve both the regeneration layers and the active EC working body, thus improving the AER performance. Because of the ease of applying an electric field to the EC unit in the disks, the device geometry can be more sophisticated than that based on magnetocaloric and barocaloric effects (*55*, *90*).

In addition to all-solid EC devices, solid–fluid coupling heat exchange was also used to boost efficient heat transfer. By using the commercial EC terpolymer as a cooling core that directly transfers the heat to the reciprocal air flow, the United Technologies Research Center (UTRC) reported a heat exchanger–free "direct air cooling device" (*91*). Different from the design of all-solid EC devices, a device using solid–fluid contact can drastically reduce the energy loss from the interface thermal resistance and eliminate the passive heat exchanger (*92*, *93*). The prototype developed

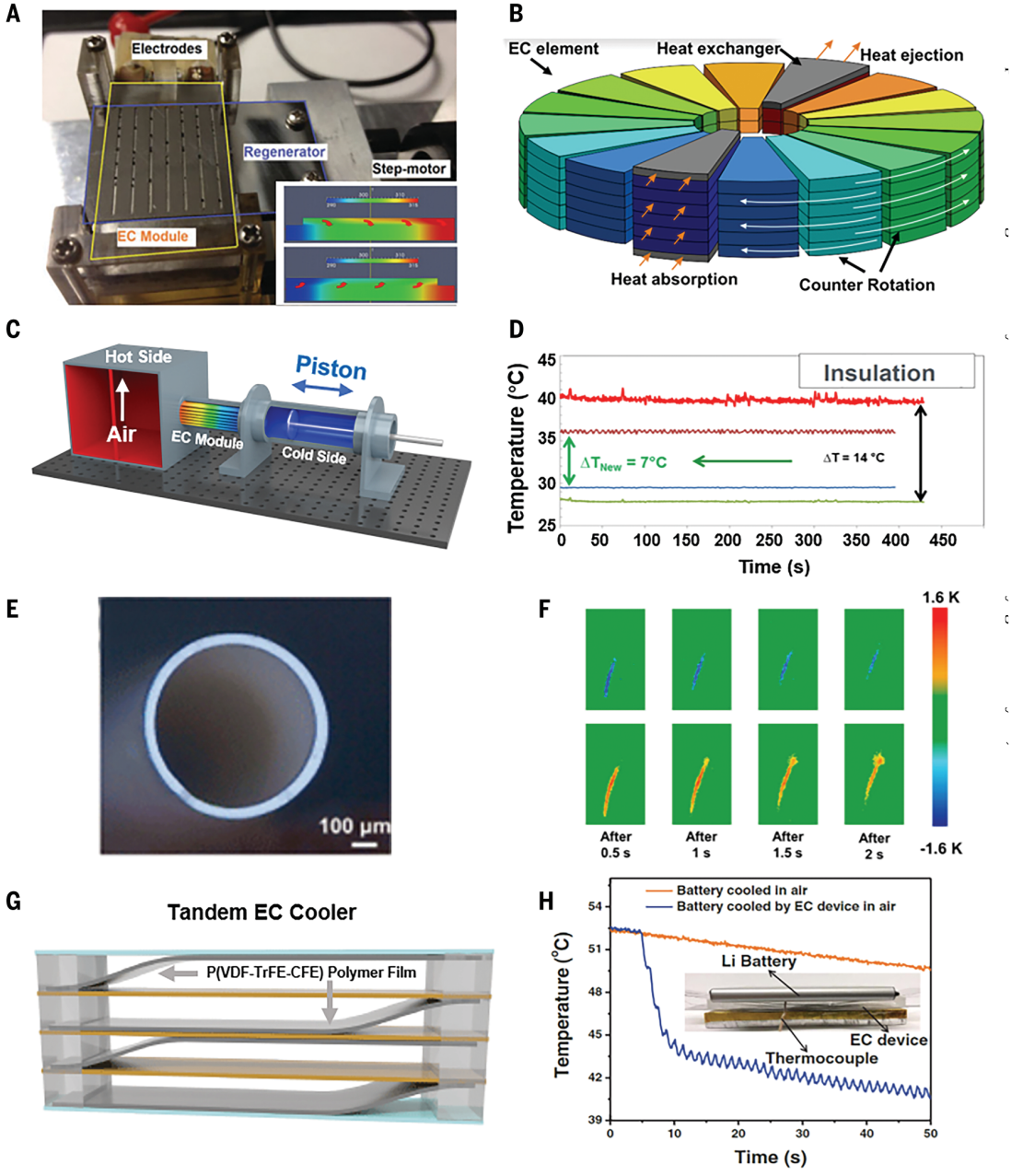


Fig. 3. Advances in EC cooling technologies using PVDF-based polymers. Prototypes of EC refrigerators using ferroelectric polymers as solid-state refrigerants. (**A**) All-solid EC cooling device using electron-beam irradiated copolymers and operating under an AER cycle (*56*). (**B**) Schematic of an ML rotary EC cooling device operating under an AER cycle. (**C** and **D**) Schematics of the EC refrigerator using a fluid (air) as the regenerator (C); a maximum temperature span of 14 K was achieved (D) (*91*). (**E** and **F**) Cooling tube made of relaxor ferroelectric polymers (E) providing heating and cooling to the fluids that pass through it (F) (*94*). (**G** and **H**) Tandem EC prototype using electrostatic oscillation (G) to operate the refrigeration cycle and introduce rapid cooling to a lithium battery (H) (*23*, *24*). [(A), (B), (C)/(D), (E)/(F), and (G)/(H) adapted with permission from (*56*) AIP Publishing, (*91*), (*94*) copyright © 2022 Elsevier, and (*23*) and (*24*) copyright © 2022 Springer Nature, respectively.]

by the UTRC demonstrated a 14°C T_{span} at room temperature (Fig. 3, C and D) (*91*), which is the highest among all EC devices (*23*, *24*, *62*, *88*, *94*, *95*). By ensuring good thermal contact during refrigeration cycling, EC cooling devices are expected to exhibit a large cooling power (>1 kW) using a device design of solid–fluid coupling (*54*). Moreover, solid–fluid heat transfer can be realized by designing EC polymeric tubes that exchange heat with the fluid running through them (Fig. 3, E and F) (*94*).

Electrostatic EC heat pumps

By integrating the electrostatic force and the ECE, an EC cooler without external motors that enhances the device coefficient of performance (COP) is possible. Ma *et al.* reported a flexible cooling device made of a P(VDF-TrFE-CFE) terpolymer stack, in which the EC polymer film stack oscillates between the cold and hot ends, driven by the electrostatic force (*23*). A cascaded polymeric EC heat pump was developed by stacking several cooling units together, which greatly enlarges the temperature span of the heat pump (Fig. 3G) (*24*) and generates cooling of 0.78 W (1100 W/kg cooling power density) at a 2.7 K temperature rise. The device demonstrated the versatility and feasibility of EC cooling technology in wearable devices for providing localized thermal management (Fig. 3H). A wearable cooler that provides an additional cooling of between ~4 and ~5 K was shown to be sufficient for personal comfort (*96*).

Despite the demonstrations of EC devices for compact and wearable cooling, their performance remains far from the requirements for practical EC cooling systems (*23*, *24*, *62*, *88*). Owing to the constraints of electric breakdown, the EC temperature change ΔT in all EC devices (including those with ceramics) is <3 K. This is too small to generate meaningful cooling given various parasitic thermal losses. In contrast, simulation results show an extrapolated device performance with a high cooling power density of 5 kW/kg at zero T_{span} or a T_{span} of >50 K at no load for EC coolers with EC P(VDF-TrFE-CFE-FA) tetrapolymers (ΔT = 7.5 K) (840 W/kg at zero T_{span} and a T_{span} of >45 K at zero cooling power with ΔT = 5.5 K for EC ceramics) (*22*, *62*).

Dielectric applications

Flexible nonvolatile memory

For flexible electronics, PVDF-based nonvolatile memories could offer high data storage density and low operation voltage (*97*–*99*). For example, ferroelectric field-effect transistors (FeFETs) (*100*) and ferroelectric tunnel junctions (FTJs) (*101*) both exhibit longer-lived nondestructive read operations (*102*). A three-terminal memristive device can be created by replacing the gate dielectric in an FET with P(VDF-TrFE) (*100*). Associated with other flexible components in FETs, the polymeric gate dielectric enabled a prototype of flexible memory, that is, an FeFET was demonstrated (Fig. 4A). The ON/OFF ratios between data "1" and "0" remain stable even for devices undergoing deformation of <1 mm bending radius or 1000 sharp folding cycles (*98*).

The successful demonstrations of mechanical flexibility in polymer FeFET inspired the development of smart wearable electronics. A rational approach to device integration of PVDF FeFETs into wearable electronics is to fabricate the devices on fibers, followed by the manufacturing of sewn fabrics (Fig. 4B). Stretching the fabric by 100% or crumpling the fabrics will not influence the memory functions (*103*).

In addition to the flexibility of PVDF-based memory, two critical performance parameters await improvement: a high data storage density and a low operation voltage. PVDF-based polymers are thermoplastic materials that are easily shaped at high temperatures. Nanoimprint technologies have been used to fabricate P(VDF-TrFE) nanoarrays with a single storage unit size of <100 nm, exhibiting the possibility of the data storage density reaching several gigabytes per square inch (*104*, *105*). A more practicable design for high-density data storage is vertical 3D stacking of FeFETs (Fig. 4C), in which every single unit can work independently (*106*, *107*).

The development of multilevel data storage in which a single memory unit can store more than two types ("0" and "1") of data was demonstrated in a PVDF-based FeFET, which also exhibits potential for enhancing the storage density (*106*, *108*). By modulating the gate voltage to yield various remnant polarizations of the ferroelectric layers, the semiconductor channels can reflect distinctive resistivities, resulting in multilevel storage features. Moreover, the high transparency of PVDF-based polymer thin films for visible light presents the possibility of combining optical switching of the electrical properties of the semiconductor layer and electrical switching of the ferroelectric polarization of the dielectric layer to realize multifunctional memories with stable multilevel storage features (*109*).

One major disadvantage of using PVDF-based polymers in memory is their high coercive field, ~50 MV/m. Therefore, a nanoscale thin film is required to make the operation voltage match the common operation voltage in integrated circuits. Many studies have reported the existence of ferroelectricity in PVDF-based materials at the nanoscale (*110*), confirming the capability of PVDF-based materials to work as memory in ultrathin films with a low switching voltage. Because of the ultrathin film of the ferroelectric layer, PVDF-based FTJs exhibit very low operation voltages; for example, Au/P(VDF-TrFE)/ITO FTJ memristor devices can execute "write" and "erase" operations at 0.5 and −0.4 V, respectively, along with "read" at 0.01 V (*111*).

Energy storage film capacitors

The high polarization together with high breakdown strength of PVDF-based polymers broaden their applications to film capacitors, given that their energy density can be easily higher than 10 J/cm^3, compared to the value of 5 J/cm^3 for the state-of-art film capacitor with biaxially oriented polypropylene (BOPP) (*112*, *113*). This is ideal for capacitor applications that require reduced device size, for example, implantable cardioverter defibrillators (ICDs) (*114*). Several major medical device companies are actively exploring such an option to replace the current tantalum electrolytic capacitors in ICDs. One critical step for dielectric and energy storage applications is to eliminate the polarization hysteresis in PVDF-based ferroelectric polymers and achieve a large and hysteresis-free polarization change.

Cost is another critical consideration for film capacitors. Hence, most studies have focused on PVDF-based copolymers with various defects, such as CTFE and hexafluoropropylene (HFP), which have yielded promising results. Moreover, Ren *et al.* reported that, even in neat PVDF, an ultrahigh discharged energy density of 50.2 J/cm^3 at 1 GV/m can be achieved in continuously folded films (Fig. 4D) (*115*). The major challenge of developing practical PVDF-based capacitors is to achieve a delicate balance between a high energy density and a low dielectric loss or low charge carrier conduction. One viable approach is to use ML films of PVDF with a high breakdown strength and a low-loss linear dielectric polymer, such as polycarbonate (PC) and poly(ethylene terephthalate) (PET). By optimizing the composition, layer thicknesses, and interfacial layers, ML film capacitors (Fig. 4E) with high energy density (~16 J/cm^3), low loss (~0.006), and improved thermal stability (up to 140°C) have been demonstrated (*116*, *117*). Another approach for capacitive energy storage relies on polymer and inorganic nanofiller composites, which use high-dielectric constant ceramic fillers to enhance the apparent permittivity. Per the mixing rules, this method usually requires a high filler loading (>10 vol %) to obtain a sizable increase in the permittivity. However, the major shortcoming of using nanocomposites is the inevitable aggregation of nanofillers, particularly at high filler loadings, leading to ready dielectric breakdown and mechanical brittleness (*118*, *119*). Thakur *et al.* and Zhang *et al.* reported the discovery of dilute nanocomposites in which nanofillers at ultralow loadings (< 0.3 vol %) generate large enhancements in both the dielectric constant and electric breakdown strength while maintaining a low dielectric loss in several high-temperature polymers (*120*, *121*). An

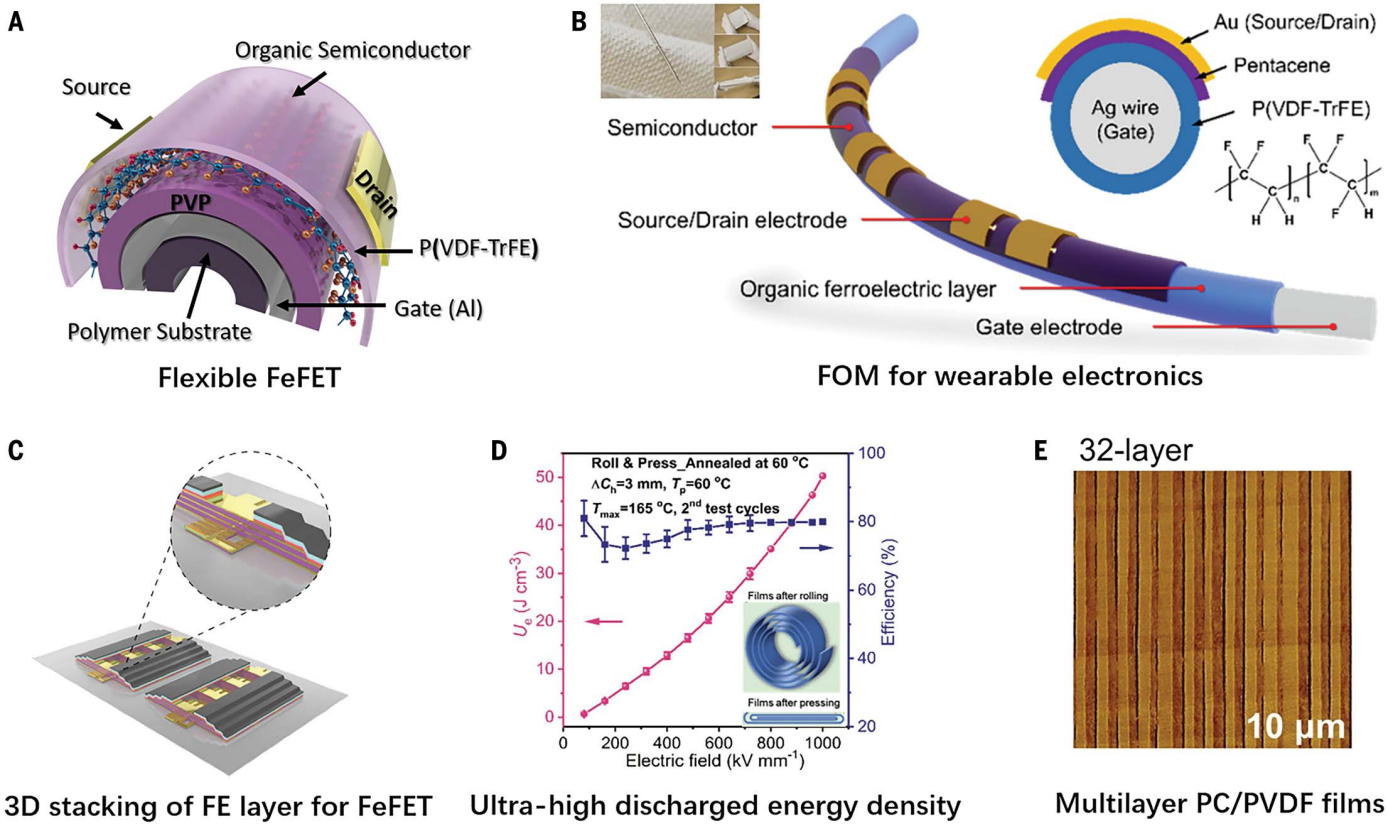


Fig. 4. Advanced dielectric applications for PVDF-based polymers. (**A**) Schematic of highly bent FeFET memory with P(VDF-TrFE) as the gate dielectric (*98*). (**B**) Schematic of fiber organic memory (FOM) based on P(VDF-TrFE). The inset presents a fabric with the FOM integrated (*103*). (**C**) Schematic of 3D stacking of a ferroelectric layer, P(VDF-TrFE), in a FeFET device (*106*). (**D**) Discharged energy density for PVDF films after continuous folding. The inset shows that the film was prepared by rolling and pressing (*115*). (**E**) Atomic force microscopy image of PC/PVDF ML films, 32 layers with a thickness of 400 nm for each layer (*116*). [(A), (B), (C), (D), and (E) adapted with permission from (*98*), (*103*) copyright © 2019 American Chemical Society, (*106*) copyright © 2019 Elsevier, and (*115*) and (*116*) copyright © 2017 Elsevier, respectively.]

analogous phenomenon is also observed in P(VDF-HFP) copolymers, in which only 0.8 vol % surface-hydroxylated $Cd_{1-x}Zn_xSe_{1-y}S_y$ nanodots generate a >100% enhancement in the dielectric constant together with a slight enhancement in the breakdown strength (*122*).

Conclusion and outlook

The advances in PVDF-based polymer ferroelectrics from ferroelectric P(VDF-TrFE) in the 1980s to P(VDF-TrFE-CFE-FA) tetrapolymers in the early 2020s demonstrate the promise and impact of molecular and nano- and meso-structure engineering in facile tailoring of the polarization processes to obtain the desired cross-coupling effect and generate the effect(s) at low electric fields. The tetrapolymer results for electrothermal and EM couplings and associated polarization responses reveal an unexplored horizon of polymer ferroelectric research (*22*, *27*). The polarization in polymer ferroelectrics originates from different sources and processes over different length scales from molecules to nano- and meso-structure morphologies and micro- and macroscale polarization domains, endowing the materials with multifunctional properties.

For ferroelectric polymers, although dipoles in polymer chains are the key to polarization, dilute nanocomposite phenomena reveal that for the same polymer dipoles, a minute amount of nanofillers (<0.2 vol %) can induce local and nanoscale polymer chain morphology changes that lead to marked enhancements in the dielectric constant (and, consequently, polarization) and electric breakdown strength (*120*, *121*). In functional materials, employing minute amounts of dopants and defects to tailor, establish, and enhance the desired functional performance is well known and widely used. The advances in PVDF-based polymer ferroelectrics and dilute nanocomposites also demonstrated the effectiveness of minor defects in selectively tailoring and influencing these polarization processes, leading to high energy conversion efficiency in desired cross-coupling effects, such as EM coupling, the ECE, and dielectric properties.

Even after many decades of research and development of polymer ferroelectrics, there are still many unknowns regarding the polarization processes in polymer dielectrics and polymer ferroelectrics at the molecular scale, nanoscale, and mesoscale, which present great opportunities for future research and development considering the rich polymer chemistry in constructing polymer molecules and nano- and mesoscale structures and the easy fabrication. Closely coupled experiments, theories, and characterization studies will lead to important insights into and uncover multiscale polarization processes and their cross-couplings to mechanical, thermal, magnetic, and optical effects. The hope is that these discoveries lead to enhanced performance, such as ultrahigh coupling and piezoelectric effect with no DC bias (Table 1), a giant ECE at much lower electric fields than that in Table 2, and tuning of the ferroelectric switching fields to any desired level. Consequently, advances in the material performance will translate into the better performance for the various highlighted applications, for example, wearable devices operated below the 50 V safe operation voltage defined by the Occupational Safety and

Health Administration. Moreover, the potential exists for generating emerging applications with high performance, such as flexible and wearable magnetoelectric sensors for biomedical applications (*123*) or large-area flexible and robust multifunctional sensor arrays for hospital and nursing home beds.

In contrast to the widely available inorganic ferroelectrics that cover a broad range of applications, after more than five decades of research, PVDF-based ferroelectrics are still the lone polymer ferroelectric family that has been successfully used in many applications. Although ferroelectricity has been observed in odd-numbered nylons (in the 1980s) and many other polymers (*8*, *9*), the low EM coupling and especially electroactive properties sensitive to environmental moisture prevent their practical use. Active polymer research exploiting the rich polarization processes and their underlying molecular structures, nanostructures and mesostructures for PVDF-based polymers has the potential to discover and develop next-generation classes of polymer ferroelectrics that are user-friendly, can be operated at much higher temperatures, and generate high electroactive couplings that rival those of PVDF-based polymers and even inorganic ferroelectrics.

REFERENCES AND NOTES

1. K. Uchino, *Ferroelectric Devices* (CRC Press, ed. 2, 2018).
2. M. E. Lines, A. M. Glass, *Principles and Applications of Ferroelectrics and Related Materials* (Oxford Univ. Press, 2001).
3. H. Kawai, The piezoelectricity of poly (vinylidene fluoride). *Jpn. J. Appl. Phys.* **8**, 975 (1969). doi: 10.1143/JJAP.8.975
4. G. T. Davis, J. E. McKinney, M. G. Broadhurst, S. C. Roth, Electric-field-induced phase changes in poly(vinylidene fluoride). *J. Appl. Phys.* **49**, 4998–5002 (1978). doi: 10.1063/1.324446
5. R. G. Kepler, R. A. Anderson, Ferroelectricity in polyvinylidene fluoride. *J. Appl. Phys.* **49**, 1232–1235 (1978). doi: 10.1063/1.325011
6. T. Furukawa, M. Date, E. Fukada, Y. Tajitsu, A. Chiba, Ferroelectric behavior in the copolymer of vinylidenefluoride and trifluoroethylene. *Jpn. J. Appl. Phys.* **19**, L109–L112 (1980). doi: 10.1143/JJAP.19.L109
7. T. Yagi, M. Tatemoto, J.-i. Sako, Transition behavior and dielectric properties in trifluoroethylene and vinylidene fluoride copolymers. *Polym. J.* **12**, 209–223 (1980). doi: 10.1295/polymj.12.209
8. A. J. Lovinger, Ferroelectric polymers. *Science* **220**, 1115–1121 (1983). doi: 10.1126/science.220.4602.1115; pmid: 17818472
9. H. S. Nalwa, Ed., *Ferroelectric Polymers: Chemistry, Physics, and Applications* (CRC Press, 1995).
10. L. E. Cross, Relaxor ferroelectrics. *Ferroelectrics* **76**, 241–267 (1987). doi: 10.1080/00150198708016945
11. Q. M. Zhang, V. Bharti, X. Zhao, Giant electrostriction and relaxor ferroelectric behavior in electron-irradiated poly (vinylidene fluoride-trifluoroethylene) copolymer. *Science* **280**, 2101–2104 (1998). doi: 10.1126/science.280.5372.2101; pmid: 9641912
12. F. Xia *et al.*, High electromechanical responses in a poly (vinylidene fluoride–trifluoroethylene–chlorofluoroethylene) terpolymer. *Adv. Mater.* **14**, 1574–1577 (2002). doi: 10.1002/1521-4095(20021104)14:21<1574::AID-ADMA1574>3.0.CO;2-#
13. R. Newnham, V. Sundar, R. Yimnirun, J. Su, Q. Zhang, Electrostriction: Nonlinear electromechanical coupling in solid dielectrics. *J. Phys. Chem. B* **101**, 10141–10150 (1997). doi: 10.1021/jp971522c
14. B. Neese *et al.*, Large electrocaloric effect in ferroelectric polymers near room temperature. *Science* **321**, 821–823 (2008). doi: 10.1126/science.1159655; pmid: 18687960
15. X. Li *et al.*, Pyroelectric and electrocaloric materials. *J. Mater. Chem. C* **1**, 23–37 (2013). doi: 10.1039/C2TC00283C
16. F. S. Foster, K. A. Harasiewicz, M. D. Sherar, A history of medical and biological imaging with polyvinylidene fluoride (PVDF) transducers. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **47**, 1363–1371 (2000). doi: 10.1109/58.883525; pmid: 18238682
17. Y. Wu *et al.*, Insect-scale fast moving and ultrarobust soft robot. *Sci. Robot.* **4**, eaax1594 (2019). doi: 10.1126/scirobotics.aax1594; pmid: 33137774
18. W. Yang *et al.*, All-fiber tribo-ferroelectric synergistic electronics with high thermal-moisture stability and comfortability. *Nat. Commun.* **10**, 5541 (2019). doi: 10.1038/s41467-019-13569-5; pmid: 31804506
19. Y. Chen *et al.*, An all-scale hierarchical architecture induces colossal room-temperature electrocaloric effect at ultralow electric field in polymer nanocomposites. *Adv. Mater.* **32**, e1907927 (2020). doi: 10.1002/adma.201907927; pmid: 32567148
20. W. Yan *et al.*, Single fibre enables acoustic fabrics via nanometre-scale vibrations. *Nature* **603**, 616–623 (2022). doi: 10.1038/s41586-022-04476-9; pmid: 35296860
21. PolyK Technologies, LLC, Regenerative face mask with long shelf life and long service time for effective coronavirus filtration, Dept. of Health and Human Services, Centers for Disease Control and Prevention, SBIR Phase I, R43OH012416 (2022); https://www.sbir.gov/node/2344785.
22. X. Qian *et al.*, High-entropy polymer produces a giant electrocaloric effect at low fields. *Nature* **600**, 664–669 (2021). doi: 10.1038/s41586-021-04189-5; pmid: 34937898
23. R. Ma *et al.*, Highly efficient electrocaloric cooling with electrostatic actuation. *Science* **357**, 1130–1134 (2017). doi: 10.1126/science.aan5980; pmid: 28912240
24. Y. Meng *et al.*, A cascade electrocaloric cooling device for large temperature lift. *Nat. Energy* **5**, 996–1002 (2020). doi: 10.1038/s41560-020-00715-3
25. Q. V. Duong, V. P. Nguyen, F. Domingues Dos Santos, S. T. Choi, Localized fretting-vibrotactile sensations for large-area displays. *ACS Appl. Mater. Interfaces* **11**, 33292–33301 (2019). doi: 10.1021/acsami.9b09691; pmid: 31411459
26. T. T. Le *et al.*, Piezoelectric nanofiber membrane for reusable, stable, and highly functional face mask filter with long-term biodegradability. *Adv. Funct. Mater.* **32**, 2113040 (2022). doi: 10.1002/adfm.202113040
27. X. Chen *et al.*, Relaxor ferroelectric polymer exhibits ultrahigh electromechanical coupling at low electric field. *Science* **375**, 1418–1422 (2022). doi: 10.1126/science.abn0936; pmid: 35324308
28. "IEEE standard on piezoelectricity: ANSI/IEEE Std 176-1987" in *Piezoelectricity*, C. Zwick Rosen, B. V. Hiremath, R. Newham, Eds. (American Institute of Physics, 1992), pp. 227–228.
29. Z. Y. Cheng, Q. M. Zhang, Field-activated electroactive polymers. *MRS Bull.* **33**, 183–187 (2008). doi: 10.1557/mrs2008.43
30. G. J. Velders, D. W. Fahey, J. S. Daniel, M. McFarland, S. O. Andersen, The large contribution of projected HFC emissions to future climate forcing. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 10949–10954 (2009). doi: 10.1073/pnas.0902817106; pmid: 19549868
31. P. Hawken, Ed., *Drawdown: The Most Comprehensive Plan Ever Proposed to Reverse Global Warming* (Penguin, 2017).
32. J. Shi *et al.*, Electrocaloric cooling materials and devices for zero-global-warming-potential, high-efficiency refrigeration. *Joule* **3**, 1200–1225 (2019). doi: 10.1016/j.joule.2019.03.021
33. P. Kobeco, I. Kurtchatov, Dielectric properties of Rochelle salt crystal. *Z. Phys.* **66**, 192–205 (1930).
34. G. G. Wiseman, J. K. Kuebler, Electrocaloric effect in ferroelectric Rochelle salt. *Phys. Rev.* **131**, 2023–2027 (1963). doi: 10.1103/PhysRev.131.2023
35. Y. V. Sinyavsky, N. Pashkov, Y. Gorovoy, G. Lugansky, L. Shebanov, The optical ferroelectric ceramic as working body for electrocaloric refrigeration. *Ferroelectrics* **90**, 213–217 (1989). doi: 10.1080/00150198908211296
36. Y. V. Sinyavsky, V. M. Brodyansky, Experimental testing of electrocaloric cooling with transparent ferroelectric ceramic as a working body. *Ferroelectrics* **131**, 321–325 (1992). doi: 10.1080/00150199208223433
37. X. Moya, S. Kar-Narayan, N. D. Mathur, Caloric materials near ferroic phase transitions. *Nat. Mater.* **13**, 439–450 (2014). doi: 10.1038/nmat3951; pmid: 24751772
38. S.-G. Lu *et al.*, Organic and inorganic relaxor ferroelectrics with giant electrocaloric effect. *Appl. Phys. Lett.* **97**, 162904 (2010). doi: 10.1063/1.3501975
39. X. Li *et al.*, Tunable temperature dependence of electrocaloric effect in ferroelectric relaxor poly (vinylidene fluoride-trifluoroethylene-chlorofluoroethylene terpolymer. *Appl. Phys. Lett.* **99**, 052907 (2011). doi: 10.1063/1.3624533
40. F. Xu *et al.*, Domain wall motion and its contribution to the dielectric and piezoelectric properties of lead zirconate titanate films. *J. Appl. Phys.* **89**, 1336–1348 (2001). doi: 10.1063/1.1325005
41. X. Chen *et al.*, Giant electrostriction enabled by defect-induced critical phenomena in relaxor ferroelectric polymers. *Macromolecules* **56**, 690–696 (2023). doi: 10.1021/acs.macromol.2c01952
42. S. Anwar *et al.*, Solution-processed transparent ferroelectric nylon thin films. *Sci. Adv.* **5**, eaav3489 (2019). doi: 10.1126/sciadv.aav3489; pmid: 31453321
43. Q. Li, Q. Wang, Ferroelectric polymers and their energy-related applications. *Macromol. Chem. Phys.* **217**, 1228–1244 (2016). doi: 10.1002/macp.201500503
44. A. J. Lovinger, D. D. Davis, R. E. Cais, J. M. Kometani, The role of molecular defects on the structure and phase transitions of poly (vinylidene fluoride). *Polymer* **28**, 617–626 (1987). doi: 10.1016/0032-3861(87)90478-2
45. G. Rui *et al.*, Giant spontaneous polarization for enhanced ferroelectric properties of biaxially oriented poly (vinylidene fluoride) by mobile oriented amorphous fractions. *J. Mater. Chem. C* **9**, 894–907 (2021). doi: 10.1039/D0TC04632A
46. A. V. Bune *et al.*, Two-dimensional ferroelectric films. *Nature* **391**, 874–877 (1998). doi: 10.1038/36069
47. M. Guo *et al.*, Toroidal polar topology in strained ferroelectric polymer. *Science* **371**, 1050–1056 (2021). doi: 10.1126/science.abc4727; pmid: 33674493
48. Y. Liu *et al.*, Ferroelectric polymers exhibiting behaviour reminiscent of a morphotropic phase boundary. *Nature* **562**, 96–100 (2018). doi: 10.1038/s41586-018-0550-z; pmid: 30283102
49. V. S. Bystrov *et al.*, Molecular modeling of the piezoelectric effect in the ferroelectric polymer poly(vinylidene fluoride) (PVDF). *J. Mol. Model.* **19**, 3591–3602 (2013). doi: 10.1007/s00894-013-1891-z; pmid: 23729009
50. S. G. Lu, Q. M. Zhang, Electrocaloric materials for solid-state refrigeration. *Adv. Mater.* **21**, 1983–1987 (2009). doi: 10.1002/adma.200802902
51. X. Li *et al.*, Giant electrocaloric effect in ferroelectric poly (vinylidenefluoride-trifluoroethylene) copolymers near a first-order ferroelectric transition. *Appl. Phys. Lett.* **101**, 132903 (2012). doi: 10.1063/1.4756697
52. T. C. Chung, A. Petchsuk, Synthesis and properties of ferroelectric fluoroterpolymers with Curie transition at ambient temperature. *Macromolecules* **35**, 7678–7684 (2002). doi: 10.1021/ma020504c
53. H. Xu *et al.*, Ferroelectric and electromechanical properties of poly (vinylidene-fluoride–trifluoroethylene–chlorotrifluoroethylene) terpolymer. *Appl. Phys. Lett.* **78**, 2360–2362 (2001). doi: 10.1063/1.1358847
54. J. Shi *et al.*, Numerical evaluation of a kilowatt-level rotary electrocaloric refrigeration system. *Int. J. Refrig.* **121**, 279–288 (2021). doi: 10.1016/j.ijrefrig.2020.09.011
55. X. Moya, N. D. Mathur, Caloric materials for cooling and heating. *Science* **370**, 797–803 (2020). doi: 10.1126/science.abb0973; pmid: 33184207
56. H. Gu *et al.*, A chip scale electrocaloric effect based cooling device. *Appl. Phys. Lett.* **102**, 122904 (2013). doi: 10.1063/1.4799283
57. J. Qian, J. Jiang, Y. Shen, Enhanced electrocaloric strength in P (VDF-TrFE-CFE) by decreasing the crystalline size. *J. Materiomics* **5**, 357–362 (2019). doi: 10.1016/j.jmat.2019.04.001
58. G. Zhang *et al.*, Ferroelectric polymer nanocomposites with complementary nanostructured fillers for electrocaloric cooling with high power density and great efficiency. *ACS Appl. Energy Mater.* **1**, 1344–1354 (2018). doi: 10.1021/acsaem.8b00052
59. J. Qian *et al.*, Interfacial coupling boosts giant electrocaloric effects in relaxor polymer nanocomposites: In situ characterization and phase-field simulation. *Adv. Mater.* **31**, e1801949 (2019). pmid: 30537017
60. Y. Liu *et al.*, Chirality-induced relaxor properties in ferroelectric polymers. *Nat. Mater.* **19**, 1169–1174 (2020). doi: 10.1038/s41563-020-0724-6; pmid: 32601482
61. S. Qian *et al.*, Not-in-kind cooling technologies: A quantitative comparison of refrigerants and system performance. *Int. J. Refrig.* **62**, 177–192 (2015). doi: 10.1016/j.ijrefrig.2015.10.019

62. A. Torelló *et al*., Giant temperature span in electrocaloric regenerator. *Science* **370**, 125–129 (2020). doi: 10.1126/science.abb8045; pmid: 33004522
63. B. Nair *et al*., Large electrocaloric effects in oxide multilayer capacitors over a wide temperature range. *Nature* **575**, 468–472 (2019). doi: 10.1038/s41586-019-1634-0; pmid: 31597164
64. D. H. Kim, H. S. Park, M. S. Kim, The effect of the refrigerant charge amount on single and cascade cycle heat pump systems. *Int. J. Refrig.* **40**, 254–268 (2014). doi: 10.1016/j.ijrefrig.2013.10.002
65. F. Le Goupil *et al*., Enhanced electrocaloric response of vinylidene fluoride–based polymers via one-step molecular engineering. *Adv. Funct. Mater.* **31**, 2007043 (2021). doi: 10.1002/adfm.202007043
66. S. T. Choi, J. O. Kwon, F. Bauer, Multilayered relaxor ferroelectric polymer actuators for low-voltage operation fabricated with an adhesion-mediated film transfer technique. *Sens. Actuators A Phys.* **203**, 282–290 (2013). doi: 10.1016/j.sna.2013.08.049
67. E. Häsler, L. Stein, G. Harbauer, Implantable physiological power supply with PVDF film. *Ferroelectrics* **60**, 277–282 (1984). doi: 10.1080/00150198408017528
68. J. Kymissis, C. Kendall, J. Paradiso, N. Gershenfeld, "Parasitic power harvesting in shoes," in *Digest of Papers. Second International Symposium on Wearable Computers (Cat. No. 98EX215)*, Pittsburgh, PA, 19 and 20 October 1998 (IEEE, 1998), pp. 132–139.
69. A. Ranjan, C. Peng, S. Wagle, F. Melandsø, A. Habib, High-frequency acoustic imaging using adhesive-free polymer transducer. *Polymers* **13**, 1462 (2021). doi: 10.3390/polym13091462; pmid: 33946539
70. C. Yang, S. Song, F. Chen, N. Chen, Fabrication of PVDF/$BaTiO_3$/CNT piezoelectric energy harvesters with bionic balsa wood structures through 3D printing and supercritical carbon dioxide foaming. *ACS Appl. Mater. Interfaces* **13**, 41723–41734 (2021). doi: 10.1021/acsami.1c11843; pmid: 34431292
71. S. Panda *et al*., Biocompatible $CaTiO_3$-PVDF composite-based piezoelectric nanogenerator for exercise evaluation and energy harvesting. *Nano Energy* **102**, 107682 (2022). doi: 10.1016/j.nanoen.2022.107682
72. Y. Wang *et al*., MXenes for energy harvesting. *Adv. Mater.* **34**, e2108560 (2022). doi: 10.1002/adma.202108560; pmid: 35048436
73. F. Jiang *et al*., Stretchable, breathable, and stable lead-free perovskite/polymer nanofiber composite for hybrid triboelectric and piezoelectric energy harvesting. *Adv. Mater.* **34**, e2200042 (2022). doi: 10.1002/adma.202200042; pmid: 35231951
74. H. Hatcher, Energy-harvesting clothes. *Nat. Rev. Mater.* **7**, 256 (2022). doi: 10.1038/s41578-022-00437-w
75. N. Sezer, M. Koç, A comprehensive review on the state-of-the-art of piezoelectric energy harvesting. *Nano Energy* **80**, 105567 (2021). doi: 10.1016/j.nanoen.2020.105567
76. J. Fu, Y. Hou, X. Gao, M. Zheng, M. Zhu, Highly durable piezoelectric energy harvester based on a PVDF flexible nanocomposite filled with oriented $BaTi_2O_5$ nanorods with high power density. *Nano Energy* **52**, 391–401 (2018). doi: 10.1016/j.nanoen.2018.08.006
77. Y. M. Yousry *et al*., Theoretical model and outstanding performance from constructive piezoelectric and triboelectric mechanism in electrospun PVDF fiber film. *Adv. Funct. Mater.* **30**, 1910592 (2020). doi: 10.1002/adfm.201910592
78. F. Mokhtari *et al*., Wearable electronic textiles from nanostructured piezoelectric fibers. *Adv. Mater. Technol.* **5**, 1900900 (2020). doi: 10.1002/admt.201900900
79. H. Gao *et al*., High-performance flexible yarn for wearable piezoelectric nanogenerators. *Smart Mater. Struct.* **27**, 095018 (2018). doi: 10.1088/1361-665X/aad718
80. K. Castkova *et al*., Structure–properties relationship of electrospun PVDF fibers. *Nanomaterials* **10**, 1221 (2020). doi: 10.3390/nano10061221; pmid: 32585824
81. L. Persano *et al*., High performance piezoelectric devices based on aligned arrays of nanofibers of poly (vinylidenefluoride-co-trifluoroethylene). *Nat. Commun.* **4**, 1633 (2013). doi: 10.1038/ncomms2639; pmid: 23535654
82. K. Shi, B. Sun, X. Huang, P. Jiang, Synergistic effect of graphene nanosheet and $BaTiO_3$ nanoparticles on performance enhancement of electrospun PVDF nanofiber mat for flexible piezoelectric nanogenerators. *Nano Energy* **52**, 153–162 (2018). doi: 10.1016/j.nanoen.2018.07.053
83. C. Peng, M. Chen, H. Wang, J. Shen, X. Jiang, P(VDF-TrFE) thin-film-based transducer for under-display ultrasonic fingerprint sensing applications. *IEEE Sens. J.* **20**, 11221–11228 (2020). doi: 10.1109/JSEN.2020.2997375
84. S. Biswas, Y. Visell, Haptic perception, mechanics, and material technologies for virtual reality. *Adv. Funct. Mater.* **31**, 2008186 (2021). doi: 10.1002/adfm.202008186
85. M. Innocenti, W. Bruno, S. Merighi, V. Parkula, F. Jeanneau, "The science of touch in electronics haptics, it used to be all about resonant frequency," Passive Components Networking Symposium, Milan, Italy, 7 to 10 September 2021.
86. H. Gu, X.-S. Qian, H.-J. Ye, Q. M. Zhang, An electrocaloric refrigerator without external regenerator. *Appl. Phys. Lett.* **105**, 162905 (2014). doi: 10.1063/1.4898812
87. H. Gu *et al*., Simulation of chip-size electrocaloric refrigerator with high cooling-power density. *Appl. Phys. Lett.* **102**, 112901 (2013). doi: 10.1063/1.4796184
88. Y. Wang *et al*., A high-performance solid-state electrocaloric cooling system. *Science* **370**, 129–133 (2020). doi: 10.1126/science.aba2648; pmid: 33004523
89. T. Zhang, X.-S. Qian, H. Gu, Y. Hou, Q. Zhang, An electrocaloric refrigerator with direct solid to solid regeneration. *Appl. Phys. Lett.* **110**, 243503 (2017). doi: 10.1063/1.4986508
90. H. Hou, S. Qian, I. Takeuchi, Materials, physics and systems for multicaloric cooling. *Nat. Rev. Mater.* **7**, 633–652 (2022). doi: 10.1038/s41578-022-00428-x
91. S. R. Annapragada, P. Verma, A. Sur, W. Xie, United Technologies Research Center, "High-Efficiency Solid-State Heat Pump Module: 2017 Building Technologies Office Peer Review" (US Department of Energy, 2017); https://www.energy.gov/sites/default/files/2017/04/f34/7_312111_Annapragada_031517-1430.pdf
92. D. Guo *et al*., Design and modeling of a fluid-based micro-scale electrocaloric refrigeration system. *Int. J. Heat Mass Transf.* **72**, 559–564 (2014). doi: 10.1016/j.ijheatmasstransfer.2014.01.043
93. U. Plaznik *et al*., Bulk relaxor ferroelectric ceramics as a working body for an electrocaloric cooling device. *Appl. Phys. Lett.* **106**, 043903 (2015). doi: 10.1063/1.4907258
94. H. Cui *et al*., Flexible microfluidic electrocaloric cooling capillary tube with giant specific device cooling power density. *Joule* **6**, 258–268 (2022). doi: 10.1016/j.joule.2021.12.010
95. Y. Bo *et al*., Electrostatic actuating double-unit electrocaloric cooling device with high efficiency. *Adv. Energy Mater.* **11**, 2003771 (2021). doi: 10.1002/aenm.202003771
96. W. Ji, B. Cao, Y. Geng, Y. Zhu, B. Lin, Study on human skin temperature and thermal evaluation in step change conditions: From non-neutrality to neutrality. *Energy Build.* **156**, 29–39 (2017). doi: 10.1016/j.enbuild.2017.09.037
97. K. L. Kim *et al*., Epitaxial growth of thin ferroelectric polymer films on graphene layer for fully transparent and flexible nonvolatile memory. *Nano Lett.* **16**, 334–340 (2016). doi: 10.1021/acs.nanolett.5b03882; pmid: 26618802
98. R. H. Kim *et al*., Non-volatile organic memory with sub-millimetre bending radius. *Nat. Commun.* **5**, 3583 (2014). doi: 10.1038/ncomms4583; pmid: 24709956
99. M. Xu *et al*., Gate-controlled multi-bit nonvolatile ferroelectric organic transistor memory on paper substrates. *J. Mater. Chem. C* **7**, 13477–13485 (2019). doi: 10.1039/C9TC04356J
100. R. C. G. Naber *et al*., High-performance solution-processed polymer ferroelectric field-effect transistors. *Nat. Mater.* **4**, 243–248 (2005). doi: 10.1038/nmat1329
101. B. B. Tian *et al*., Tunnel electroresistance through organic ferroelectrics. *Nat. Commun.* **7**, 11502 (2016). doi: 10.1038/ncomms11502; pmid: 27143121
102. H. Li, R. Wang, S.-T. Han, Y. Zhou, Ferroelectric polymers for non-volatile memory devices: A review. *Polym. Int.* **69**, 533–544 (2020). doi: 10.1002/pi.5980
103. M. Kang *et al*., Low-voltage organic transistor memory fiber with a nanograined organic ferroelectric film. *ACS Appl. Mater. Interfaces* **11**, 22575–22582 (2019). doi: 10.1021/acsami.9b03564; pmid: 31148447
104. Y. Chen *et al*., High-resolution structural mapping and single-domain switching kinetics in 2D-confined ferroelectric nanodots for low-power FeRAM. *Nanoscale* **12**, 11997–12006 (2020). doi: 10.1039/D0NR02210A; pmid: 32463061
105. Z. Hu, M. Tian, B. Nysten, A. M. Jonas, Regular arrays of highly ordered ferroelectric polymer nanostructures for non-volatile low-voltage memories. *Nat. Mater.* **8**, 62–67 (2009). doi: 10.1038/nmat2339; pmid: 19060889
106. S. Hwang *et al*., Multi-level operation of three-dimensionally stacked non-volatile ferroelectric polymer memory with high-performance hole-injection layer. *Org. Electron.* **75**, 105394 (2019). doi: 10.1016/j.orgel.2019.105394
107. S. K. Hwang, S. M. Cho, K. L. Kim, C. Park, 3D-stacked vertical channel nonvolatile polymer memory. *Adv. Electron. Mater.* **1**, 1400042 (2015). doi: 10.1002/aelm.201400042
108. M. Carroli *et al*., Multiresponsive non-volatile memories based on optically switchable ferroelectric organic field-effect transistors. *Adv. Mater.* **33**, e2007965 (2021). doi: 10.1002/adma.202007965; pmid: 33656201
109. S. He, M. Guo, Y. Wang, Y. Liang, Y. Shen, An optical/ferroelectric multiplexing multidimensional nonvolatile memory from ferroelectric polymer. *Adv. Mater.* **34**, 2202181 (2022). doi: 10.1002/adma.202202181
110. Z. Yin, B. Tian, Q. Zhu, C. Duan, Characterization and application of PVDF and its copolymer Films prepared by spin-coating and Langmuir–Blodgett method. *Polymers* **11**, 2033 (2019). doi: 10.3390/polym11122033; pmid: 31817985
111. S. Majumdar, B. Chen, Q. H. Qin, H. S. Majumdar, S. van Dijken, Electrode dependence of tunneling electroresistance and switching stability in organic ferroelectric P(VDF-TrFE)-based tunnel junctions. *Adv. Funct. Mater.* **28**, 1703273 (2018). doi: 10.1002/adfm.201703273
112. B. Chu *et al*., A dielectric polymer with high electric energy density and fast discharge speed. *Science* **313**, 334–336 (2006). doi: 10.1126/science.1127798; pmid: 16857935
113. N. Meng *et al*., Ultrahigh β-phase content poly(vinylidene fluoride) with relaxor-like ferroelectricity for high energy density capacitors. *Nat. Commun.* **10**, 4535 (2019). doi: 10.1038/s41467-019-12391-3; pmid: 31628311
114. J. P. DiMarco, Implantable cardioverter-defibrillators. *N. Engl. J. Med.* **349**, 1836–1847 (2003). doi: 10.1056/NEJMra035432; pmid: 14602883
115. X. Ren *et al*., Ultra-high energy density integrated polymer dielectric capacitors. *J. Mater. Chem. A* **10**, 10171–10180 (2022). doi: 10.1039/D1TA09045C
116. E. Baer, L. Zhu, 50th anniversary perspective: Dielectric phenomena in polymers and multilayered dielectric films. *Macromolecules* **50**, 2239–2256 (2017). doi: 10.1021/acs.macromol.6b02669
117. T. Ju *et al*., Enhancing breakdown strength and lifetime of multilayer dielectric films by using high temperature polycarbonate skin layers. *Energy Storage Mater.* **45**, 494–503 (2022). doi: 10.1016/j.ensm.2021.12.009
118. H. Li *et al*., Enabling high-energy-density high-efficiency ferroelectric polymer nanocomposites with rationally designed nanofillers. *Adv. Funct. Mater.* **31**, 2006739 (2021). doi: 10.1002/adfm.202006739
119. G. Zhang, Q. Li, E. Allahyarov, Y. Li, L. Zhu, Challenges and opportunities of polymer nanodielectrics for capacitive energy storage. *ACS Appl. Mater. Interfaces* **13**, 37939–37960 (2021). doi: 10.1021/acsami.1c04991; pmid: 34370438
120. Y. Thakur *et al*., Enhancement of the dielectric response in polymer nanocomposites with low dielectric constant fillers. *Nanoscale* **9**, 10992–10997 (2017). doi: 10.1039/C7NR01932G; pmid: 28752176
121. T. Zhang *et al*., A highly scalable dielectric metamaterial with superior capacitor performance over a broad temperature. *Sci. Adv.* **6**, eaax6622 (2020). doi: 10.1126/sciadv.aax6622; pmid: 32042896
122. L. Li *et al*., Significant improvements in dielectric constant and energy density of ferroelectric polymer nanocomposites enabled by ultralow contents of nanofillers. *Adv. Mater.* **33**, e2102392 (2021). doi: 10.1002/adma.202102392; pmid: 34302399
123. C.-W. Nan, M. Bichurin, S. Dong, D. Viehland, G. Srinivasan, Multiferroic magnetoelectric composites: Historical perspective, status, and future directions. *J. Appl. Phys.* **103**, 031101 (2008). doi: 10.1063/1.2836410
124. K. Zou *et al*., Giant room-temperature electrocaloric effect of polymer-ceramic composites with orientated $BaSrTiO_3$ nanofibers. *Nano Lett.* **22**, 6560–6566 (2022). doi: 10.1021/acs.nanolett.2c01776; pmid: 35947031
125. M.-D. Li *et al*., Thermal management of chips by a device prototype using synergistic effects of 3-D heat-conductive network and electrocaloric refrigeration. *Nat. Commun.* **13**, 5849 (2022). doi: 10.1038/s41467-022-33596-z; pmid: 36195612
126. G. Qian *et al*., The electrocaloric effect of PBZ/PVDF flexible composite film near room temperature. *J. Mater. Sci.*

Mater. Electron. **32**, 12001–12016 (2021). doi: 10.1007/s10854-021-05831-8

127. Z. Wang *et al*., Enhanced electrocaloric effect within a broad temperature range in lead-free polymer composite films by blending the rare-earth doped $BaTiO_3$ nanopowders. *Adv. Compos. Hybrid Mater.* **4**, 469–477 (2021). doi: 10.1007/s42114-021-00252-x

128. P. Bai *et al*., An active pixel-matrix electrocaloric device for targeted and differential thermal management. *Adv. Mater.* **35**, e2209181 (2023). doi: 10.1002/adma.202209181; pmid: 36690602

ACKNOWLEDGMENTS

Funding: Q.M.Z. and X.C. acknowledge support from the US Office of Naval Research under award N00014-19-1-2028. X.Q. thanks the National Natural Science Foundation of China (grant 52076127), the Natural Science Foundation of Shanghai (grants 20ZR1471700 and 22JC1401800), and the State Key Laboratory of Mechanical System and Vibration (grant MSVZD202211). L.Z. acknowledges financial support from the US National Science Foundation (grant DMR-2103196). **Competing interests**: Q.M.Z. and X.C. have filed a PCT patent application (PCT/US2022/032214) at Penn State on EM tetrapolymers and related EM devices. X.Q. has filed a patent application (WO/2022/257747) on tetrapolymers and ECE. L.Z. has no competing interests.
License information: Copyright © 2023 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

Submitted 1 December 2022; accepted 5 April 2023
10.1126/science.adg0902

RESEARCH ARTICLE SUMMARY

CANCER

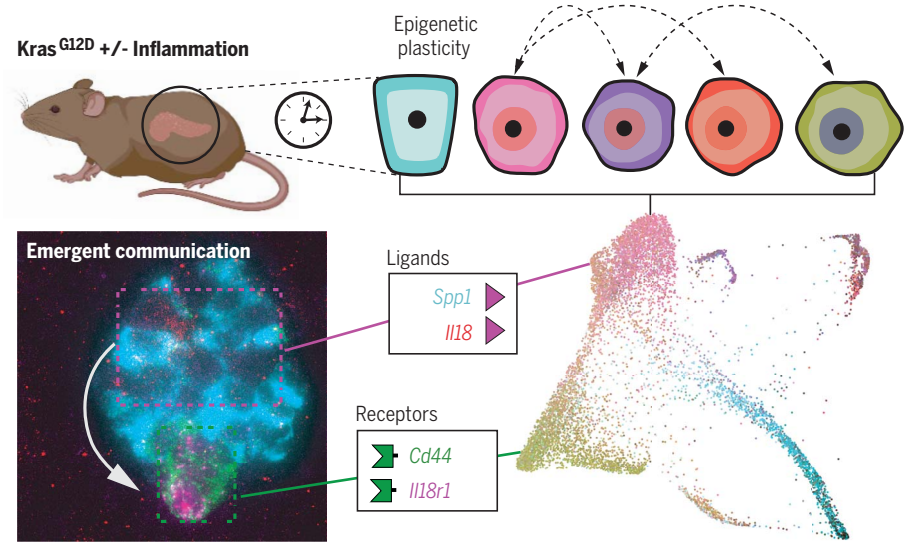
Epigenetic plasticity cooperates with cell-cell interactions to direct pancreatic tumorigenesis

Cassandra Burdziak†, Direna Alonso-Curbelo†, Thomas Walle, José Reyes, Francisco M. Barriga, Doron Haviv, Yubin Xie, Zhen Zhao, Chujun Julia Zhao, Hsuan-An Chen, Ojasvi Chaudhary, Ignas Masilionis, Zi-Ning Choo, Vianne Gao, Wei Luan, Alexandra Wuest, Yu-Jui Ho, Yuhong Wei, Daniela F. Quail, Richard Koche, Linas Mazutis, Ronan Chaligné, Tal Nawy, Scott W. Lowe*, Dana Pe'er*

INTRODUCTION: Virtually all cancers begin with genetic alterations in healthy cells, but mounting evidence suggests that nongenetic events such as environmental signaling play a crucial role in unleashing tumorigenesis. In the pancreas, epithelial cells harboring an activating mutation in the *Kras* proto-oncogene can remain phenotypically normal until an inflammatory event that drives cellular plasticity and tissue remodeling. The inflammation-driven molecular, cellular, and tissue changes that precede and direct tumor formation remain poorly understood.

RATIONALE: Understanding tumorigenesis requires a high-resolution view of events spanning cancer progression. We leveraged genetically engineered mouse models (GEMMs), single-cell genomics (RNA sequencing and assay for transposase-accessible chromatin sequencing), and imaging technologies to measure pancreatic epithelial cell states across physiological, premalignant, and malignant stages. To analyze this rich and complex dataset, we developed computational and functional approaches to characterize epigenetic plasticity and to infer cell-cell communication impacts on tissue remodeling.

RESULTS: Our data revealed that early in tumorigenesis, *Kras*-mutant cells are capable of acquiring multiple highly reproducible cell states that are undetectable in normal or regenerating pancreata. Several such states align with experimentally validated cells of origin of neoplastic lesions, some of which display a high degree of plasticity upon inflammatory insult. These diverse *Kras*-mutant cell populations are defined by distinct chromatin accessibility patterns and undergo inflammation-driven cell fate transitions that precede preneoplastic and premalignant lesion formation. Furthermore, a subset of early *Kras*-mutant cell states exhibit marked similarity to either benign or malignant fates that emerge weeks to months later; for instance, *Kras*-mutant *Nestin*-positive progenitor-like cells display accessible chromatin near genes active in malignant tumors.



A lens into epigenetic plasticity and its consequences on tumorigenesis. A genetically engineered mouse model enabled a longitudinal view of cell states in early pancreatic cancer progression, revealing plasticity endowed through epigenetic priming. Plastic cell states express distinct sets of communication genes involved in tumorigenesis-specific communication, as visualized by highly multiplexed imaging. Mouse illustration was created with BioRender (https://biorender.com/).

We defined and quantified epigenetic plasticity as the diversity in transcriptional phenotypes that is enabled or restricted by a given epigenetic accessibility landscape. These plastic cell states are enriched for open chromatin near cell-cell communication genes encoding ligands and cell-surface receptors, suggesting an increased propensity to communicate with the microenvironment. Given the rapid remodeling of both the epithelial and immune compartments during inflammation, we hypothesize that this epigenetically enabled communication is a major driver of tumorigenesis. We found that the premalignant epithelium displays extraordinary modularity with respect to communication gene coexpression patterns, with distinct cell subpopulations each expressing a unique set of receptors and ligands that define the nature of incoming and outgoing signals that they can receive and send.

Through the development of Calligraphy, an algorithm that utilizes this receptor-ligand modularity to robustly infer the cell-cell communication underlying tissue remodeling, we showed that the enhanced signaling repertoire of early neoplastic tissue endows specific plastic epithelial populations with greater capability for cross-talk, including numerous communication routes with immune populations. As one example, we identified a feedback loop between inflammation-driven *Kras*-mutant epithelial and immune cell states involving interleukin-33 (IL-33), which was previously implicated in pancreatic tumorigenesis. Using a new GEMM that enables spatiotemporally controlled suppression of epithelial *Il33* expression during mutant *Kras*-initiated neoplasia, we functionally demonstrated that the loop initiated by epithelial IL-33 directs exit from a highly plastic, inflammation-induced epithelial state, enabling progression toward typical neoplasia.

CONCLUSION: Multimodal single-cell profiling of tumorigenesis in mouse models identified the cellular and tissue determinants of pancreatic cancer initiation, and a rigorous quantification of plasticity enabled the discovery of plasticity-associated gene programs. We found that *Kras*-mutant subpopulations markedly increase epigenetic plasticity upon inflammation, reshaping their communication potential with immune cells and establishing aberrant cell-cell communication loops that drive their progression toward neoplastic lesions. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: peerd@mskcc.org (D.P.); lowes@mskcc.org (S.W.L.)
†These authors contributed equally to this work.
Cite this article as C. Burdziak *et al.*, *Science* **380**, eadd5327 (2023). DOI: 10.1126/science.add5327

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.add5327

RESEARCH ARTICLE

CANCER

Epigenetic plasticity cooperates with cell-cell interactions to direct pancreatic tumorigenesis

Cassandra Burdziak[1,2]†, Direna Alonso-Curbelo[3,4]†, Thomas Walle[1,5,6,7], José Reyes[1,3], Francisco M. Barriga[3], Doron Haviv[1,2], Yubin Xie[1,2], Zhen Zhao[3,8], Chujun Julia Zhao[1,9], Hsuan-An Chen[3], Ojasvi Chaudhary[1,10], Ignas Masilionis[1,10], Zi-Ning Choo[1], Vianne Gao[1,2], Wei Luan[3], Alexandra Wuest[3], Yu-Jui Ho[3], Yuhong Wei[11], Daniela F. Quail[11], Richard Koche[12], Linas Mazutis[1,9,13], Ronan Chaligné[1,10], Tal Nawy[1], Scott W. Lowe[3,14]*, Dana Pe'er[1,14]*

The response to tumor-initiating inflammatory and genetic insults can vary among morphologically indistinguishable cells, suggesting as yet uncharacterized roles for epigenetic plasticity during early neoplasia. To investigate the origins and impact of such plasticity, we performed single-cell analyses on normal, inflamed, premalignant, and malignant tissues in autochthonous models of pancreatic cancer. We reproducibly identified heterogeneous cell states that are primed for diverse, late-emerging neoplastic fates and linked these to chromatin remodeling at cell-cell communication loci. Using an inference approach, we revealed signaling gene modules and tissue-level cross-talk, including a neoplasia-driving feedback loop between discrete epithelial and immune cell populations that was functionally validated in mice. Our results uncover a neoplasia-specific tissue-remodeling program that may be exploited for pancreatic cancer interception.

The initial events by which tissues diverge from normalcy to form benign neoplasms and malignant tumors remain poorly understood. It is well established that this process is driven by genetic mutations (*1*); however, the discovery of prevalent cancer driver mutations in phenotypically normal epithelia (*2*) challenges the classic notion of cancer pathogenesis and underscores the essential role of cellular and environmental context (*3*–*5*). Indeed, nonmutagenic environmental insults promote tumor initiation in mice (*6*, *7*), and chronic inflammatory conditions substantially increase cancer risk in humans (*8*, *9*). These events can have heterogeneous effects even among morphologically indistinguishable and genetically identical cells from the same tissue (*10*). Genetic tracing studies similarly reveal that all such cells are not equally prone to undergo neoplastic and malignant reprogramming (*11*). This heterogeneity suggests that for tumorigenesis to proceed, select mutant cells either have or gain an enhanced ability to change cell states, a phenomenon known as cellular plasticity (*12*, *13*).

Developmental, regenerative, and pathologic plasticity is largely determined at the chromatin level as increases or decreases in the repertoire of transcriptional programs that can be accessed by a given cell (*13*, *14*). Cells showing a high degree of plasticity, such as stem cells, often have a more "open" or accessible chromatin landscape that becomes restricted during differentiation (*15*, *16*). Previous work has used dedifferentiation with respect to normal cell states to characterize cancer cell plasticity with single-cell genomics in lung cancer models (*17*, *18*). However, we still do not know how plasticity emerges in the earliest stages of tumorigenesis, particularly in concert with the environmental insults that accelerate these initiating events. Learning how plasticity is triggered to arise in premalignant tissues and how it contributes to early tumor evolution is paramount to understanding and intercepting cancer at its earliest stages.

Pancreatic ductal adenocarcinoma (PDAC) is typically diagnosed too late for curative treatment and arises from cooperativity between genetic and epigenetic reprogramming events (*19*). Unlike more genetically heterogeneous cancers, PDAC is mainly initiated by an activating mutation in the proto-oncogene *KRAS*. However, *KRAS*-mutant epithelial cells can remain phenotypically normal and depend on inflammatory stimuli (pancreatitis) to transform into preneoplastic and neoplastic lesions (*20*–*22*). We (*23*) and others (*24*, *25*) have reported that oncogenic KRAS cooperates with inflammation to trigger large-scale chromatin-remodeling events that promote tumor initiation. However, important questions remain: How does mutant *KRAS*-mediated plasticity give rise to neoplastic lineages and enable their subsequent evolution to invasive disease? What are the critical cell-intrinsic and cell-extrinsic factors that determine a cell's propensity to acquire a plastic and ultimately a tumorigenic cell state? Understanding the answers to these questions may point to intervention strategies to prevent progression.

To shed light on neoplastic plasticity in PDAC, we compared physiological, premalignant, and malignant cell-state heterogeneity using single-cell genomics, applying computational methods and functional perturbation in autochthonous genetically engineered mouse models (GEMMs) that accurately recapitulate many aspects of the human disease. Beyond providing a comprehensive charting of epithelial dynamics from normal metaplasia through malignant tissue states, our approach allowed us to expose, quantify, and perturb early plasticity traits endowed by oncogene-environment interactions and define molecular, cellular, and tissue-level principles of premalignant tumor evolution.

Results

Targeted high-resolution profiling of epithelial dynamics during damage-induced neoplasia

The study of epithelial dynamics in pancreatic cancer has been limited by the inability to capture early and transitional cell states, which tend to be rare, short-lived, and difficult to identify. To characterize the full spectrum of epithelial cell states in both normal and pathological tissue remodeling, we generated a single-cell transcriptomic [single-cell RNA sequencing (scRNA-seq)] atlas of healthy, regenerating, benign neoplastic, and malignant epithelia using GEMMs that faithfully model cancer from initiation to metastasis. Our GEMMs incorporate a *Ptf1a*-Cre–dependent mKate2 fluorescent reporter to enrich pancreatic epithelial cells (*23*, *26*, *27*), allowing us to comprehensively profile pancreatic epithelial dynamics in well-defined tissue states. Specifically, we profiled pancreatic epithelial cells from healthy pancreas (N1) undergoing reversible metaplasia associated with normal regeneration after injury (N2) and the metaplasia-neoplasia-adenocarcinoma sequence that initiates PDAC in the presence of mutant *Kras* (K1 to K6) (Fig. 1A,

[1]Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA. [2]Tri-Institutional Training Program in Computational Biology and Medicine, Weill Cornell Medicine, New York, NY 10065, USA. [3]Cancer Biology and Genetics Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA. [4]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, 08028 Barcelona, Spain. [5]Clinical Cooperation Unit Virotherapy, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. [6]Department of Medical Oncology, National Center for Tumor Diseases, Heidelberg University Hospital, 69120 Heidelberg, Germany. [7]German Cancer Consortium (DKTK), 69120 Heidelberg, Germany. [8]Department of Pathology, Molecular and Cell-Based Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. [9]Department of Biomedical Engineering, Columbia University, New York, NY 10027, USA. [10]Alan and Sandra Gerry Metastasis and Tumor Ecosystems Center, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA. [11]Rosalind and Morris Goodman Cancer Institute, McGill University, Montreal, Quebec H3A 1A3, Canada. [12]Center for Epigenetics Research, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA. [13]Institute of Biotechnology, Life Sciences Centre, Vilnius University, 02158 Vilnius, Lithuania. [14]Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA.
*Corresponding author. Email: peerd@mskcc.org (D.P.); lowes@mskcc.org (S.W.L.)
†These authors contributed equally to this work.

fig. S1A, and table S1). In this setting, as in human cancer (*10*), *Kras*-mutant metaplasia is accelerated by an inflammatory insult (pancreatitis) (preneoplasia; K1 to K2 transition), proceeds to benign pancreatic intraepithelial neoplasia, (PanIN; K3 and K4), and ultimately, malignant PDAC (K5) and distal metastases (K6; Fig. 1A and fig. S1A).

Using a lineage-tagging reporter to enrich for epithelial (mKate2$^+$) cells, we captured both abundant and rare constituents of normal, regenerating, and *Kras*-mutant epithelia, such as progenitor-like tuft cells (*Pou2f3*$^+$*Dclk1*$^+$), epithelial-to-mesenchymal transition (EMT)–like cells (*Zeb1*$^+$), neuroendocrine cells (*Syp*$^+$*Chga*$^+$), and other previously reported cell subpopulations (*28*–*30*) (Fig. 1B; fig. S1, B to E; and table S2). We also characterized highly granular routes of acinar-to-ductal metaplasia (ADM) associated with regeneration and tumor initiation (*23*, *31*) (fig. S2, A and B). Compared with healthy and regenerating pancreata, we uncovered a staggering expansion of new cell states that emerge during the earliest stages of mutant *Kras*-driven tumorigenesis (K1 and K2; Fig. 1B; and figs. S1, B to E, and 2, A to C). Despite such heterogeneity, the distinct cell states captured within premalignant tissues were reproducible across biological replicates (individual mice) (fig. S2D). In stark contrast, and consistent with studies analyzing human PDAC (*32*), malignant tumors isolated from different mice showed extensive intertumor variability, only sharing one small cell cluster (fig. S2, D and E).

We next used diffusion maps (*33*) to characterize the major axes of transcriptional variation in our data, ordering cells along components associated with coherent gene expression patterns. The top component of variation closely matches progression from normal to regenerating, early tumorigenic, and finally late-stage disease, and is consistent with gene signatures that distinguish advanced human PDAC from normal pancreas (*34*, *35*). Specifically, genes up-regulated in human and mouse PDAC rise along the first diffusion component, whereas normal pancreas programs are down-regulated (Fig. 1C). Consistent with prior reports analyzing bulk RNA-seq data (*23*), the combined effects of *Kras* mutation and injury-driven inflammation are sufficient to induce signatures of human PDAC in premalignancy, as early as 24 to 48 hours postinjury (hpi) (Fig. 1, C and D). However, the added granularity of our single-cell analyses revealed that cancer-specific signatures are not induced uniformly across premalignant epithelial cell states; for example, some rare early *Kras*-mutant cells express high levels of EMT gene programs [*Zeb1* and *Vim* (*30*, *36*)] (Fig. 1D and fig. S1E). *Kras*-mutant cell states are also observed with varying degrees of dedifferentiation (down-regulation of acinar genes) and reactivation of developmental (*Clu*) or oncogenic (mutant *Kras* and *Myc*) programs (Fig. 1D). Thus, in the presence of inflammation, *Kras*-mutant pancreatic epithelial cells rapidly undergo specific and highly reproducible changes that endow select subpopulations with the capacity to activate disease-relevant programs long before malignant progression.

Aberrant, highly plastic cell states emerge early in PDAC progression

To map the cellular origins and processes underlying this diversity in transcriptional cell states, we first visualized heterogeneity in all *Kras*-mutant epithelia using a force-directed layout (FDL), which emphasizes cell-state transitions along axes toward malignancy. As expected, we found that the *Kras*-mutant pancreatic epithelium undergoes progressive gene expression changes that activate metaplastic [*Clu*$^+$*Krt19*$^+$ (*22*)], neoplastic [*Agr2*$^+$*Muc5ac*$^+$*Tff1*$^+$ (*37*)], and ultimately, invasive cancer [*Foxa1*$^+$ (*38*)] programs (Fig. 2A). The relatively low heterogeneity in apoptotic or proliferative signatures (fig. S3A) implies that much of the change in observed cellular states (K1 to K4) is likely caused by cell-state transitions rather than population dynamics.

To better characterize sources of cell-state variation, we applied CellRank (*39*), a data-driven approach that infers transcriptional dynamics from cell-cell similarity coupled with RNA velocity information (*27*, *40*, *41*). RNA velocities derived from the proportion of spliced to unspliced transcripts in each cell can indicate likely future states in neighboring phenotypic space. CellRank integrates directional information from per-cell velocity estimates with standard pseudotime inference based on cell-cell similarity to infer global transcriptional dynamics that can robustly pinpoint the origins of cell-state trajectories (*39*). Applying CellRank to early *Kras*-mutant cells acutely responding to an inflammatory insult (K1 and K2) identified multiple states that potentially act as distinct origins for the observed heterogeneity (fig. S3B).

The four inferred origin or "apex" states include one differentiated acinar (*Ptf1a*$^+$) and three dedifferentiated (*Nes*$^+$, *Aldh1b1*$^+$, *Tff2*$^+$) populations. Most of these states align with independent genetic lineage-tracing studies that demonstrate their "cell-of-origin" potential individually (*26*, *42*–*45*) (fig. S3, B and C). These relationships are further supported by single-molecule fluorescence in situ hybridization (smFISH) data derived from inflamed *Kras*-mutant tissue (K2), which revealed clear transitional states (*Nes*$^+$*Msn*$^+$*Anxa10*$^+$) in lesions containing both apex cells (*Nes*$^+$*Msn*$^+$) and the gastric-like cells (*Anxa10*$^+$) predominant in neoplastic tissue weeks later (K3 and K4; fig. S3, D and E).

Moreover, several of the inferred apex states are highly responsive to inflammation, with apparent cell-state shifts along the cell-cell similarity graph emerging in the context of tissue injury (K2). For instance, during pancreatitis, well-differentiated acinar cells generate a metaplastic population with transcriptional features that are intermediate for acinar (*Zg16* and *Cpa1*) and tumorigenesis-associated (*S100a6*) programs within 24 hpi (ADM-PDAC "bridge") (fig. S3F), and *Nes*$^+$ progenitor-like cells shift into a state showing reduced activation of tumor-suppressive programs (*Cdkn2a*). Our findings thus suggest that oncogenic *Kras* enables the emergence of diverse high-potential states not observed in healthy nor regenerating pancreata (Fig. 1B and fig. S2), each exhibiting distinct responses to inflammatory triggers, but all up-regulating cancer-associated programs (Fig. 1D).

An epigenetic basis for high plasticity states

Given the important role for chromatin dynamics in driving neoplasia (*23*), we hypothesized that the expansive phenotypic diversity in *Kras*-mutant apex states and their injury-driven progeny arises through a diversification of permissive chromatin states. To determine how chromatin dynamics correspond to changes observed in our longitudinal scRNA-seq atlas, we first analyzed bulk assay for transposase-accessible chromatin sequencing (ATAC-seq) data matching the above tissue stages (*23*, *27*). The dominant principal components of variation revealed accessibility patterns specific to each stage of progression (Fig. 2B). Compared with samples from normal pancreas epithelium (N1 and N2), the chromatin landscapes of preneoplastic *Kras*-mutant epithelia (K1 and K2) reproducibly shift toward states acquired in early neoplasia and sustained in advanced disease (K3 to K6). Nevertheless, we observed that the chromatin landscapes of benign neoplastic lesions (K3 and K4) and malignant tumors (K5 and K6) are highly divergent. Consistent with this, large sets of regulatory elements ("chromatin modules") exhibit mutually exclusive accessibility patterns across benign and malignant stages, with one set of ATAC-seq peaks showing increased accessibility in benign lesions but not in malignant counterparts (the benign neoplasia chromatin module) and another set behaving in the opposite manner (the malignant chromatin module; Fig. 2C and table S3). The modular structure of these data suggests that chromatin accessibility at benign (K3 and K4) and malignant (K5 and K6) stages corresponds to discrete, stable cell states, which may underlie the clearly distinct morphologies (fig. S1A) and expression patterns (fig. S2D) characteristic of cells of these advanced stages.

Mapping diverging benign neoplasia and malignant chromatin accessibility modules to single-cell gene expression reveals a concordant pattern. Genes proximal to benign neoplasia

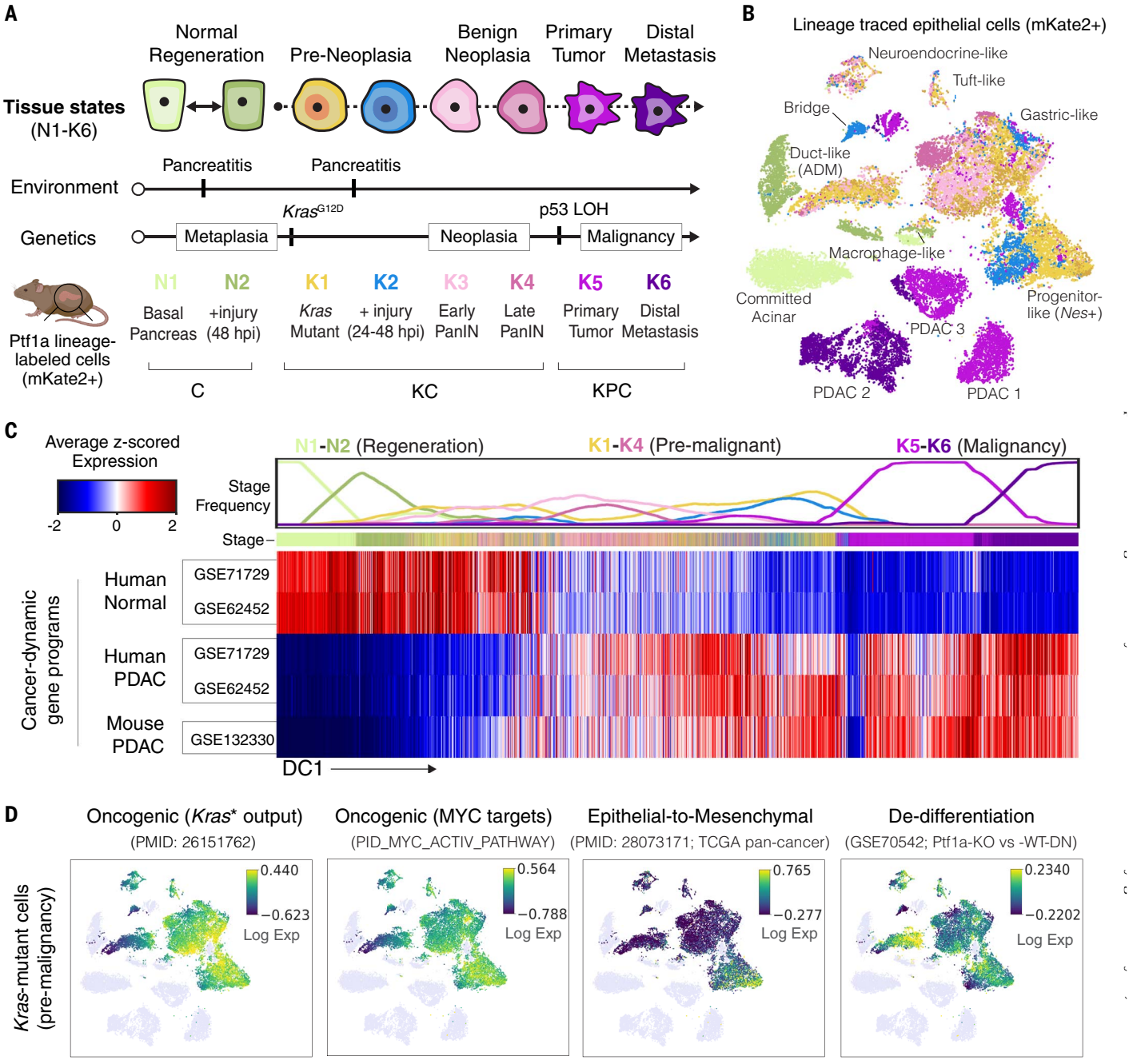


Fig. 1. A single-cell transcriptomic atlas of pancreatic regeneration and tumorigenesis. (**A**) Experimental design for tissue collection. GEMMs expressing *Ptf1a*-Cre enable FACS-based enrichment of mKate2-labeled exocrine pancreas epithelial cells (*23*). mKate2^{+} cells were isolated from wild-type *Kras* mice without injury with caerulein (N1) or 48 hpi (N2), and from *Kras*^{G12D} mice (KC genotype) without injury (K1) and 24 to 48 hours (K2) or 3 weeks after caerulein (K3, PanIN stage), as well as uninjured older KC mice (K4). PDAC primary tumors (K5) and liver and lung metastases (K6) were harvested from KC mice with a *p53* floxed (*p53*$^{fl/+}$) or mutant (*p53*$^{R172H/+}$) allele (KPC genotype). Mouse illustration was created with BioRender (https://biorender.com/). (**B**) t-distributed stochastic neighbor embedding (tSNE) visualization of pancreatic epithelial scRNA-seq profiles from all collected stages (n = 17 mice), colored as in (A) and labeled by cell state (*27*). ADM denotes cells undergoing acinar-to-ductal metaplasia (*31*) and "bridge" denotes cells between acinar-like and malignant programs, which express genes from both. (**C**) Expression of PDAC-associated gene sets (rows) across all pancreatic epithelial cells (columns) (*34*, *35*). Cells are ordered by the first diffusion component (DC1), representing the major axis of progression from normal (N1) to metastatic (K6) states. Plot at top displays frequency (from 0 to 1) of cells per stage in bins of n = 2000 cells. Gene set score for each cell is computed as the average of log-normalized expression, z scored for each gene to obtain a comparable scale. Heatmap is standardized to compare cells within each gene set. (**D**) tSNE plots as in (B), with premalignant (K1 to K4) *Kras*-mutant cells colored by the expression of genes (from left to right) up-regulated in bulk RNA-seq of *Kras*-mutant (*Kras**) pancreas relative to normal (*67*), associated with Myc activity (*68*), EMT (*36*), or down-regulated upon *Ptf1a* knockout (*67*). Colors are scaled from the 5th to the 95th percentile of expression.

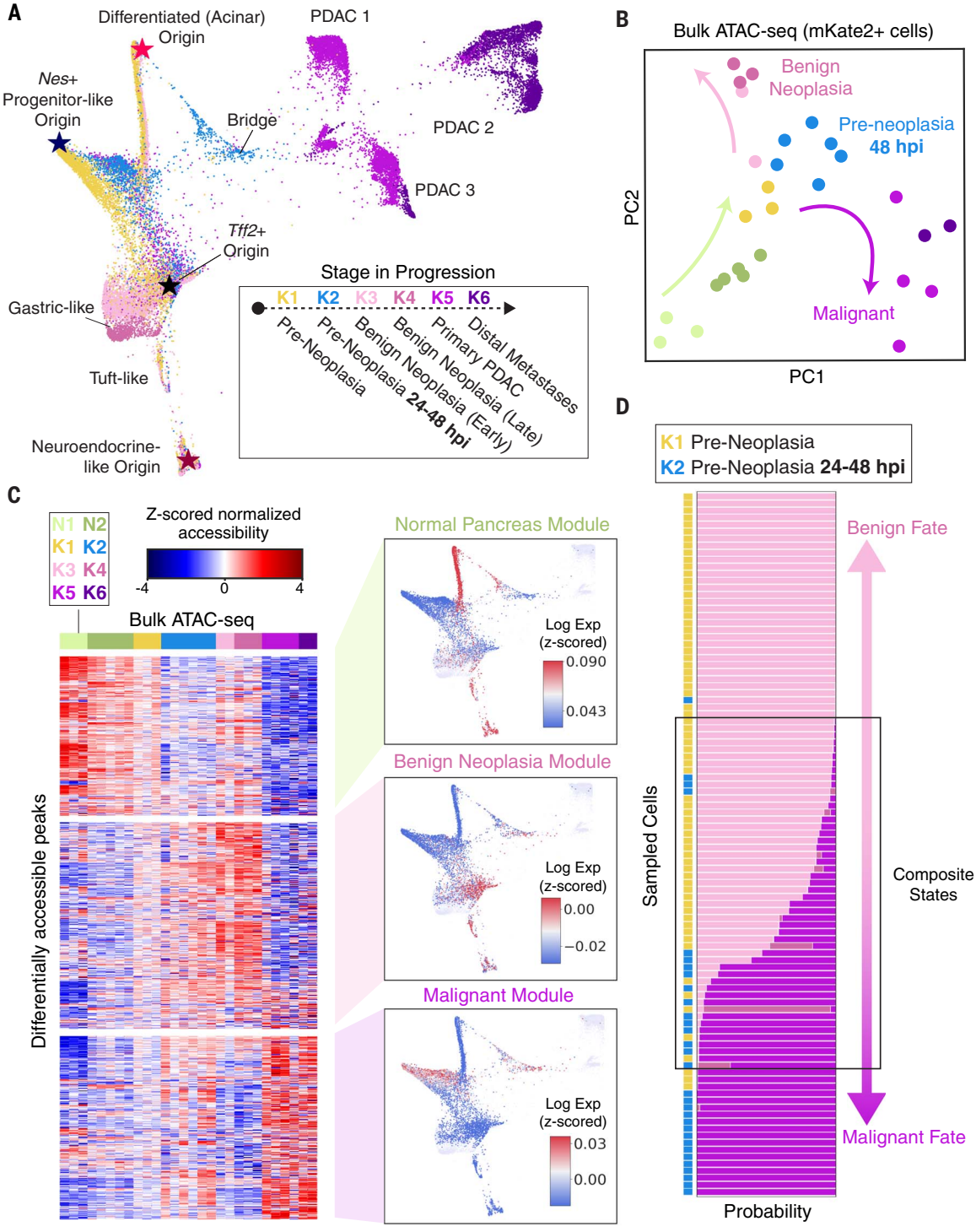


Fig. 2. Differential epigenetic priming of *Kras*-mutant cells. (**A**) FDL of all *Kras*-mutant scRNA-seq profiles (K1 to K6; n = 11 mice). Cells are colored by stage as in Fig. 1A. Stars highlight "apex" states inferred by CellRank (*39*) (fig. S3B). (**B**) Principal component analysis (PCA) of bulk ATAC-seq profiles from pancreatic epithelial cells. Each point shows the position of a single biological replicate (individual mouse), colored by stage as in (A). Arrows indicate a transition upon injury and *Kras* mutation (N1 to N2 and K1 to K2; green arrow) and a divergence between benign neoplastic (K3 to K4; pink arrow) and malignant (K5 to K6; purple arrow) stages. (**C**) Left: Chromatin accessibility along progression. Subsets of differentially accessible ATAC-seq peaks (rows) are organized into three modules by clustering (*27*); bulk ATAC-seq replicates (columns) are ordered and colored by stage as in (A). Peaks organize into distinct accessibility patterns, denoted as chromatin modules (*27*). Right: Expression of genes corresponding to chromatin accessibility modules in preneoplastic cells (K1 and K2). FDL map as in (A), colored by module expression score computed by z scoring each cell to emphasize dominant gene programs per cell and averaging genes nearest to module peaks. Color (expression scores) are scaled between the 40th and the 90th percentiles. (**D**) Probability of classifying preneoplastic cells (K1 and K2) as more similar to benign neoplastic (K3 and K4) or malignant (K5 and K6) scRNA-seq profiles, based on expression similarity. Sampled cells (rows) are ordered from highest benign fate probability (top) to highest malignant fate probability (bottom); bars represent probability of classification from 0 to 1 to K3, K4, K5, or K6 labels, colored as in (A). A fraction of cells exhibit composite states with probability for both fates.

module loci are up-regulated in premalignant disease (K3 and K4), and genes proximal to malignant module loci increase in malignant disease (K5) (fig. S4A). Chromatin modules associated with advanced stages are induced very early in tumor development, such that benign neoplasia chromatin module– and malignant chromatin module–associated genes are expressed and restricted to transcriptionally distinct populations within 24 to 48 hpi (K1 and K2; Fig. 2C and table S3). Furthermore, bulk ATAC-seq data show an initial increase in accessible chromatin in both modules from samples collected at 48 hpi or without injury (K1 and K2), well before the emergence of benign lesions or malignant disease (K3 to K6; Fig. 2C). These observations imply that the transcriptional diversity of preneoplastic *Kras*-mutant cells is established at the chromatin level and involves benign neoplasia or malignant module activation before the development of PanINs or PDAC. This activation requires both oncogenic *Kras* and inflammation, because these programs are not similarly accessible or expressed in normal regeneration (N1 and N2) (Fig. 2C and fig. S4A).

The early establishment of a permissive chromatin landscape (K1 and K2) that is later specified into a restricted, distinct set of accessible regulatory elements (K3 to K6) is reminiscent of cell-fate determination occurring in developmental systems (*16*). We thus refer to the cell states that preneoplastic *Kras*-mutant cells may eventually acquire as "cell fates" associated with benign neoplasia (K3 and K4) or malignancy (K5 and K6). We postulated that preneoplastic *Kras*-mutant cells (K1 and K2) expressing programs associated with the chromatin landscape of a single distinct fate (benign or malignant) may be epigenetically primed toward that fate, conferring greater propensity to acquire its phenotype over time or in response to certain exogenous triggers. We further reasoned that similarities between the transcriptomes of *Kras*-mutant cells from preneoplastic (K1 and K2) and later neoplastic stages (K3 to K6) would indicate such fate potential. We therefore developed a classification-based approach that first identifies gene expression patterns that accurately discriminate between cell populations in benign lesions or cancer, and then uses these patterns to assign cell-fate probabilities to preneoplastic cells on the basis of the activation of fate-associated genes. Specifically, we trained a logistic-regression classifier to distinguish between benign neoplasia (K3 and K4) and malignancy (K5 and K6), and used it to classify preneoplastic (K1 and K2) cells (*27*). This classifier was highly accurate (99%) in assigning fate to PanIN and PDAC cells of known fate and identified a set of discriminative genes that have been linked to either fate (fig. S4B).

Applied to preneoplastic cells, this approach indeed pinpointed *Kras*-mutant cells that are strongly skewed toward one or the other fate (Fig. 2D and fig. S4C). Most preneoplastic cells are classified as only having the potential to acquire a single fate, with cells responding to inflammation (K2) assigned a higher probability of acquiring a malignant fate. We also identified an interesting set of *Kras*-mutant cells that are not well classified (Fig. 2D and fig. S4D), most of which express a composite program of otherwise divergent fate-associated genes (fig. S4E). These dual-primed subpopulations exist largely in the absence of tissue damage and overlap with initiating apex states (*Ptf1a*$^+$ acinar and *Nes*$^+$ progenitor) captured independently by CellRank (fig. S3, B and C).

Collectively, our results imply that tumorigenesis can proceed from multiple well-differentiated or progenitor-like states, and that their neoplastic progression is not dictated solely by cell-intrinsic determinants (*Kras* gene mutation), but rather is affected by inflammatory signals that epigenetically prime them toward diverse fates that can be predicted early in disease progression.

Epigenetic plasticity is enhanced by inflammation

To map the epigenomic landscape at higher resolution, we generated single-cell chromatin accessibility (scATAC-seq) profiles of preneoplastic (K1), preneoplastic inflamed (K2), benign neoplastic (K3), and adenocarcinoma (K5) epithelia. Consistent with an epigenetic basis for the observed premalignant diversity (Fig. 2C), we found considerable heterogeneity in chromatin accessibility within *Kras*-mutant epithelial cells at each stage (Fig. 3A and fig. S5A). A major axis of variation in accessibility reproduced the divergence between benign and malignant fates seen in bulk analyses (Fig. 3B and fig. S5B). Substantial variation in accessibility near fate-associated genes occurred across both stages and clusters (fig. S5C), with *Kras*-mutant apex cells exhibiting a composite state defined by open chromatin at benign-associated and malignant-associated loci. This pattern extends to variation in open chromatin near other genes that define the benign and malignant chromatin modules (fig. S5D). These data support bona fide epigenetic priming of divergent fate-associated programs in early, preneoplastic *Kras*-mutant cells.

To better connect primed chromatin landscapes to their transcriptional outputs, we next sought to integrate scATAC-seq and scRNA-seq profiles from comparable stages. Clustering and cell-state annotation demonstrated that cell states derived from scRNA-seq data largely match those derived from scATAC-seq data at the broad cluster level, including those corresponding to *Nr5a2*$^+$ acinar cells, *Neurod1*$^+$ neuroendocrine cells, *Pou2f3*$^+$ tuft cells, and *Nes*$^+$ progenitor cells (fig. S5A). However, we also found substantial epigenomic heterogeneity within each scATAC-seq cluster. To explore this heterogeneity in more detail, we applied an algorithm that aggregates highly similar cells into granular cell states, or metacells (*27*, *46*, *47*). Metacells provide much higher resolution than clusters, but aggregate cells sufficiently to reduce sparsity and improve statistical power for comparison.

After separately identifying metacells for each scRNA-seq and scATAC-seq modality, we developed a framework to map between them that is based on similarity between a gene's expression and its proximal chromatin accessibility (fig. S6, A to C) (*27*). This integrative analysis showed the expected correspondence between the accessibility and expression programs of comparable cell states (Fig. 3C). However, we also observed extensive off-diagonal correspondence, indicating that features of the chromatin landscape are shared across diverse gene expression states. Specifically, we found diverse premalignant transcriptional states (tuft, neuroendocrine, progenitor, and gastric) to broadly correlate with the ADM epigenomic state, reflecting the known acinar history of these *Ptf1a* lineage–sorted cells (*23*, *26*). In other cases, these correspondences may indicate widespread transcriptional "poising" of regulatory elements near unexpressed genes. Such effects were particularly evident in apex *Nes*$^+$ progenitor cells, which exist in preneoplastic tissues at 48 hpi, but establish chromatin landscapes that are highly correlated with those of late-stage malignant populations (Fig. 3C and fig. S5B). Although these similarities can be partially explained by lineage relationships within each data modality, the greater off-diagonal (inter-cell-state) correlation existing across modalities (fig. S6D) suggests that subpopulations of preneoplastic *Kras*-mutant cells are epigenetically primed to engage neoplasia transcriptional programs later in progression (Fig. 2C).

We sought to quantify the degree of epigenetic plasticity, which we define as the amount of diversity in transcriptional phenotypes that is enabled (or restricted) by a given chromatin accessibility landscape. To first determine these potential transcriptional phenotypes, we used a simple classifier to identify gene expression patterns that discriminate cell states. Assuming that proximal open chromatin conveys the potential for a gene's activation, we then applied the classifier to predict cell states based on accessibility proximal to genes rather than gene expression (Fig. 3D). We reasoned that for a given epigenomic state, uncertainty in such predictions serves as a measure of epigenetic plasticity. Following this logic, high plasticity is characterized by many accessible loci that define multiple discrete transcriptional states and thus produce high classifier prediction uncertainty. By contrast, low plasticity is defined by restricted potential diversity and prediction certainty. Applying this approach

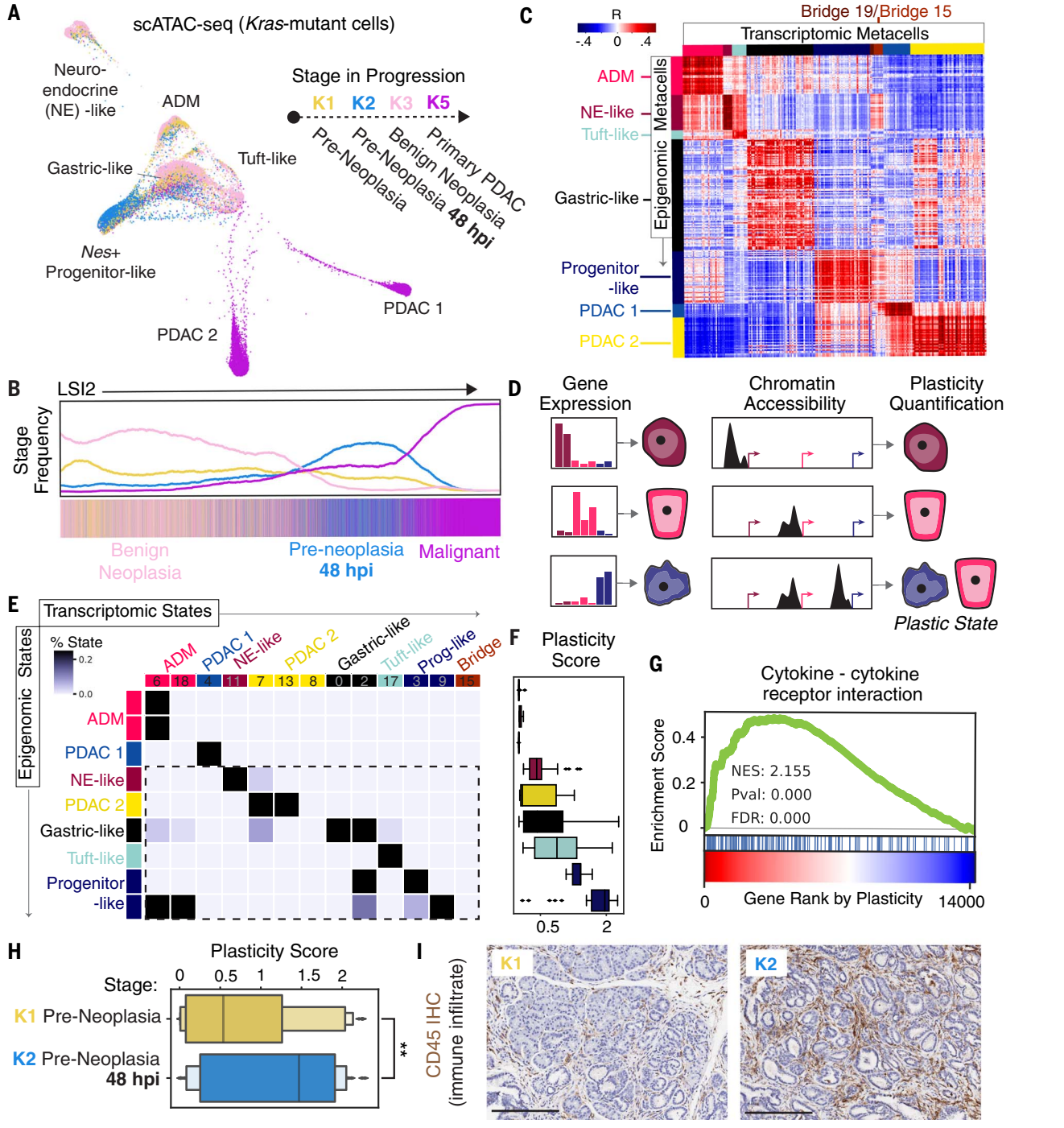


Fig. 3. *Kras*-mutant cells display elevated epigenetic plasticity, which is associated with cell-cell communication propensity. (**A**) FDL of scATAC-seq profiles from *Kras*-mutant epithelial cells from premalignant (K1 to K3) and malignant (K5) stages (n = 9 mice) colored by stage. (**B**) Frequency of cells from each stage along second high-variance component from latent semantic indexing (LSI) of scATAC-seq profiles (*65*). (**C**) Pairwise Pearson correlation coefficients of metacells from scATAC-seq ArchR gene accessibility scores (rows) and scRNA-seq expression values (columns). Annotated cell states, determined by refined PhenoGraph clustering of scRNA-seq (fig. S1C) and scATAC-seq (fig. S5A), are colored according to their annotation. Blocks of positive correlation along diagonal represent similar cell states across the two modalities, and off-diagonal correlations indicate similarity across distinct cell states. (**D**) Cartoon of classifier-based approach to quantify plasticity (*27*). (**E**) Classifier confusion matrix based on procedure in (D). Cell states, determined by scRNA-seq (fig. S1C) and scATAC-seq (fig. S5A) metacell clusters, are colored as in labels in (C). Values represent number of metacells from an epigenomic cluster that classify to a transcriptomic cluster, normalized within each row. Dashed box highlights high plasticity epigenomic states. (**F**) Plasticity scores for epigenomic clusters in (E). Boxes represent interquartile range (IQR) of plasticity scores for all epigenomic metacells assigned to that cluster, computed as per-cell Shannon entropy in the classifier's predicted probability distribution across transcriptomic states. Lines represent medians and whiskers represent 1.5 × IQR. (**G**) GSEA plot based on Spearman rank correlation between plasticity score and each gene's accessibility score. (**H**) Plasticity scores for epigenomic metacells from K1 and K2, showing significant increase in plasticity in K2 (P = 0.006, one-tailed t test, t = 2.5511). (**I**) IHC of CD45 (brown) marking immune cells in K1 and K2 tissue, showing increase in immune infiltrate in response to injury. Scale bar, 200 μm.

to epigenomic metacells identified populations of varying plasticity (Fig. 3, E and F, and fig. S6E), with the most plastic states exhibiting substantial overlap with the apex cells identified by CellRank (such as *Nes*$^{+}$ progenitor-like and *Tff2*$^{+}$ gastric-like cells) and experimentally validated cells-of-origin from lineage tracing studies [such as *Pou2f3*$^{+}$ tuft cells (*28*, *42*–*44*)] (Fig. 3F). Some of these states arise largely from preneoplastic conditions (K1 and K2), aligning with observations on priming toward future neoplastic states. All plastic states identified in this analysis have no clear analog in normal or regenerating pancreata (Fig. 1B and fig. S2, A and B), although a deeper exploration of physiological plasticity would be required to fully contrast normal and disease mechanisms.

To expose potential unifying features of distinct plastic cell states, we used gene set enrichment analysis (GSEA) (*48*) to identify gene signatures within populations displaying high plasticity scores (table S4). This analysis revealed a robust and consistent up-regulation of sets related to cell-cell communication (fig. S6F), with the cytokine–cytokine receptor interaction gene set yielding the top association (normalized enrichment score = 2.155, adjusted *P* value = 0.000) (Fig. 3G). A substantial fraction of these plasticity-associated genes encoded inflammatory mediators, receptors, or ligands involved in cell-cell communication, including those previously associated with malignant progression [*Csf2*, *Cxcl1*, and *Cxcr2* (*49*)] (table S5). Accordingly, plasticity increases significantly (*P* = 0.006; one-tailed *t* test, t = 2.5511) upon injury in the context of *Kras* mutation (K2 versus K1) (Fig. 3H), suggesting an interplay between highly plastic cells and immune infiltrates flooding the preneoplastic tissue environment in this context (Fig. 3I). Together, our results indicate that epithelial plasticity in premalignant cells is directly associated with an increased, epigenetically encoded propensity for ligand-receptor–mediated communication with the immune microenvironment.

Calligraphy charts cell-state–specific communication repertoires and their interactions

The dominance of the association between plasticity and cell-cell communication drove us to investigate how heterotypic interactions may result from plasticity or enhance it in the *Kras*-mutant pancreatic epithelium. We hypothesized that chromatin remodeling of receptor and ligand gene loci (hereafter referred to as "communication genes") contributes to plasticity in preneoplasia by enabling cells to respond to inductive signals from the environment.

The delineation of communication events requires an assessment of the communication propensities of each cell state (defined by its expressed receptors and ligands), which may then be used to link interacting cell states on the basis of prior knowledge of receptor-ligand (R-L)–binding partners. As a first step, we characterized communication gene accessibility and expression across cell states of the premalignant epithelium using the scATAC-seq and scRNA-seq datasets generated above. Each plastic cell state reveals substantial variability in chromatin accessibility near communication genes, consistent with distinct molecular repertoires for potential communication (fig. S7A). To identify trends of coordinated gene expression, we searched for coexpression between any two communication genes (testing all combinations of two receptors, two ligands, and one of each) in individual cells across the premalignant epithelium, and found a high degree of block structure in pairwise coexpression. This pattern implies that communication capabilities are driven by "communication modules": sets of communication genes that are mutually expressed in the same cell populations (Fig. 4A).

We next sought to infer actual cell-cell signaling interactions that may occur between cells expressing different communication modules. Although several methods have been developed to predict cellular interactions from single-cell data (*50*), their inference relies on weak signals arising from the noisy expression of a single cognate R-L pair across fixed cell states. We therefore developed our own approach, Calligraphy, which leverages the observed modularity in communication gene expression to infer potential cell-cell signaling events (*27*). Calligraphy first identifies communication modules, thereby establishing the incoming and outgoing communication that each cell state can participate in. Next, Calligraphy identifies communication events between cell states on the basis of prior knowledge of cognate R-L–binding partners. Unlike previous methods, which test interactions on individual R-L pairs, Calligraphy draws inferences across entire sets of genes, making the output insensitive to noise in any single gene (*50*).

Using Calligraphy (*27*), we obtained seven communication modules of genes that are coexpressed across the premalignant pancreatic epithelium (Fig. 4A and table S6) and are reproducible in situ in smFISH data (fig. S7B). Mapping the average expression of communication genes back onto the premalignant epithelium revealed that most cells express a single dominant module, making it possible to annotate cells by their corresponding module (Fig. 4B). Cell states defined solely by communication gene expression coincide with those identified by clustering the entire transcriptome (fig. S7, C to E). We observed similar patterns in scATAC-seq data, where each subpopulation maintains open chromatin around genes from distinct communication modules, supporting an epigenetic basis for the emergence of these programs (fig. S7F).

Normal and regenerating epithelial cells (N1 and N2) showed much less diversity in communication module expression compared with their *Kras*-mutant counterparts, with most cells maintaining very low module expression (Fig. 4C). Among cells with wild-type *Kras*, a small injury-induced population does express communication modules associated with the gastric (E6) cell state, likely reflecting the expected trans-/dedifferentiation of acinar cells under inflammatory conditions (*51*). These cells also expressed high levels of transcripts derived from a mutant *Kras* gene signature (*23*), implying that high-level RAS signaling may play a role in normal regeneration and that KRAS mutation might stabilize such transient, injury-induced communication modules.

Moreover, communication modules established in the premalignant pancreas are maintained in advanced cancers (K5 and K6), with most cells expressing at least one of the gastric (E6), progenitor (E7), or bridge (E3) modules (Fig. 4D). These modules (as well as their corresponding cell states) are observed in an analogous mouse model of preneoplasia with activation of mutant *Kras* in adult acinar cells (*52*) (fig. S8). Furthermore, these communication modules are conserved in human PDAC derived from multiple patients (*32*) (Fig. 4E). The distinct behavior of communication modules in early neoplasia across multiple model systems and their persistence in advanced murine and human PDAC implies a functional role in pancreatic tumorigenesis.

Extensive epithelial-immune interactions drive oncogenic tissue remodeling

The distinct communication modules that arise in *Kras*-mutant epithelial cells during inflammation compared with normal regeneration implicate one or more signaling nodes in early PDAC development (*20*–*22*). Tissue damage produces inflammation and changes in immune cell state and composition that contribute to neoplasia (*53*); thus, we investigated how mutant *Kras*-driven epithelial communication modules interact with infiltrating and tissue-resident immune cells. scRNA-seq analysis of immune cells (CD45^{+} sorted) from *Kras*-mutant tissues before (K1) and after (K2 and K3) induction of pancreatitis identified all of the expected immune subtypes, including both abundant (macrophage) and rare [regulatory T (T_{reg}) cells and innate lymphoid cell (ILC)] types (fig. S9) (*27*). As expected, injury-induced inflammation causes substantial remodeling of the immune cell landscape, including the enrichment or depletion of specific lymphoid and myeloid cell states (fig. S10A).

Applying Calligraphy to these data identified consistent and structured communication modules defining distinct immune populations. To achieve even greater resolution, we ran Calligraphy separately on T cells/ILCs/natural killer

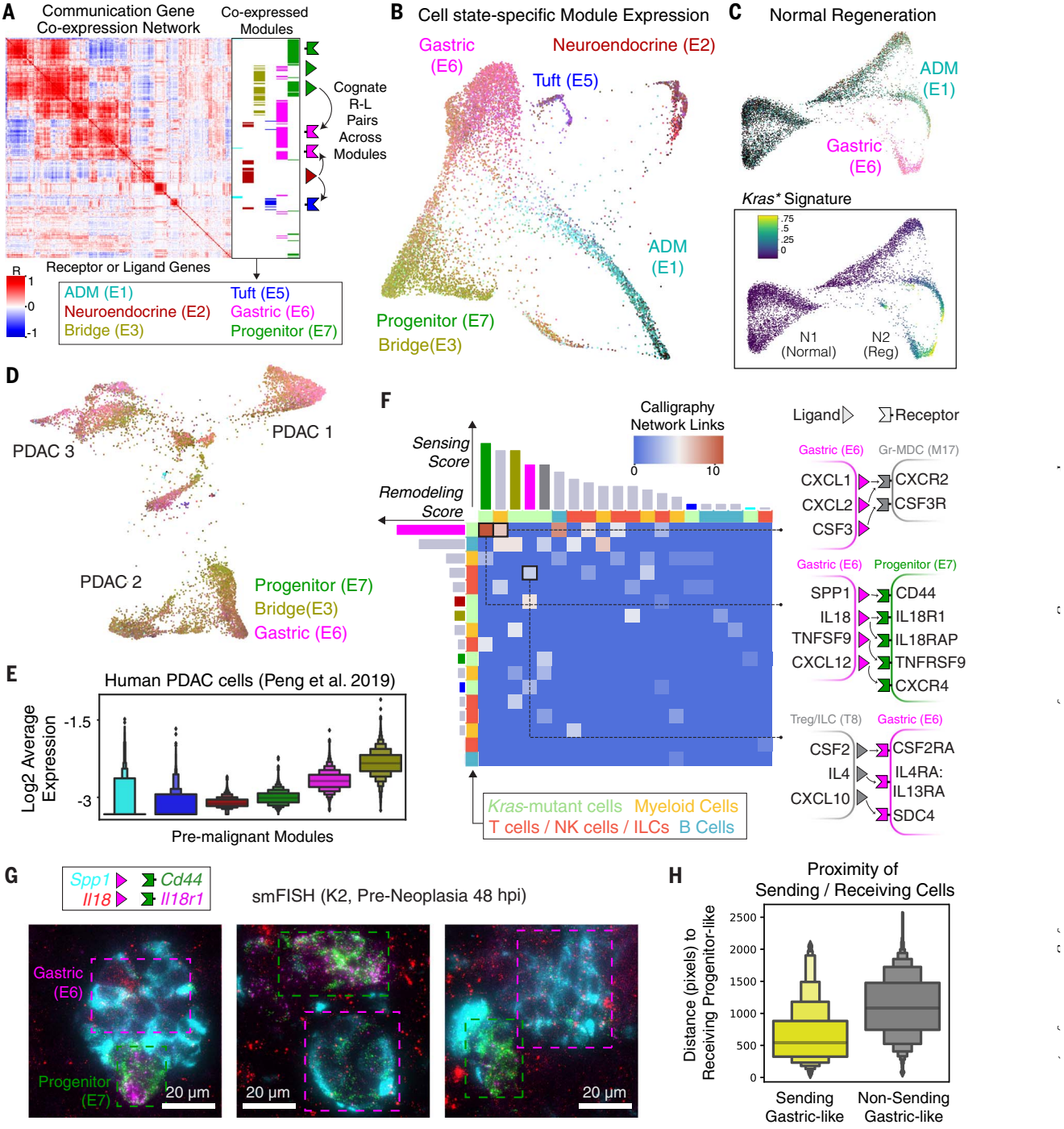


Fig. 4. Inferred epithelial-immune cross-talk in plastic neoplastic states. (**A**) Calligraphy-inferred "communication modules" in premalignant *Kras*-mutant epithelial cells (K1 to K3, n = 6 mice). Each row or column represents one receptor or ligand; value at intersection indicates correlation in expression (Pearson's r) of that gene pair across premalignant cells. Blocks of highly correlated genes denote partially overlapping modules (annotated at right) that tend to coexpress in the same cell states. Schematic (far right) describes the second step of Calligraphy (fig. S11A). (**B**) FDL of K1 to K3 epithelial cells with color values based on relative communication module gene expression (*27*). (**C**) FDL of *Kras*^{WT} pancreas cells before and after injury (N1 and N2, n = 4 mice), colored by K1 to K3 communication module expression as in (B) (top) or *Kras* mutant signature gene expression [bottom, (*23*)], scaled between the 1st and the 99th percentile. (**D**) FDL of malignant cells (K5 and K6, n = 3 mice), colored as in (B). (**E**) Communication module expression in human pancreatic tumor scRNA-seq data (*32*), colored as in (B). (**F**) Pairwise cross-talk between communication modules inferred by Calligraphy from epithelial or immune scRNA-seq data (one module per row or column), colored gray for immune or as in (A) for epithelial modules. Heat values represent number of inferred cognate R-L pair interactions across each communicating module pair; some contributing receptors or ligands are shown at right. Bars quantify total inferred edge counts, representing remodeling (row) or sensing (column) interactions for that module. (**G**) Two smFISH fields of view reveal the spatial proximity of sending (magenta box) to receiving (green box) mKate2^{+} epithelial cells. The expression of two gastric (E6) module ligands (cyan and red), as well as two progenitor (E7) receptors (magenta and green) overlap spatially in these three examples. Scale bars, 20 μm. (**H**) Distance between each receiving progenitor cell (*Il18r1*^{hi}*Cd44*^{hi}) and double-positive sending gastric cell (*Il18*^{hi}*Spp1*^{hi}), versus randomly selected nonsending gastric cells (*Il18*^{lo}*Spp1*^{lo}).

(NK) cells, myeloid cells, and B cells, and found numerous modules containing known regulators, as well as candidates for pancreatic tumorigenesis (fig. S10, B and C). For example, T cell/ILC/NK cell module 8 is highly expressed in ILC2, ILC3/LTi, and T_{reg} cells; these cells express the receptor and co-receptor *Il1rl1*/*Il1rap* for interleukin-33 (IL-33), a ligand that accelerates the formation of mucinous PanIN lesions (*23*).

Reasoning that such rapid immune and epithelial remodeling could arise through heterotypic cross-talk in pancreata undergoing neoplastic transformation, we used a feature in Calligraphy to nominate potential cell-cell interactions that drive this process (fig. S11A) (*27*). To limit our search to tumorigenesis-specific cross-talk after tissue inflammation, we filtered Calligraphy modules to retain those cognate R-L pairs in which at least one partner is selectively up-regulated in *Kras*-mutant (K2) relative to *Kras*-wild type (N2) epithelial cells (*23*). This filtering reduced the space of possible interacting molecules from 340 total communication genes down to 55 receptors and 46 ligands potentially involved in tumorigenesis-associated communication. We then assumed that two cell states potentially interact if their associated modules are enriched in the number of shared cognate R-L pairs spanning them (*27*). Whereas CellphoneDB (*50*) predicts a highly dense network of 720 out of 729 (98.8%) possible interactions, involving nearly all pairwise combinations of cell states, Calligraphy identified a sparser, more interpretable network of potential neoplasia-specific interactions between the *Kras*-mutant epithelium and the immune environment (5.6% of possible interactions; fig. S11, B and C, and table S7).

Within Calligraphy's context-specific network were apparent "master communication hubs" that participate in numerous interactions. We calculated a receiving score (i.e., the ability to sense the environment through expressed receptors) and a transmission score (i.e., the ability to remodel the environment through expressed ligands) based on the number of Calligraphy's statistically significant incoming and outgoing edges for each module ($P < 0.1$) (*27*). The two most prominent hubs for transmitting and receiving interactions are the epithelial gastric (E6) and progenitor (E7) modules, respectively (Fig. 4F), which correspond to the "high-plasticity" populations identified above. These same communication hubs are enriched in advanced mouse and human PDAC (Fig. 4, D and E).

To validate predicted communication networks, we mapped interacting modules to their spatial context, leveraging smFISH data including probes for transcripts marking distinct cell states and their corresponding communication genes. We identified cells with concordant communication gene expression patterns across space for module E6-expressing gastric cells (*Anxa10*$^+$*Il18*$^+$*Spp1*$^+$) and module E7-expressing progenitor cells (*Nes*$^+$*Il18rl*$^+$*Cd44*$^+$) (Fig. 4G and fig. S7B) in preneoplastic cells subject to inflammation (K2). Calculating distances between receiving (progenitor-like) cells and their closest sending (gastric-like) cells against a random control set of gastric-like cells that do not express these ligands showed a significant enrichment ($P < 0.01$, *t* test) of receiving cells in the vicinity of their interacting sending cells (Fig. 4H).

Neoplastic tissue remodeling involves feedback communication loops

One of our most notable observations is the marked remodeling of epithelial and immune compartments within 24 to 48 hpi (figs. S2D and S10A). This remodeling is highly reproducible and expansive, involving numerous new cell states that are quickly adopted by most cells. The dynamics of such a rapid and robust response suggest a feedback loop by which immune cell intermediates may amplify tumor-promoting epithelial cues (*54*). We probed the Calligraphy module-module interaction network to systematically enumerate cycles involving any epithelial or immune subsets and identified only one feedback loop in the system (Fig. 5A) (*27*). The loop includes the epithelial gastric (E6) hub module, which is maintained in late disease, and further involves cytokines and receptors with reported roles in KRAS-driven pancreatic tumorigenesis, including IL-1A, IL-33, and IL-4RA (*55*, *56*). Specifically, it engages T_{reg} and ILC2 cells through IL-33 signaling before feeding back to *Kras*-mutant epithelial cells (Fig. 5B).

Il33 is expressed during pancreatitis by a small subset (4%) of TFF1/ANXA10$^+$*Kras*-mutant gastric module–expressing epithelial cells and is predicted to initiate signaling to T_{regs} and ILC2s (module T8) by binding with its cognate receptor *Il1rl1* and coreceptor *Il1rap* (Fig. 5B and fig. S12, A to D). Supporting the relevance of these interactions, immunofluorescence (IF) data reveal that IL-33–expressing epithelial cells (IL-33$^+$ mKate2$^+$) and rare T_{regs} (Foxp3$^+$) are in close spatial proximity in *Kras*-mutant pancreata under injury conditions (distance versus randomly permuted positions, $P < 0.01$, *t* test, for all IF images collected across five independent mice) (Fig. 5, C and D, and fig. S12E). Subsequently, many receiving cells that express *Il1rl1* (module T8) also express the T helper cell 2 (T_H2) cytokine gene *Il4* (odds ratio = 21.47, $P = 9.88 \times 10^{-35}$, Fisher's exact test), consistent with the known role of IL-33 in triggering T_H2-type immune responses (*57*). Module T8 cells then apparently signal through IL-4 (*Il4*) back to the gastric module (E6) through the IL-4 receptor (*Il4ra*), thereby closing the loop and potentially propagating signals to other module-expressing cells from both immune and epithelial compartments (Fig. 5B).

The broad expression of IL-4RA across *Kras*-mutant epithelial cell states, including gastric-like cell, tuft cell, and *Nes*$^+$ progenitor-like cell populations (Fig. 5B and fig. S12, A and C), implies that this signaling loop has a system-wide impact on premalignant tissue (45% of premalignant epithelial cells appear affected). By contrast, few wild-type normal pancreas cells express both sending (*Il33*) and receiving (*Il4ra*) factors and do so at low levels, even during injury-induced regeneration (fig. S12, F and G).

A particular strength of Calligraphy is its ability to construct communication circuits linking a cascade of signaling events between multiple communication modules in a serial fashion, thus mapping cell populations that are potentially both directly and indirectly affected by epithelial-derived IL-33. Calligraphy predicts that IL-33–driven communication has a large impact on premalignant tissue, and the proportion of affected tissue increases through signaling cascades between communication modules that each use multiple cognate R-L pairs, ultimately reaching most of the premalignant pancreas (72% of cells; Fig. 5, E and F). Although it is unlikely that the IL-33 loop is solely responsible for KRAS-driven tumor progression in the context of the high complexity of observed intercellular communication in the premalignant tissue, the number of populations that appear directly and indirectly affected by epithelial IL-33 expression suggests that this communication circuit plays an important role in driving tumorigenesis.

KRAS-dependent IL-33 feedback loop directs rapid tissue remodeling in early tumorigenesis

Whereas previous studies have implicated stroma-derived IL-33 in disease phenotypes (*58*, *59*), Calligraphy identified a feedback loop driven by IL-33 expressed from *Kras*-mutant epithelial cells. To determine how IL-33 derived specifically from the epithelium contributes to early neoplasia, we developed a GEMM that enables specific *Il33* suppression in lineage-traced *Kras*-mutant epithelial cells. Animals were produced from multiallelic embryonic stem cells engineered to harbor a conditional *Kras*^{G12D} allele together with a doxycycline (dox)–inducible, green fluorescent protein (GFP)–coupled short hairpin RNA (shRNA) capable of suppressing *Il33* (KC-sh*Il33*), allowing potent *Il33* suppression in the epithelial compartment after dox administration (Fig. 6, A and B). In addition, a separate cohort of animals was produced harboring a neutral shRNA (shRen) as controls. scRNA-seq and spatial imaging [imaging mass cytometry (IMC)] were performed on each model assessing epithelial and immune compartments at an early time point (48 hpi), when inflammation unleashes neoplastic remodeling (K2), and later

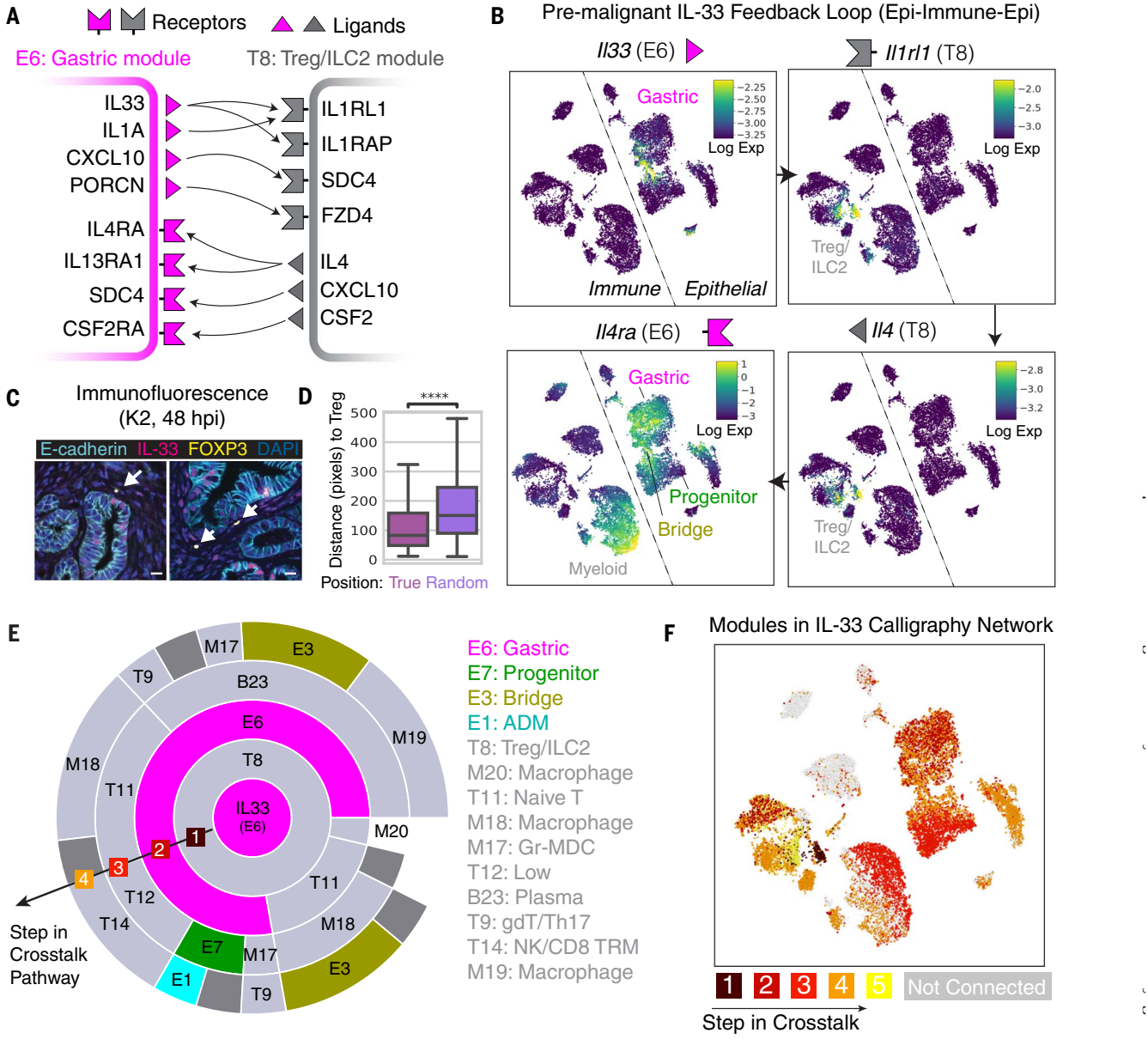


Fig. 5. *Kras*-mutant epithelial states participate in a feedback loop with immune populations. (**A**) A feedback loop identified by Calligraphy in the pre-malignant pancreas. Arrows depict cognate R-L interactions. (**B**) tSNEs of immune and epithelial scRNA-seq data from premalignant stages (K1 to K3, n = 6 mice), displaying imputed expression (*69*) of key genes from the loop in (A). Arrows between plots indicate sequential steps of the loop. (**C**) Co-IF images showing coexpression of IL-33 and E-cadherin (an epithelial marker) and apposition of FOXP3-expressing T_{regs} (arrows) and IL-33–expressing epithelial cells. Scale bar, 10 μm. (**D**) Distance in pixels (0.325 μm per pixel) of IL-33$^+$ epithelial cells to T_{regs} against a null model of spatial distribution in co-IF data pooled across all biological replicates from K2 tissue. Distances are calculated between each IL-33$^+$ epithelial (E-cadherin$^+$) cell and its closest T_{reg} (CD3$^+$ FOXP3$^+$). Asterisks, significant difference (P < 0.0001, one-tailed, unpaired t test) compared with random distances calculated by permuting epithelial cells. (**E**) IL-33–centric cross-talk paths originating from epithelial gastric module E6 (central circle, magenta), with each outward concentric circle illustrating possible communication paths from inner to outer modules based on links inferred by Calligraphy. Arc length is proportional to the number of inner-module ligands that can bind to cognate receptors in the outer module. (**F**) tSNE as in (B), colored according to the step in which communication events from the IL-33–centric path in (E) reach the module expressed by that cell. Cells are assigned to their most highly expressed module, and each module is scored by the earliest step in which it appears along any paths through the Calligraphy network emanating from E6-derived IL-33. Cells expressing modules that are not downstream of IL-33 are colored in gray.

[3 weeks postinjury (wpi)], when PanIN lesions normally emerge (K3). As expected, IL-33 expression remains intact in nonepithelial pancreatic cells (*Il33*$^+$ *Vim*$^+$ or *Il33*$^+$ *aSMA*$^+$) (fig. S13, A to D) in sh*Il33* animals on dox, and is specifically abrogated in *Kras* mutant cells expressing the gastric markers TFF1 and ANXA10 (Fig. 6B and fig. S13, A and B).

Analysis of coembedded scRNA-seq data derived from control or KC-sh*Il33* on-dox mice show that IL-33 perturbation profoundly shifted the observed cell states within both epithelial and immune compartments. We applied the Milo algorithm (*60*), which characterizes such local shifts (as opposed to loss or gain of entire

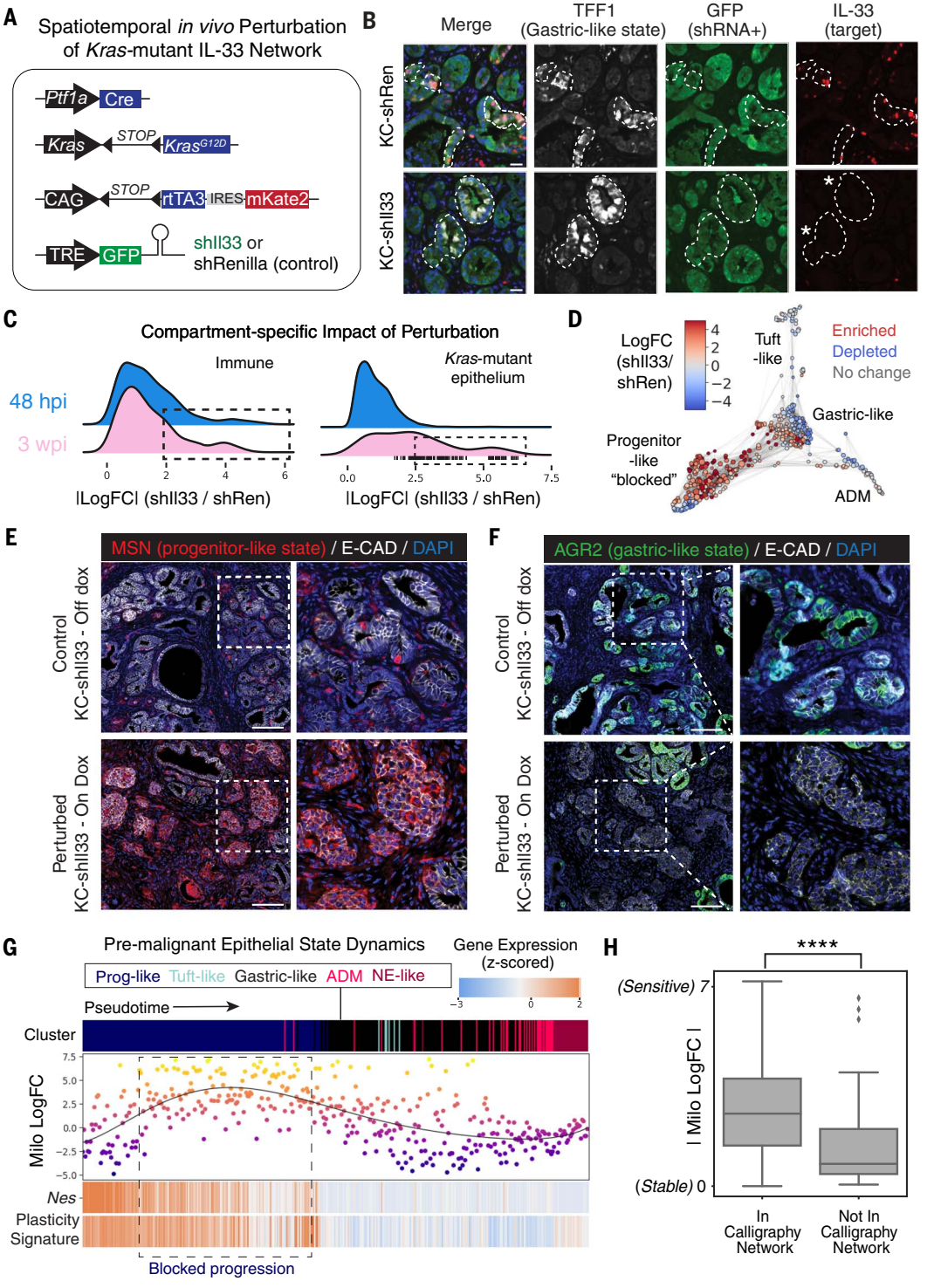


Fig. 6. Spatiotemporal in vivo perturbation of *Il33* impairs neoplastic progression. (**A**) Mouse models for inducible repression of *Il33* (KC-sh*Il33*, two independent strains) or Renilla control (KC-shRen), restricted to *Kras*-mutant epithelial cells by *Ptf1a*-Cre expression. (**B**) Representative IF of pancreata from control (top) or KC-sh*Il33* (bottom) mice placed on dox at 5 weeks of age and analyzed 9 days later at the 48 hpi time point (K2). *Kras*-mutant epithelial cells, not surrounding stroma, express the *Il33* shRNA marked by GFP in KC-sh*Il33*; TFF1 marks the cell state in which IL-33 is activated at 48 hpi in control but not sh*Il33* animals. Dashed lines demark the epithelium-stroma boundary; asterisks highlight suppression of *Il33* in TFF1^{+} metaplastic cells of KC-sh*Il33* mice. DAPI marks nuclei (blue). Scale bar, 20 μm. (**C**) Milo (*60*) log-fold change (logFC) magnitudes across cell neighborhoods (n = 5 mice), with higher values indicating a greater impact of IL-33 perturbation. Rightward distribution shifts (dotted lines) indicate a larger impact on particular cell states; vertical dashed lines indicate neighborhoods with significant (adjusted $P < 0.1$) shifts according to Milo, appearing only in K3 epithelia. (**D**) FDL of Milo neighborhoods colored by logFC of abundance in sh*Il33* samples relative to controls at the late (3 wpi) time point. (**E** and **F**) Representative IF in pancreata from KC-sh*Il33* mice placed on-dox (bottom) or off-dox (top) at 3 wpi (K3), showing aberrant accumulation of progenitor-like state (MSN^{+}) in epithelial cells (E-cadherin^{+}) of IL-33-perturbed animals at 48 hpi (E) and depletion of gastric-like (AGR2^{+}) states upon epithelial IL-33 suppression (F). DAPI marks nuclei (blue). Scale bar, 100 μm. (**G**) Impact of *Il33* perturbation across *Kras*-mutant epithelial neighborhoods at K3 (3 wpi). Top, pseudotime-ordered neighborhoods (columns) colored by cell state. Middle, neighborhoods plotted and colored by Milo logFC; higher logFC denotes greater abundance in sh*Il33* relative to control. Bottom, *Nes*- and plasticity-associated gene expression (Fig. 3F) (*27*); heatmap colors are scaled to ±2 SD from mean. (**H**) Milo logFC of neighborhoods mapped to modules that are (left) or are not (right) downstream of Calligraphy's IL-33–centric network; asterisks indicate significance ($P = 1.24 \times 10^{-7}$, unpaired, one-tailed t test).

clusters) by grouping similar cells into "neighborhoods" and identifying those that are differentially abundant between perturbed and control conditions. Consistent with an epithelial-to-immune cross-talk, we found that *Il33* suppression in the *Kras*-mutant epithelium results in rapid remodeling of the immune landscape, with multiple immune subpopulations shifting in abundance by 48 hpi (Fig. 6C). Epithelial remodeling is delayed by comparison; a lack of substantial changes at the early time point is followed by substantial remodeling of many epithelial cell states at 3 wpi (K3). By this time, the perturbation of IL-33–mediated cross-talk generates evidence of neoplastic

epithelial remodeling, with a shift to more cells in progenitor-like and fewer in gastric-like states (Fig. 6D and fig. S13, E and F), which was also seen in the IF data (Fig. 6, E and F).

To gain insights into the impact of IL-33 on the dynamics of cell-state transitions, we used Palantir (*61*) to infer a pseudotime ordering of epithelial neighborhoods at 3 wpi, beginning from the *Nes*$^{+}$ progenitor state (fig. S13E). Ordering Milo log-fold-change values along this pseudotime axis confirms that more IL-33–perturbed epithelial cells accumulate in earlier states expressing progenitor markers (*Nes*) and other genes associated with a plastic state (plasticity score correlation $P < 0.01$) (Fig. 6, D and G). Although *Il33* is expressed in only a small fraction of *Kras*-mutant epithelial cells, *Il33* perturbation results in marked changes in the cell-state composition of the premalignant pancreas, apparently by preventing the transition from a plastic progenitor-like state into distinct PanIN populations, such as the gastric-like cells that are normally abundant in unperturbed epithelia by 3 wpi.

The widespread changes in cell state caused by IL-33 perturbation support Calligraphy's prediction of the relevance of this feedback loop. To more directly link the specific predicted interacting partners with observed perturbation-induced changes, we mapped each Milo neighborhood to Calligraphy communication modules (fig. S13G) and evaluated the extent to which cell states predicted to be downstream of IL-33–mediated cross-talk overlap with cell states affected by the perturbation. Qualitatively, we found the largest impact of *Il33* perturbation on progenitor and bridge modules, both of which are predicted to be downstream of IL-33 and to express IL-4 receptor (*Il4ra*) (Fig. 5, B and E), whereas the only two modules not downstream of IL-33 are those with the smallest effect sizes (E2 and E5). Quantitatively, cell states predicted by Calligraphy to participate in the IL-33 network were more significantly affected by the *Il33* perturbation ($P = 1.24 \times 10^{-7}$, one-sided *t* test, t = –5.25) (Fig. 6H). These results functionally validate Calligraphy as an approach to infer both communication circuits and the specific subpopulations affected (directly and indirectly) upon perturbation of such networks.

Discussion

Although much is known about the molecular processes affecting tumor progression to advanced PDAC, pancreatic cancer is diagnosed late, and the paucity of molecular studies of early neoplasia has left us with little knowledge of how it emerges from a relatively homogeneous epithelium. By combining single-cell sequencing of mouse models with computational analysis, we found that permissive chromatin states in *Kras*-mutant cells diversify the communication programs available to preneoplastic tissue, expanding downstream cross-talk throughout the tumor microenvironment. Moreover, in the *Kras*-mutant context, epigenetic reprogramming and the emergence of cancer-driving populations is very dynamic, occurring within 2 days of insult by inflammation.

Mutation is known to drive plasticity in lung cancer through the loss of AT1 or AT2 lineage identity and acquisition of a phenotype intermediate between these states (*17*, *18*). In the pancreas, a similar loss of acinar identity and gain of an intermediate acinar-ductal state have long been observed in both tumorigenesis and regeneration; thus, traditional notions of plasticity are insufficient to describe its contribution to disease. We defined plasticity as the potential of a cell to manifest diverse future fates, motivating a generalizable plasticity score that tracks with the degree of epigenetic priming. This score nominated several highly plastic cell states, in which open chromatin unlocks access to multiple distinct gene programs observed in benign lesions or malignant disease and revealed that inflammation enhances plasticity across these states.

To better elucidate the emergence of plastic states, we sought to reconcile prior work proposing different cells of origin for neoplasia. Our GEMMs harbor mutant *Kras* in all acinar cells, allowing us to comprehensively explore which states can initiate tumorigenesis. Using CellRank (*39*), we traced the origins of epithelial transcriptional diversity to multiple "apex" progenitor populations that correspond with experimentally determined cells of origin. These populations also exhibit high plasticity scores and unify prior work by suggesting that neoplasia can arise from multiple *Kras*-mutant cell states through distinct responses to inflammation. Moreover, our *Ptf1a*-Cre model traces this diversity back to a predominantly acinar-like state, supported by the fact that nearly all premalignant epithelial cell states have an acinar-like chromatin state (epigenetic "memory"), which itself maps to a CellRank-predicted apex state. Whereas non-acinar lineage cells can also undergo neoplastic transformation in mice (*21*, *62*), our results agree with the observed loss of normal lineage identity upon *Kras* mutation and inflammation (*63*) and reveal apex states that may emerge after this transition. Among apex states, multiple unbiased analyses in particular support a *Nes*$^{+}$ progenitor–like state (*44*, *45*), which displays PDAC-associated chromatin alterations, expresses progenitor-associated genes, and scores highest for our plasticity metric—all hallmarks of highly plastic cells.

Our plasticity score was most correlated with cytokine and receptor genes, implying that plastic populations are primed to signal and/or respond to the environment. Addressing this, we investigated how cell-cell communication may drive rapid tissue remodeling. Prior communication inference approaches failed to find specific signals among the large number of cytokines and receptors expressed across cell populations. We therefore developed Calligraphy to leverage modularity in gene expression for greater power and robustness over testing individual R-L gene pairs, allowing us to focus on neoplasia-specific communication networks. Calligraphy identified modules of coexpressed communication genes that unexpectedly mapped one-to-one to transcriptional cell states, implying that communication is critical for establishing cell-state diversity within the premalignant pancreas. These networks were largely absent from normal pancreas, with only one being induced in a rare subpopulation of cells that emerges upon tissue damage. The same module has the highest propensity for tissue remodeling and persists in advanced murine and human cancers, demonstrating that cancer commandeers gene programs used during normal regeneration.

Our analyses revealed a feedback loop initiated by IL-33 signaling from *Kras*-mutant epithelial cells expressing the gastric module to T_H2 cytokine-expressing T_{regs} and ILC2s, which signal back to the epithelium (among multiple other routes). These findings link previous results on the relevance of T_H2 signaling in PDAC tumorigenesis (*56*) to those on the role IL-33 in this process (*23*). Spatial analysis revealed proximal localization of signaling populations in the loop, and epithelial *Il33* knockdown in a GEMM impaired inflammation-driven remodeling of plastic populations, blocking the emergence of gastric-like state cells that are otherwise abundant in PanIN lesions. This mechanism can be driven by solely epithelium-derived IL-33 despite the high stromal IL-33 expression previously implicated in disease phenotypes (*58*, *59*). Further, the results of in vivo *Il33* perturbation support Calligraphy inference by matching predictions of which populations are perturbed, to what degree, and in what temporal sequence. Other modules defined herein are also likely to have functional importance. Future work can extend this approach to other niche components such as fibroblasts or endothelial cells and should expose additional communication with the potential for therapeutic or diagnostic exploitation.

PDAC is frequently detected too late for curative intervention, and a detailed understanding of early neoplastic events may enable the development of rational strategies to prevent, detect, and intercept tumors before they progress to an intractable stage. Our results show that GEMMs can be used to study and perturb early events, revealing epigenetically plastic cell states in neoplasia that are not observed in the normal or regenerative pancreas. Further efforts to understand neoplasia-specific

communication networks driving PDAC initiation hold promise for the development of therapeutics that block early cancer progression, and may also be effective against advanced disease.

Methods summary

Experimental design

Samples from GEMMs were collected to span the entire range of PDAC progression (K1 to K6), as well as regenerating pancreata (N1 and N2). Additional samples were collected from GEMMs enabling selective genetic perturbation of premalignant *Kras*-mutant cells. Table S8 summarizes experimental conditions (*27*). All animal experiments were performed in accordance with the institutional animal care and use committee–approved protocol (11-06-018).

Generation of bulk and single-cell omics data

Tissue dissociation and cell preparations for bulk and single-cell ATAC-seq were performed as previously described (*23*). For scRNA-seq analysis, fluorescence-activated cell sorted (FACS) epithelial or immune cells were encapsulated and processed following 10x Genomics user manual (Reagent Kit 3′ v2) as described previously (*27*).

Spatial and immunophenotyping data

Tissues were processed and stained for imaging [IF/immunohistochemistry (IHC), IMC, hematoxylin & eosin (H&E), and smFISH] or FACS analyses (*27*). Tables S9 and S10 summarize panels used for multiplexed IMC and smFISH. IMC data were collected using the Hyperion Imaging System and CyTOF software version 7.0.8493.0 (Fluidigm). smFISH imaging was performed on a Nikon Ti2 inverted microscope. FACS data were acquired using a five-laser BD Biosciences LSRFortessa flow cytometer and analyzed using FlowJo version 10.0 software.

Computational analysis

scRNA-seq data were processed with SEQC (*64*), filtered with a custom pipeline (*27*), and log library size normalized. scATAC-seq data were processed with ArchR (*65*). Processed transcriptomic and epigenomic datasets were analyzed with custom Python scripts for visualization, cell-state annotation, metacell inference, multimodal integration, plasticity scoring, and Calligraphy communication inference, among other analyses fully described in (*27*). smFISH image analysis was performed on maximum projection images with segmentation on the DAPI channel using Mesmer (*66*) and Python code for phenotyping and spatial analyses.

REFERENCES AND NOTES

1. B. Vogelstein *et al*., Cancer genome landscapes. *Science* **339**, 1546–1558 (2013). doi: 10.1126/science.1235122; pmid: 23539594
2. I. Martincorena *et al*., High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880–886 (2015). doi: 10.1126/science.aaa6806; pmid: 25999502
3. N. Wijewardhane, L. Dressler, F. D. Ciccarelli, Normal somatic mutations in cancer transformation. *Cancer Cell* **39**, 125–129 (2021). doi: 10.1016/j.ccell.2020.11.002; pmid: 33220180
4. D. Hanahan, Hallmarks of cancer: New dimensions. *Cancer Discov.* **12**, 31–46 (2022). doi: 10.1158/2159-8290.CD-21-1059; pmid: 35022204
5. A. S. Nam, R. Chaligne, D. A. Landau, Integrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics. *Nat. Rev. Genet.* **22**, 3–18 (2021). doi: 10.1038/s41576-020-0265-5; pmid: 32807900
6. C. Guerra *et al*., Chronic pancreatitis is essential for induction of pancreatic ductal adenocarcinoma by K-Ras oncogenes in adult mice. *Cancer Cell* **11**, 291–302 (2007). doi: 10.1016/j.ccr.2007.01.012; pmid: 17349585
7. C. Carrière, A. L. Young, J. R. Gunn, D. S. Longnecker, M. Korc, Acute pancreatitis markedly accelerates pancreatic cancer progression in mice expressing oncogenic Kras. *Biochem. Biophys. Res. Commun.* **382**, 561–565 (2009). doi: 10.1016/j.bbrc.2009.03.068; pmid: 19292977
8. A. B. Lowenfels *et al*., Hereditary pancreatitis and the risk of pancreatic cancer. *J. Natl. Cancer Inst.* **89**, 442–446 (1997). doi: 10.1093/jnci/89.6.442; pmid: 9091646
9. L. M. Coussens, Z. Werb, Inflammation and cancer. *Nature* **420**, 860–867 (2002). doi: 10.1038/nature01322; pmid: 12490959
10. V. Giroux, A. K. Rustgi, Metaplasia: Tissue injury adaptation and a precursor to the dysplasia-cancer sequence. *Nat. Rev. Cancer* **17**, 594–604 (2017). doi: 10.1038/nrc.2017.68; pmid: 28860646
11. R. Maddipati, B. Z. Stanger, Pancreatic cancer metastases harbor evidence of polyclonality. *Cancer Discov.* **5**, 1086–1097 (2015). doi: 10.1158/2159-8290.CD-15-0120; pmid: 26209539
12. S. R. Torborg, Z. Li, J. E. Chan, T. Tammela, Cellular and molecular mechanisms of plasticity in cancer. *Trends Cancer* **8**, 735–746 (2022). doi: 10.1016/j.trecan.2022.04.007; pmid: 35618573
13. W. A. Flavahan, E. Gaskell, B. E. Bernstein, Epigenetic plasticity and the hallmarks of cancer. *Science* **357**, eaal2380 (2017). doi: 10.1126/science.aal2380; pmid: 28729483
14. M. A. Dawson, The cancer epigenome: Concepts, challenges, and therapeutic opportunities. *Science* **355**, 1147–1152 (2017). doi: 10.1126/science.aam7304; pmid: 28302822
15. W. Xie *et al*., Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134–1148 (2013). doi: 10.1016/j.cell.2013.04.022; pmid: 23664764
16. C. A. Gifford *et al*., Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell* **153**, 1149–1163 (2013). doi: 10.1016/j.cell.2013.04.037; pmid: 23664763
17. L. M. LaFave *et al*., Epigenomic state transitions characterize tumor progression in mouse lung adenocarcinoma. *Cancer Cell* **38**, 212–228.e13 (2020). doi: 10.1016/j.ccell.2020.06.006; pmid: 32707078
18. N. D. Marjanovic *et al*., Emergence of a high-plasticity cell state during lung cancer evolution. *Cancer Cell* **38**, 229–246.e13 (2020). doi: 10.1016/j.ccell.2020.06.012; pmid: 32707077
19. P. Storz, Acinar cell plasticity and development of pancreatic ductal adenocarcinoma. *Nat. Rev. Gastroenterol. Hepatol.* **14**, 296–304 (2017). doi: 10.1038/nrgastro.2017.12; pmid: 28270694
20. C. Guerra *et al*., Pancreatitis-induced inflammation contributes to pancreatic cancer by inhibiting oncogene-induced senescence. *Cancer Cell* **19**, 728–739 (2011). doi: 10.1016/j.ccr.2011.05.011; pmid: 21665147
21. S. Y. G. Friedlander *et al*., Context-dependent transformation of adult pancreatic cells by oncogenic K-Ras. *Cancer Cell* **16**, 379–389 (2009). doi: 10.1016/j.ccr.2009.09.027; pmid: 19878870
22. J. P. Morris IV, D. A. Cano, S. Sekine, S. C. Wang, M. Hebrok, β-catenin blocks Kras-dependent reprogramming of acini into pancreatic cancer precursor lesions in mice. *J. Clin. Invest.* **120**, 508–520 (2010). doi: 10.1172/JCI40045; pmid: 20071774
23. D. Alonso-Curbelo *et al*., A gene-environment-induced epigenetic program initiates tumorigenesis. *Nature* **590**, 642–648 (2021). doi: 10.1038/s41586-020-03147-x; pmid: 33536616
24. E. Del Poggetto *et al*., Epithelial memory of inflammation limits tissue damage while promoting pancreatic tumorigenesis. *Science* **373**, eabj0486 (2021). doi: 10.1126/science.abj0486; pmid: 34529467
25. Y. Li *et al*., Mutant Kras co-opts a proto-oncogenic enhancer network in inflammation-induced metaplastic progenitor cells to initiate pancreatic cancer. *Nat. Cancer* **2**, 49–65 (2021). doi: 10.1038/s43018-020-00134-z; pmid: 35121887
26. Y. Kawaguchi *et al*., The role of the transcriptional regulator Ptf1a in converting intestinal to pancreatic progenitors. *Nat. Genet.* **32**, 128–134 (2002). doi: 10.1038/ng959; pmid: 12185368
27. Materials and methods are available as supplementary materials.
28. C. B. Westphalen *et al*., Dclk1 defines quiescent pancreatic progenitors that promote injury-induced regeneration and tumorigenesis. *Cell Stem Cell* **18**, 441–455 (2016). doi: 10.1016/j.stem.2016.03.016; pmid: 27058937
29. S. Sinha *et al*., PanIN neuroendocrine cells promote tumorigenesis via neuronal cross-talk. *Cancer Res.* **77**, 1868–1879 (2017). doi: 10.1158/0008-5472.CAN-16-0899; pmid: 28386018
30. A. D. Rhim *et al*., EMT and dissemination precede pancreatic tumor formation. *Cell* **148**, 349–361 (2012). doi: 10.1016/j.cell.2011.11.025; pmid: 22265420
31. J. L. Kopp *et al*., Identification of Sox9-dependent acinar-to-ductal reprogramming as the principal mechanism for initiation of pancreatic ductal adenocarcinoma. *Cancer Cell* **22**, 737–750 (2012). doi: 10.1016/j.ccr.2012.10.025; pmid: 23201164
32. J. Peng *et al*., Author Correction: Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Res.* **29**, 777 (2019). doi: 10.1038/s41422-019-0212-1; pmid: 31409908
33. R. R. Coifman, S. Lafon, Diffusion maps. *Appl. Comput. Harmon. Anal.* **21**, 5–30 (2006). doi: 10.1016/j.acha.2006.04.006
34. S. Yang *et al*., A novel MIF signaling pathway drives the malignant character of pancreatic cancer by targeting NR3C2. *Cancer Res.* **76**, 3838–3850 (2016). doi: 10.1158/0008-5472.CAN-15-2841; pmid: 27197190
35. R. A. Moffitt *et al*., Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat. Genet.* **47**, 1168–1178 (2015). doi: 10.1038/ng.3398; pmid: 26343385
36. D. L. Gibbons, C. J. Creighton, Pan-cancer survey of epithelial-mesenchymal transition markers across the Cancer Genome Atlas. *Dev. Dyn.* **247**, 555–564 (2018). doi: 10.1002/dvdy.24485; pmid: 28073171
37. R. Jahan *et al*., Trefoil factor(s) and CA19.9: A promising panel for early detection of pancreatic cancer. *EBioMedicine* **42**, 375–385 (2019). doi: 10.1016/j.ebiom.2019.03.056; pmid: 30956167
38. J.-S. Roe *et al*., Enhancer reprogramming promotes pancreatic cancer metastasis. *Cell* **170**, 875–888.e20 (2017). doi: 10.1016/j.cell.2017.07.007; pmid: 28757253
39. M. Lange *et al*., CellRank for directed single-cell fate mapping. *Nat. Methods* **19**, 159–170 (2022). doi: 10.1038/s41592-021-01346-6; pmid: 35027767
40. V. Bergen, M. Lange, S. Peidli, F. A. Wolf, F. J. Theis, Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020). doi: 10.1038/s41587-020-0591-3; pmid: 32747759
41. G. La Manno *et al*., RNA velocity of single cells. *Nature* **560**, 494–498 (2018). doi: 10.1038/s41586-018-0414-6; pmid: 30089906
42. M. Ioannou *et al*., ALDH1B1 is a potential stem/progenitor marker for multiple pancreas progenitor pools. *Dev. Biol.* **374**, 153–163 (2013). doi: 10.1016/j.ydbio.2012.10.030; pmid: 23142317
43. E. Mameishvili *et al*., Aldh1b1 expression defines progenitor cells in the adult pancreas and is required for Kras-induced pancreatic cancer. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 20679–20688 (2019). doi: 10.1073/pnas.1901075116 pmid: 31548432
44. C. Carrière, E. S. Seeley, T. Goetze, D. S. Longnecker, M. Korc, The Nestin progenitor lineage is the compartment of origin for pancreatic intraepithelial neoplasia. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 4437–4442 (2007). doi: 10.1073/pnas.0701117104; pmid: 17360542
45. C. Carrière, A. L. Young, J. R. Gunn, D. S. Longnecker, M. Korc, Acute pancreatitis accelerates initiation and progression to pancreatic cancer in mice expressing oncogenic Kras in the nestin cell lineage. *PLOS ONE* **6**, e27725 (2011). doi: 10.1371/journal.pone.0027725; pmid: 22140463
46. Y. Baran *et al*., MetaCell: Analysis of single-cell RNA-seq data using K-nn graph partitions. *Genome Biol.* **20**, 206 (2019). doi: 10.1186/s13059-019-1812-2; pmid: 31604482
47. S. Persad *et al*., SEACells: Inference of transcriptional and epigenomic cellular states from single-cell genomics data. bioRxiv 2022.04.02.486748 [Preprint] (2022); https://doi.org/10.1101/2022.04.02.486748.
48. A. Subramanian *et al*., Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression

profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005). doi: 10.1073/pnas.0506580102; pmid: 16199517
49. Y. Pylayeva-Gupta, K. E. Lee, C. H. Hajdu, G. Miller, D. Bar-Sagi, Oncogenic Kras-induced GM-CSF production promotes the development of pancreatic neoplasia. *Cancer Cell* **21**, 836–847 (2012). doi: 10.1016/j.ccr.2012.04.024; pmid: 22698407
50. M. Efremova, M. Vento-Tormo, S. A. Teichmann, R. Vento-Tormo, CellPhoneDB: Inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020). doi: 10.1038/s41596-020-0292-x; pmid: 32103204
51. O. Strobel *et al*., In vivo lineage tracing defines the role of acinar-to-ductal transdifferentiation in inflammatory ductal metaplasia. *Gastroenterology* **133**, 1999–2009 (2007). doi: 10.1053/j.gastro.2007.09.009; pmid: 18054571
52. Y. Schlesinger *et al*., Single-cell transcriptomes of pancreatic preinvasive lesions and cancer reveal acinar metaplastic cells' heterogeneity. *Nat. Commun.* **11**, 4516 (2020). doi: 10.1038/s41467-020-18207-z; pmid: 32908137
53. H. Gonzalez, C. Hagerling, Z. Werb, Roles of the immune system in cancer: From tumor initiation to metastatic progression. *Genes Dev.* **32**, 1267–1284 (2018). doi: 10.1101/gad.314617.118; pmid: 30275043
54. U. Alon, Network motifs: Theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450–461 (2007). doi: 10.1038/nrg2102; pmid: 17510665
55. S. Das, B. Shapiro, E. A. Vucic, S. Vogt, D. Bar-Sagi, Tumor cell-derived IL1β promotes desmoplasia and immune suppression in pancreatic cancer. *Cancer Res.* **80**, 1088–1101 (2020). doi: 10.1158/0008-5472.CAN-19-2080; pmid: 31915130
56. A. Alam *et al*., Fungal mycobiome drives IL-33 secretion and type 2 immunity in pancreatic cancer. *Cancer Cell* **40**, 153–167.e11 (2022). doi: 10.1016/j.ccell.2022.01.003; pmid: 35120601
57. L. Y. Drake, H. Kita, IL-33: Biological properties, functions, and roles in airway disease. *Immunol. Rev.* **278**, 173–184 (2017). doi: 10.1111/imr.12552; pmid: 28658560
58. P. Andersson *et al*., Molecular mechanisms of IL-33-mediated stromal interactions in cancer metastasis. *JCI Insight* **3**, e122375 (2018). doi: 10.1172/jci.insight.122375; pmid: 30333314
59. A. Velez-Delgado *et al*., Extrinsic KRAS signaling shapes the pancreatic microenvironment through fibroblast reprogramming. *Cell. Mol. Gastroenterol. Hepatol.* **13**, 1673–1699 (2022). doi: 10.1016/j.jcmgh.2022.02.016; pmid: 35245687
60. E. Dann, N. C. Henderson, S. A. Teichmann, M. D. Morgan, J. C. Marioni, Differential abundance testing on single-cell data using k-nearest neighbor graphs. *Nat. Biotechnol.* **40**, 245–253 (2022). doi: 10.1038/s41587-021-01033-z; pmid: 34594043
61. M. Setty *et al*., Characterization of cell fate probabilities in single-cell data with Palantir. *Nat. Biotechnol.* **37**, 451–460 (2019). doi: 10.1038/s41587-019-0068-4; pmid: 30899105
62. J. M. Bailey *et al*., p53 mutations cooperate with oncogenic Kras to promote adenocarcinoma from pancreatic ductal cells. *Oncogene* **35**, 4282–4288 (2016). doi: 10.1038/onc.2015.441; pmid: 26592447
63. A. Malinova, L. Veghini, F. X. Real, V. Corbo, Cell lineage infidelity in PDAC progression and therapy resistance. *Front. Cell Dev. Biol.* **9**, 795251 (2021). doi: 10.3389/fcell.2021.795251; pmid: 34926472
64. E. Azizi *et al*., Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell* **174**, 1293–1308.e36 (2018). doi: 10.1016/j.cell.2018.05.060; pmid: 29961579
65. J. M. Granja *et al*., ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021). doi: 10.1038/s41588-021-00790-6; pmid: 33633365
66. N. F. Greenwald *et al*., Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nat. Biotechnol.* **40**, 555–565 (2022). doi: 10.1038/s41587-021-01094-0; pmid: 34795433
67. N. M. Krah *et al*., The acinar differentiation determinant PTF1A inhibits initiation of pancreatic ductal adenocarcinoma. *eLife* **4**, e07125 (2015). doi: 10.7554/eLife.07125; pmid: 26151762
68. C. F. Schaefer *et al*., PID: The Pathway Interaction Database. *Nucleic Acids Res.* **37**, D674–D679 (2009). doi: 10.1093/nar/gkn653; pmid: 18832364
69. D. van Dijk *et al*., Recovering gene interactions from single-cell data using data diffusion. *Cell* **174**, 716–729.e27 (2018). doi: 10.1016/j.cell.2018.05.061; pmid: 29961576
70. C. Burdziak, Code for: Epigenetic plasticity cooperates with cell-cell interactions to direct pancreatic tumorigenesis, Zenodo (2023); https://doi.org/10.5281/zenodo.7874129.

ACKNOWLEDGMENTS

We thank J. Simon and the MSKCC animal facility for technical support with animal colonies; MSKCC Flow Cytometry facility for support with FACS sorting experiments; the Single Cell and Imaging Mass Cytometry Platform at Goodman Cancer Research Centre; the members of the Sloan Kettering Institute's Single Cell Analytics and Innovation Lab (SAIL) computational unit, and members of the Lowe and Pe'er laboratories for advice and discussions, in particular M. Setty and A. Chaikovsky for foundational scRNA-seq analyses and critically revising the final manuscript, respectively. We also thank J. Moffitt, B. Watson, J. Hurley, and S. Aviles for generously sharing their knowledge, protocols, and guidance to help us set up the multiplex smFISH platform in the lab and S. Gan and D. V. Kumar Yarlagadda for pivotal contributions to establishing our infrastructure for multiplex smFISH imaging. **Funding:** This work was supported by Alan and Sandra Gerry Metastasis and Tumor Ecosystems Center (GMTEC) funding and National Cancer Institute (NCI) Cancer Center Support grant P30 CA008748. C.B. is supported by a Ruth L. Kirschtein Predoctoral Fellowship (National Cancer Institute grant F31CA24690). D.A.C. is supported by La Caixa Junior Leader Fellowship LCF/BQ/PI20/11760006 and FERO-ASEICA grant BFERO2021 from the FERO Foundation and Spanish Cancer Research Association and Spanish Ministry of Science and Innovation grant PID2021-128102OA-I00. F.M.B. is supported by an Edward P. Evans Young Investigator Award. T.W. is supported by a fellowship from the DKFZ Clinician Scientist Program, supported by the Dieter Morszeck Foundation. SCIMAP is supported by the Fraser Memorial Trust and a McGill MI4 Platform grant. J.R. is a Howard Hughes Medical Institute Fellow of the Damon Runyon Cancer Research Foundation (grant DRG-2382-19). C.J.Z. is supported by National Institutes of Health grant R25 CA233208. S.W.L. is an investigator in the Howard Hughes Medical Institute and the Geoffrey Beene Chair for Cancer Biology. D.P. is an investigator in the Howard Hughes Medical Institute, is an Alan and Sandra Gerry Endowed Chair, and is supported by NCI grant U54 CA209975, National Institute of Child Health and Human Development DP1 grant HD084071, and the Starr Cancer Consortium. **Author contributions:** Conceptualization: C.B., D.A.-C., S.W.L., D.P.; Data curation: C.B., D.A.-C., T.W., J.R., D.H., Y.X., V.G., Y.-J.H.; Formal analysis: C.B., T.W., D.H., Y.X., C.J.Z., R.K.; Funding acquisition: C.B., D.A.-C., S.W.L., D.P.; Investigation: C.B., D.A.-C., T.W., J.R., F.M.B., D.H., Y.X., Z.Z., H.-A.C., O.C., I.M., W.L., A.W., L.M., R.C.; Methodology: C.B., D.A.-C., S.W.L., D.P.; Resources: Z.-N.C., Y.W., D.Q.; Software: C.B.; Supervision: S.W.L., D.P.; Visualization: C.B., D.A.-C., T.W.; Writing – original draft: C.B., D.A.-C., T.N., S.W.L., D.P.; Writing – review & editing: T.W., J.R. **Competing interests:** S.W.L. is a consultant and holds equity in Blueprint Medicines, ORIC Pharmaceuticals, Mirimus Inc., PMV Pharmaceuticals, Faeth Therapeutics, and Constellation Pharmaceuticals. D.A.-C. and S.W.L. are listed as the inventors for patent application (PTC/US2019/041670, internationally filing date 12 July 2019) covering methods for preventing or treating KRAS mutant pancreas cancer with inhibitors of type 2 cytokine signaling. D.P. is on the scientific advisory board of Insitro. T.W. reports stock ownership for Roche, Bayer, Innate Pharma, Illumina, and 10x Genomics, as well as research funding (not related to this study) from CanVirex AG, Basel Switzerland, and Institut für Klinische Krebsforschung GmbH, Frankfurt, Germany. C.B., D.A.-C., S.W.L., and D.P. are listed as inventors on a provisional patent application (63/390,075) related to methods and compositions for treating PDAC, for which Memorial Sloan Kettering Cancer Center is the applicant. **Data and materials availability:** All sequencing data have been deposited at the Gene Expression Omnibus (GEO) under accession number GSE207943. An interactive data browser to plot gene expression trends on tSNE or FDL visualizations of scRNA-seq data is accessible at http://pdac-progression-browser.us-east-1.elasticbeanstalk.com. Code for data analysis is available at github (https://github.com/dpeerlab/pdac-progression) and Zenodo (*70*). KC-shIL33 ESCs for the production of EPO-GEMMs are available from S.W.L. upon request. **License information:** Copyright © 2023 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.add5327
Materials and Methods
Figs. S1 to S13
Tables S1 to S18
References (*71*–*101*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 17 June 2022; accepted 31 March 2023
10.1126/science.add5327

RESEARCH ARTICLE SUMMARY

HUMAN GENETICS

Demographic history and genetic structure in pre-Hispanic Central Mexico

Viridiana Villa-Islas, Alan Izarraras-Gomez, Maximilian Larena, Elizabeth Mejía Perez Campos, Marcela Sandoval-Velasco, Juan Esteban Rodríguez-Rodríguez, Miriam Bravo-Lopez, Barbara Moguel, Rosa Fregel, Ernesto Garfias-Morales, Jazeps Medina Tretmanis, David Alberto Velázquez-Ramírez, Alberto Herrera-Muñóz, Karla Sandoval, Maria A. Nieves-Colón, Gabriela Zepeda García Moreno, Fernando A. Villanea, Eugenia Fernández Villanueva Medina, Ramiro Aguayo-Haro, Cristina Valdiosera, Alexander G. Ioannidis, Andrés Moreno-Estrada, Flora Jay, Emilia Huerta-Sanchez, J. Víctor Moreno-Mayar, Federico Sánchez-Quinto, María C. Ávila-Arcos*

INTRODUCTION: Between 1100 and 900 years ago, a global warming episode affected numerous civilizations worldwide. In the Americas, severe droughts reconfigured the demography of pre-Hispanic civilizations as well as the ecological landscape. We report ancient genomic data from 12 pre-Hispanic individuals from before and after this climate change episode to investigate the population dynamics at the limit between the two biocultural regions of Aridoamerica in the north and Mesoamerica in central and south Mexico.

RATIONALE: The archaeological evidence indicates that the droughts shifted the border southward between desertic Aridoamerica and the verdant and culturally rich Mesoamerica, which was home to large civilizations like the Aztecs and Mayans. This climate change allegedly led to a population replacement in the northern frontier of Mesoamerica by seminomadic hunter-gatherers from Aridoamerica. However, this hypothesis relies solely on archaeological data, so we generated ancient genomic and mitochondrial DNA data to test it. In addition, we leveraged these data to describe the ancient population structure of Mexico and to investigate the contribution of unsampled genetic lineages to the ancient genomes.

RESULTS: The pre-Hispanic population structure closely resembles the structure observed in present-day Mexico, which clearly differentiates northern and central Indigenous populations. This reflects an overall conservation of the genetic structure of the populations inhabiting the Mexican territory for at least 1400 years (which is the date of the most ancient individual in our study).

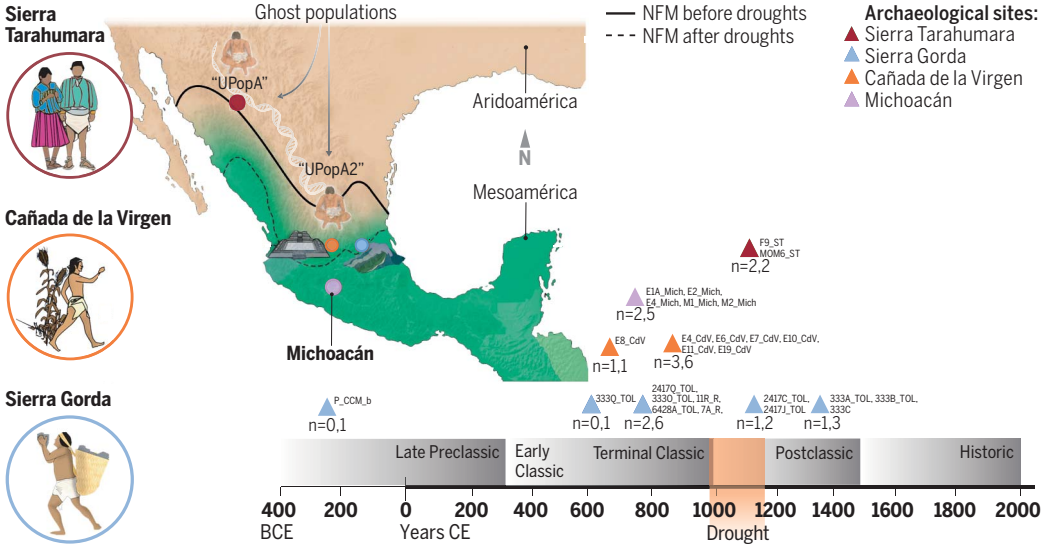
We found genetic continuity in the ancient individuals from before and after the climate change episode. This contradicts the hypothesis of population replacement by Aridoamerican groups in this region and suggests that the local population stayed in their homeland despite the long-standing droughts. The population continuity at the studied site can be explained by the favorable location in the l and humid Sierra Gorda mountain range the fact that cinnabar mining was the main economic activity, not agriculture.

We identified the contribution of two distinct unsampled "ghost" genetic ancestries to pre-Hispanic populations of northern and central Mexico, respectively. Whereas the unsampled genetic ancestry contributing to the northern genome coincides with one previously identified in a present-day population from south Mexico, the second ghost genetic ancestry was previously unknown.

CONCLUSION: The ancient genomes revealed a conservation of the genetic structure in Mexico in the past 1400 years and population continuity in the northern frontier of Mesoamerica despite the severe droughts 1100 years ago. It is likely that the mining-based economy allowed the population to subsist in their homeland during this climate change period when the border between Aridoamerica and Mesoamerica shifted southward. The identification of a new ghost genetic ancestry (UpopA2), along with that observed in ancient Sierra Tarahumara and present-day Mixe (UPopA), reveals a complex population history in the late Pleistocene in the Americas. The recovery and study of ancient genomes from Mexico, conducted ethically, can help fill important gaps in our understanding of the deep population history of the Americas. ■

The list of author affiliations is available in the full article online.
***Corresponding author. Email: mavila@liigh.unam.mx**
Cite this article as V. Villa-Islas *et al.*, *Science* 380, eadd6142 (2023). DOI: 10.1126/science.add6142

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.add6142



Population continuity and ghost genetic ancestries in pre-Hispanic Mexico. Ancient individuals who inhabited the northern frontier of Mesoamerica (NFM) before and after a 200-year period of severe droughts (shown in timeline) belong to a continuous population. This is in contrast to a previous hypothesis suggesting that hunter-gatherers from Aridoamerica replaced the populations at the NFM following the southward shift of its limits with Mesoamerica (solid and dashed lines). Individuals from Sierra Tarahumara and Cañada de la Virgen show genetic ancestry from two distinct ghost populations (UpopA and UpopA2).

RESEARCH ARTICLE

HUMAN GENETICS

Demographic history and genetic structure in pre-Hispanic Central Mexico

Viridiana Villa-Islas[1], Alan Izarraras-Gomez[1], Maximilian Larena[2], Elizabeth Mejía Perez Campos[3], Marcela Sandoval-Velasco[4,5,6], Juan Esteban Rodríguez-Rodríguez[1], Miriam Bravo-Lopez[1], Barbara Moguel[1,7], Rosa Fregel[8], Ernesto Garfias-Morales[1], Jazeps Medina Tretmanis[9], David Alberto Velázquez-Ramírez[10], Alberto Herrera-Muñóz[3], Karla Sandoval[11], Maria A. Nieves-Colón[12,13], Gabriela Zepeda García Moreno[14], Fernando A. Villanea[15], Eugenia Fernández Villanueva Medina[16], Ramiro Aguayo-Haro[16], Cristina Valdiosera[17,18], Alexander G. Ioannidis[19], Andrés Moreno-Estrada[12], Flora Jay[20], Emilia Huerta-Sanchez[9], J. Víctor Moreno-Mayar[21], Federico Sánchez-Quinto[22], María C. Ávila-Arcos[1]*

Aridoamerica and Mesoamerica are two distinct cultural areas in northern and central Mexico, respectively, that hosted numerous pre-Hispanic civilizations between 2500 BCE and 1521 CE. The division between these regions shifted southward because of severe droughts ~1100 years ago, which allegedly drove a population replacement in central Mexico by Aridoamerican peoples. In this study, we present shotgun genome-wide data from 12 individuals and 27 mitochondrial genomes from eight pre-Hispanic archaeological sites across Mexico, including two at the shifting border of Aridoamerica and Mesoamerica. We find population continuity that spans the climate change episode and a broad preservation of the genetic structure across present-day Mexico for the past 2300 years. Lastly, we identify a contribution to pre-Hispanic populations of northern and central Mexico from two ancient unsampled "ghost" populations.

Before European colonization, the territory that is present-day Mexico was home to numerous civilizations that occupied two main cultural areas: Aridoamerica in northern Mexico, inhabited mainly by hunter-gatherers, and Mesoamerica in central and southern Mexico (Fig. 1), where some of the largest pre-Hispanic agriculture-based civilizations in the Americas flourished between 2500 BCE and 1521 CE (*1*, *2*). The distinction between Aridoamerica and Mesoamerica is based on the cultural characteristics and subsistence strategies of the peoples that inhabited them, as well as the ecological features of each region (*3*, *4*). Archaeological evidence indicates that the border between these two areas shifted southward between 900–1300 CE after multidecadal droughts (*5*). This period is also known as the Medieval Warm Period in other regions of the world (*6*). The droughts allegedly led to population replacements in the northern frontier of Mesoamerica by semino-madic hunter-gatherers ("Chichimecas") from Aridoamerica (*7*) and precipitated the fall of some pre-Hispanic societies and the abandonment of Mesoamerican cities in central and southeast Mexico (*5*, *8*–*10*).

Evidence for the population replacement at the northern frontier of Mesoamerica comes solely from the archaeological record. Whether this change was the product of migration or acculturation has been debated by archaeologists for years (*8*, *11*–*13*). Studying the genetic variation of ancient populations that span this period of climate change across these two regions is thus necessary to illuminate the regional population dynamics in response to this drastic environmental change. However, ancient genomic data for pre-Hispanic populations from Mexico is very limited, with only two studies reporting 11 low-depth genomes (<0.3×) for a few individuals restricted to northern Mexico (*14*, *15*), and no available genomes from central and southern Mexico.

In this study, we report the most extensive set of complete mitochondrial genomes (mitogenomes) and shotgun genome-wide data to date from pre-Hispanic individuals from Mexico. We analyze our data jointly with publicly available datasets of other ancient Native Americans and present-day Indigenous populations from Mexico. This large compendium of ancient genomic data allows us to study the pre-Hispanic genetic structure in the territory occupied by Mexico, helping to answer long-standing questions regarding population dynamics at the northern frontier of Mesoamerica, disentangling the complex structure of central Mexico populations, and revealing a previously unknown ancient contribution of unsampled genetic lineages to Mexican populations.

Results

Sampling and pre-Hispanic genomic data generation

We screened archaeological samples from 37 pre-Hispanic individuals excavated at seven sites from within Mesoamerica in central Mexico. Three are in the Sierra Gorda in Querétaro state (Toluquilla and Ranas sites and a cave in Cadereyta de Montes), three in Michoacán state (Zaragoza, Tanhuato, and La Mina sites), and one in Guanajuato state (Cañada de la Virgen site) (Fig. 1). Furthermore, we produced additional shotgun data for two mummies from the Sierra Tarahumara from Aridoamerica, northern Mexico, who were sequenced at lower depths in a previous study (*14*) (Fig. 1). The sampling of the archeological human remains was made after approval by the Archaeology Council of the Instituto Nacional de Antropología e Historia with permit numbers 401.3S.16-2017/990 and 401.3S.16-2019/222. We sampled the minimum possible amount of tissue to avoid unnecessary destruction. For teeth, we tried to separate the root without damaging the crown and returned all leftover material to the archaeologists responsible for each collection. Although consultation with Indigenous populations for destructive analysis of human archaeological material is not a requirement by law in Mexico and the studied sites are not located within Indigenous communities, we acknowledge the need to involve Indigenous perspectives in delineating the

[1]International Laboratory for Human Genome Research, Universidad Nacional Autónoma de México (UNAM), Querétaro, México. [2]Department of Organismal Biology, Uppsala University, Uppsala, Sweden. [3]National Institute of Anthropology and History, Querétaro, Mexico. [4]Section for Evolutionary Genomics, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark. [5]Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560, USA. [6]Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Cuidad de México, Mexico. [7]Centro de Geociencias, UNAM Juriquilla, Juriquilla, Querétaro, México. [8]Department of Biochemistry, Microbiology, Cell Biology and Genetics, Universidad de La Laguna, San Cristóbal de La Laguna, Spain. [9]Center for Computational Molecular Biology, Brown University, Providence, RI, USA. [10]Institute for Medical Microbiology and Hospital Epidemiology, Hannover Medical School, 30625 Hannover, Germany. [11]Equity and Gender Office of the Centre for Research and Advanced Studies (CODIGO-C), CINVESTAV, Mexico City, Mexico. [12]Unit of Advanced Genomics, National Laboratory of Genomics for Biodiversity (LANGEBIO), CINVESTAV, Irapuato, Guanajuato, Mexico. [13]Department of Anthropology, University of Minnesota Twin Cities, Minneapolis, MN, USA. [14]National Institute of Anthropology and History, Guanajuato, Mexico. [15]Department of Anthropology, University of Colorado Boulder, Boulder, CO, USA. [16]National Institute of Anthropology and History, Michoacán, Mexico. [17]Departamento de Historia, Geografía y Comunicaciones, Universidad de Burgos, Burgos, Spain. [18]Department of History and Archaeology, La Trobe University, Melbourne, Australia. [19]Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA, USA. [20]Laboratoire Interdisciplinaire des Sciences du Numérique, Université Paris-Saclay, CNRS, INRIA, 91400 Orsay, France. [21]Lundbeck Foundation GeoGenetics Centre, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark. [22]Computational Genomics, Instituto Nacional de Medicina Genómica, Ciudad de México, Mexico.
*Corresponding author. Email: mavila@liigh.unam.mx

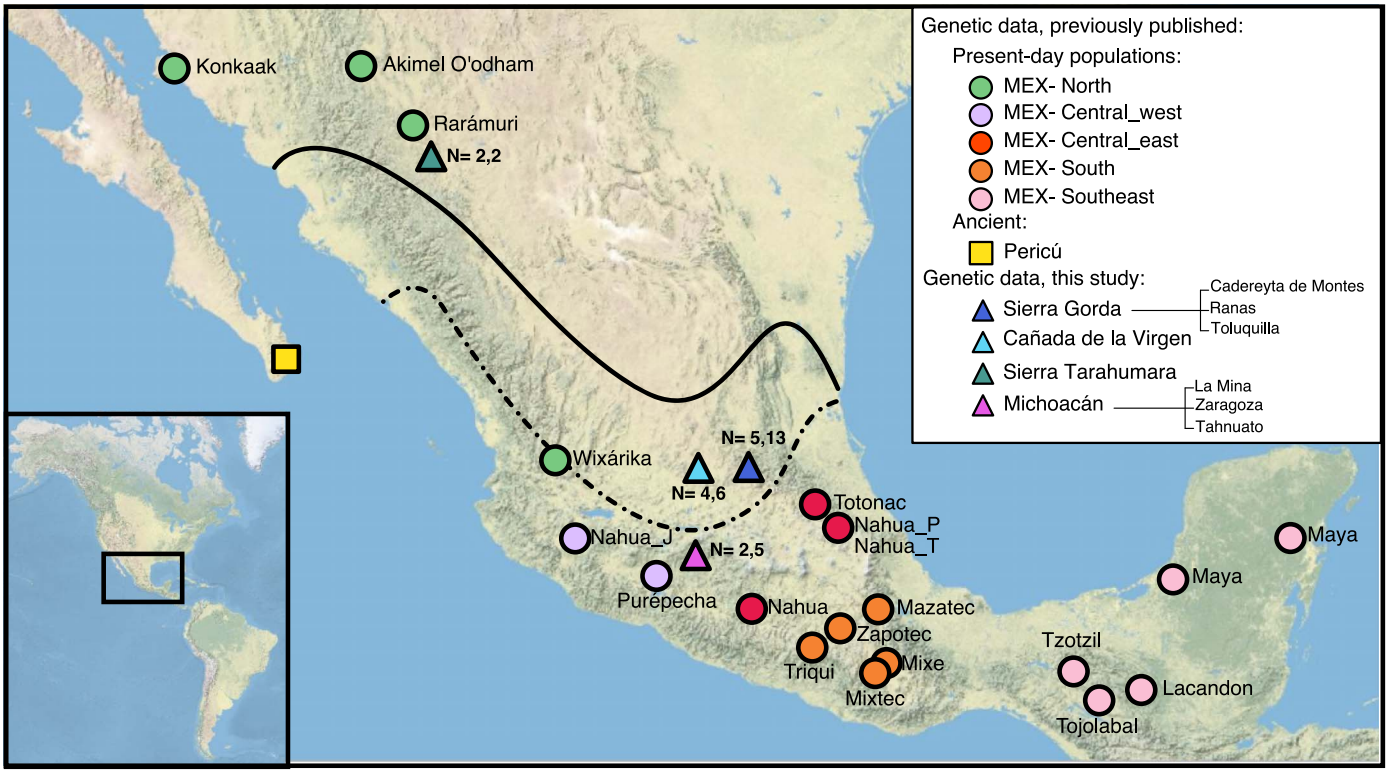


Fig. 1. Map of archeological sites and modern populations. Map of Mexico indicating the location of the pre-Hispanic sites analyzed in this study (triangles) and in previous studies (squares) (*14*, *15*), and the approximate locations of present-day Indigenous populations sampled in previous studies (circles) (*14*, *18*, *22*, *36*). The continuous line indicates the northern frontier of Mesoamerica, with Aridoamerica to the north and Mesoamerica to the south during the years ~800–900 CE. The dashed line indicates this border after the year ~1500 CE, based on (*2*). Numbers next to archaeological sites from this study indicate the number of whole-genome data and mitochondrial genomes recovered per site (separated by a comma).

regulations for these types of studies [see (*16*) for a deeper discussion on the subject]. To engage the neighboring communities, we delivered public lectures to communicate the genetic results to the village closest to the sites in Sierra Gorda.

After processing and sequencing samples, we obtained informative amounts of shotgun sequence data (0.01 to 4.7× genomes) for 12 individuals spanning a time transect of 750 years (600–1351 CE). These include 10 low-depth genomes from within Mesoamerica in central Mexico, and two from Aridoamerica, in the Sierra Tarahumara (table S1). In addition, we reconstructed the mitochondrial genomes for 27 individuals across these sites (5.7 to 1284.8x), spanning a time transect of 1600 years (320 BCE to 1351 CE). (tables S1 and S2). We included in the dataset three previously published ancient genomes (table S3) as well as nine ancient mitogenomes from northern Mexico (table S4), yielding a total of 15 ancient genomes and 36 ancient mitogenomes from the Mexican territory.

We named each sample according to burial and individual identifiers provided by archaeologists and added two additional labels. The first label includes one to three letters that refer to the archaeological site, whereas the second refers to whether the individual dates to before ("b") or after ("a") the drought period. The three pieces of information are divided by an underscore character. Throughout the text we refer to all individuals using these identifiers.

Genetic structure and diversity in uniparental markers

All chromosomal sex assignments matched the morphological sex except one (XX chromosomally and male on the basis of morphology) (table S1) (*17*). It was possible to assign Y-DNA haplogroups to 5 out of the 13 individuals assigned as XY, all of which had a Native American Q lineage. This is in agreement with previous studies on ancient and present-day Indigenous Mexican individuals (*18*–*23*). At the subhaplogroup level, we found Q1a2a1-L54 and Q1a2a1a1a1-M3, with no apparent differences between Aridoamerica and Mesoamerica (table S1).

All 27 reconstructed mitogenomes carried one of the mitochondrial DNA (mtDNA) haplogroups found in Indigenous populations of the Americas: A (n = 10), B (n = 9), C (n = 4), and D (n = 4) (*24*) (tables S1 and S4) (*17*). After merging with nine additional pre-Hispanic mitogenomes from Mexico from other studies (*14*, *15*, *25*), we found all four haplogroups in Mesoamerica (n = 25) and only C and B in Aridoamerica (n = 11) (table S4). Consistently, we observed higher nucleotide diversity (π) values in Mesoamerican populations compared with Aridoamerica (*17*).

The pre-Hispanic mtDNA haplogroup distribution resembles the one in present-day individuals (*23*, *25*, *26*), showing an overall continuity of the matrilineal genetic structure for at least 2300 years. Of the available complete ancient mtDNA genomes from Aridoamerica, 36% harbor haplogroup C, which decreases in frequency to 8% in Mesoamerica, whereas the opposite occurs for haplogroup A, with 40% of presence in Mesoamerica and 0% in Aridoamerica (table S4). This gradient has been previously observed in ancient and contemporary populations from Mexico (*18*, *19*, *23*, *25*–*31*). Moreover, we detected unreported variants in the mtDNA PhyloTree (*32*) for seven individuals assigned to the subhaplogroups A2d, B2c, and B2l (table S5) (*17*).

In Sierra Gorda, we identified subhaplogroup A2d in six individuals—four predrought and two postdrought. After quality filtering (*17*), five of them were included in a median-joining network (table S4). Four (two pre- and two postdrought) clustered together in a single

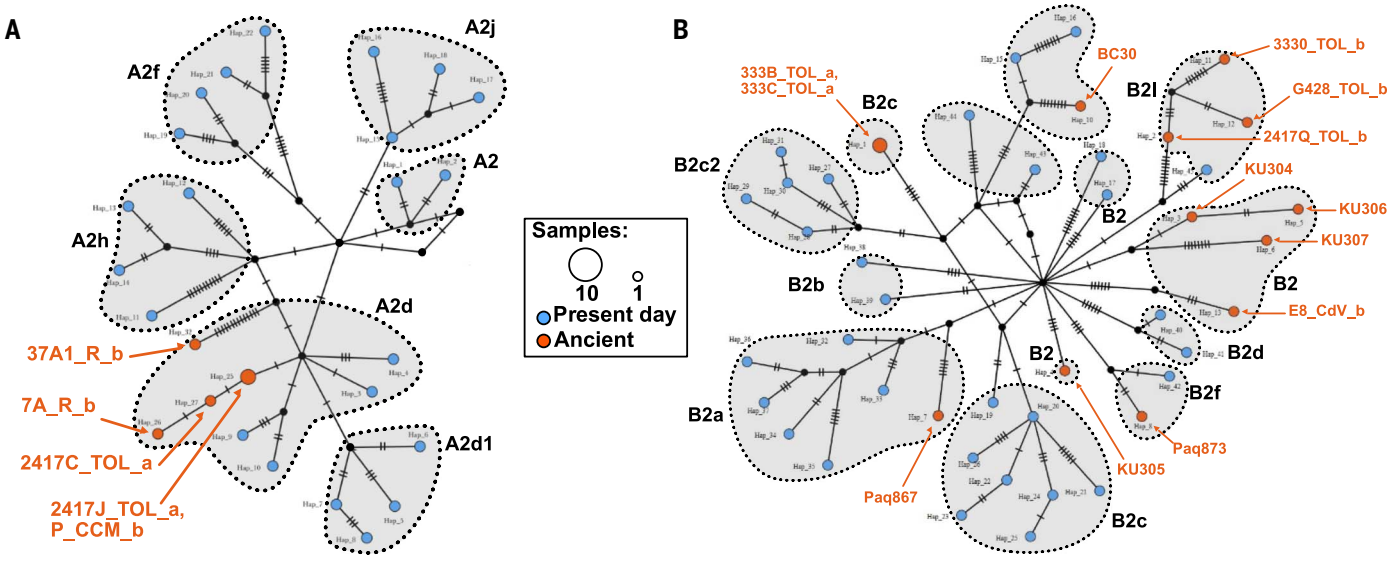


Fig. 2. mtDNA haplotype networks. (**A**) Haplotype network of mitochondrial subhaplogroups A2 including present-day and pre-Hispanic mitochondrial haplotypes from Mexico. Arrows point to individuals from Sierra Gorda within haplogroup A2d from before and after the drought period. (**B**) Haplotype network of mitochondrial subhaplogroups of the lineage B2 including present-day and pre-Hispanic mitochondrial haplotypes from Mexico. Arrows point to pre-Hispanic individuals from Cueva Candelaria, Paquimé, Sierra Gorda, and Cañada de la Virgen.

clade despite spanning a time transect of 1480 years, whereas the fifth individual remained in a different clade within A2d (Fig. 2A). Notably, P_CCM_b and 2417J_TOL_a are found in the same node within A2d. This reflects a continuity of this maternal lineage despite the severe droughts in Sierra Gorda. By contrast, the haplotype network for haplogroup B, including five Sierra Gorda individuals, shows that the three predrought individuals share subhaplogroup B2I, whereas the two postdrought individuals are related and found in a node representing subhaplogroup B2c. Because mtDNA is subject to genetic drift that leads to lineage loss, it cannot accurately inform about genetic structure or admixture events. The discrepant patterns between haplogroups A and B could be explained by a number of scenarios, which are difficult to test because of the small sample size. This highlights the need to explore autosomal data to test the population-replacement hypothesis reliably.

We further leveraged the mtDNA data to estimate the past female effective population size (Ne) using Extended Bayesian Skyline Plots (EBSP) in pre-Hispanic (n = 29) and present-day Mexico (n = 232) (tables S4 and S6), per haplogroup and all haplogroups together (fig. S6, A to H). All EBSP runs (table S7) showed wide (95%) confidence intervals (CIs) except the EBSP of only present-day Mexico, in which we observed a population expansion ~5 kyr B.P. (thousand years before present) and a population decline ~500 yr B.P., though the median values do not support the latter (fig. S6, G and H). Despite the wide CI, there is a clear distinction between the estimated female Ne per haplogroup, with haplogroup A reaching the highest present-day values (104,048), followed by D (100,812), and decreasing almost an order of magnitude for haplogroups B and C (37,155 and 7226, respectively). To improve the estimations and gain insights about the past female Ne in a wider geographical range, we merged our dataset with available ancient mtDNA sequences from South America (n = 137) (table S6). The EBSPs for haplogroups A, B, and D, respectively, also show wide CIs (fig. S6, I, J, and N), whereas haplogroup C shows a clearer pattern of population expansion starting ~15 kyr B.P. and a population decline starting ~5 kyr B.P. (fig. S6K). Consistent with our previous observation, the estimated female Ne is notably higher for haplogroups A and D than for B and C in the merged dataset (fig. S6, I to L).

Autosomal genetic structure in pre-Hispanic Mexico

We performed ADMIXTURE (*33*) and principal component analysis (PCA) (*34*, *35*) to visualize the genetic relationship and genetic structure at the autosomal level between the pre-Hispanic individuals and other ancient individuals representative of North, Central, and South America (table S3), along with present-day Native American individuals from Mexican and continental populations (table S8) (*14*, *18*, *22*, *36*, *37*). All pre-Hispanic individuals from Mexico, as well as all ancient individuals from California, Belize, and Patagonia, clustered together in PCA with present-day Indigenous populations from Mexico (fig. S7) and shared similar genetic composition in the ADMIXTURE analysis (fig. S8).

We projected the pre-Hispanic individuals onto a principal-components space of only pre-Hispanic and present-day Indigenous populations from Mexico (*14*, *18*, *22*, *36*) using an approach named "missing DNA" PCA (mdPCA) that corrects for the high missing fraction of low-depth ancient genomes (*17*). All ancient individuals clustered closely in the mdPCA with present-day populations from the same geographical region except for individual MOM6_ST_a, who is intermediate between northern and central present-day populations. Both pre- and postdrought Sierra Gorda individuals clustered together with present-day Nahua from Jalisco, Purépecha, Totonac, Nahua from Puebla, and Nahua, all of which are from central Mexico (Fig. 3A). This pattern was replicated when using traditional PCA (fig. S9) and a temporal factor analysis approach (*38*), which corrected the ancestral relationships with the dates of the pre-Hispanic individuals (fig. S10) (*17*). In agreement, the ADMIXTURE analysis assuming six ancestral components revealed that pre-Hispanic individuals show similar genetic ancestry proportions to those observed in present-day populations from the same region (Fig. 3B) and that individuals from Sierra Gorda show a homogeneous genetic composition independent of the period in which they lived (Fig. 3B).

We then performed several combinations of D- and outgroup-f3 statistics to further explore the genetic structure of ancient and present-day populations (*17*). Outgroup-f3 statistics were of the form f3(Test, Source1; YRI), where the Test is the pre-Hispanic individual under analysis and Source1 is a present-day Indigenous population. D-statistics were of the form

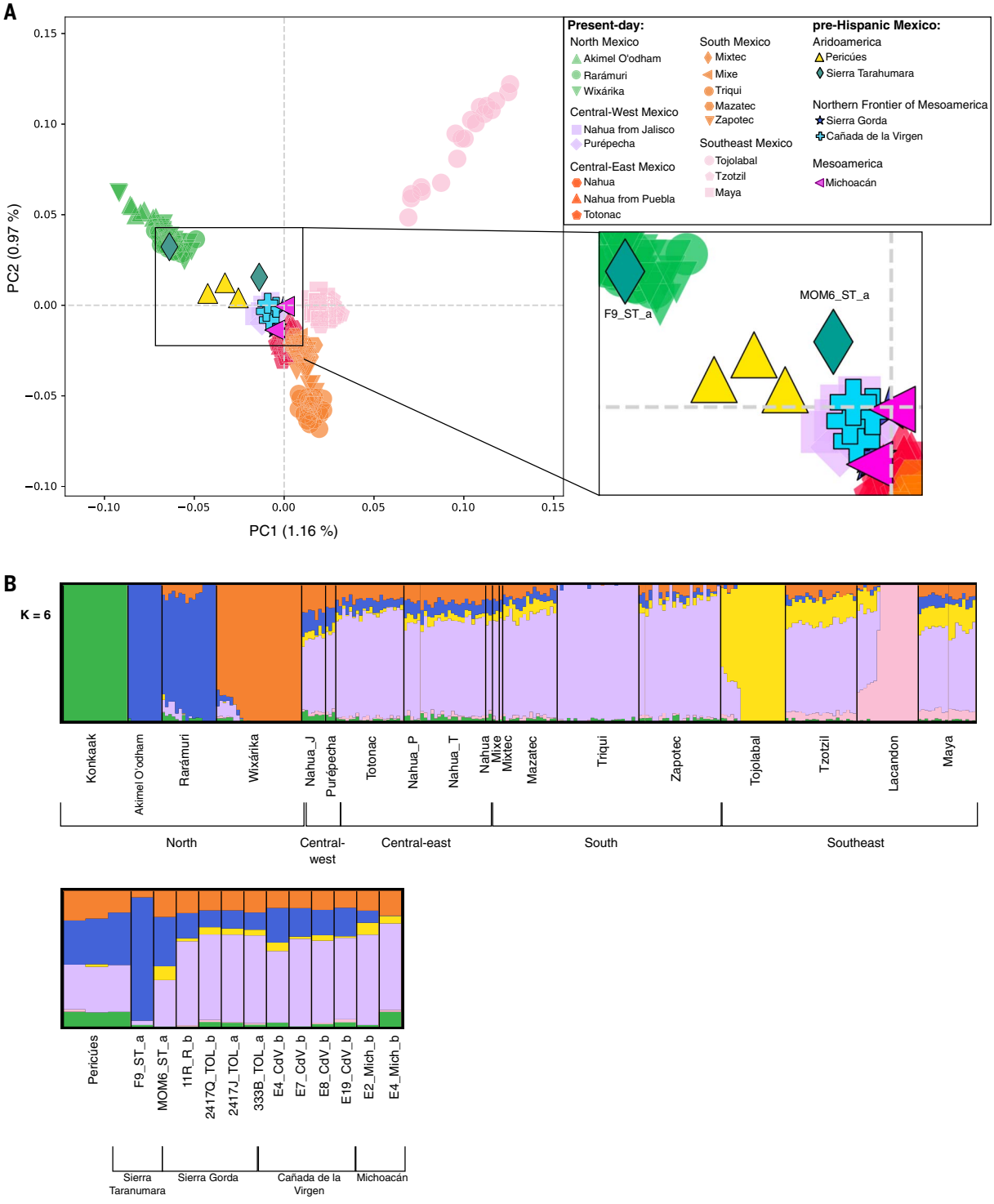


Fig. 3. Genetic structure of Indigenous and pre-Hispanic individuals. (**A**) Missing DNA Principal Component Analysis (mdPCA) of present-day Indigenous Populations with the ancient and pre-Hispanic genomes from Mexico. (**B**) Unsupervised clustering analysis of same data by ADMIXTURE with k = 6, showing green, blue, and orange components in northern populations, yellow and pink components in southeast populations, and a violet component in south and central Mexico populations.

D(Pop1, Pop2; Test, YRI), where Test is the pre-Hispanic individual under analysis, and Pop1 and Pop2 are all possible combinations of present-day Indigenous populations. YRI represents Yoruba as the outgroup population in both tests. We found that F9_ST_a shared higher genetic drift with present-day northern Mexican populations than with any other present-day population (fig. S11 and table S9) and was significantly more related to the present-day Rarámuri than to any other population when tested in the form D(Pop1, Rarámuri; F9_ST_a, YRI) (fig. S12 and table S10). However, a population continuity test that considers the read counts in the ancient individual and allele frequencies in the present-day population (*39*) rejected the null hypothesis of population continuity between F9_ST_a and

present-day Rarámuri, which had a P value of $10^{-499.3}$ (*17*).

The genetic ancestries of pre-Hispanic individuals from central Mexico are more complex. They do not seem to have a higher shared genetic drift with any present-day Indigenous population according to the f3-outgroup values, for which almost all standard errors overlapped (figs. S13 to S15 and table S9). Furthermore, in D-statistics analysis, we found mostly $D = 0$ in all possible combinations of Pop1 and Pop2 (figs. S16 to S18 and table S10), which was expected in populations with extensive gene flow between them and not completely diverged from each other.

Genetic diversity in pre-Hispanic Mexico

We estimated the conditional nucleotide diversity (CND) (*40*) and runs of homozygosity (ROH) (*41*) in pre-Hispanic and present-day Indigenous individuals from Mexico (*18*, *22*) to make inferences about past genetic diversity and effective population sizes and investigate changes in patterns of genetic variation that may have arisen after the drought period (*17*). Both the CND values and the ROH distribution show that pre-Hispanic and present-day individuals from northern Mexico (Pericúes and Akimel O'odham) have the lowest CND values and the highest sum of inferred ROH [>4 centimorgan (cM)] compared with other populations of their respective periods (tables S11 to S13). Pericú hunter-gatherers had the lowest CND values (fig. S19 and table S12), but the ROH proportion revealed a larger population size than another ancient hunter-gatherer population from Patagonia (fig. S19 and table S12) (*17*). These results agree with previous studies that report lower genetic diversity and longer ROH in hunter-gatherer populations from South America than those observed in other regions of the Americas (*42*).

Pre-Hispanic individuals from Sierra Tarahumara, Toluquilla, and Cañada de la Virgen show similar values of CND and sum of inferred ROH segments. In Toluquilla, we found that the CND value increases with the date difference between pairs of individuals, which probably reflects the accumulation of new mutations during the 489 to 680 years of difference between pre- (2417Q_TOL_b) and postdrought (2417J_TOL_a and 333B_TOL_a) individuals. Lastly, the segment-size distribution of ROHs suggests that the pre-Hispanic individuals studied here belonged to populations with small Ne (2Ne = 1600 to 6400) (fig. S20) (*17*), which is in agreement with Ne values previously calculated for present-day northern and southern populations (*43*).

Genetic continuity before and after the 900–1300 CE droughts in the Sierra Gorda

To formally test the hypothesis of population replacement in the Sierra Gorda from the northern frontier of Mesoamerica during the 900–1300 CE droughts, we applied different combinations of outgoup-f3 and D-statistics. We used individuals from Sierra Gorda from the pre- (TOL_b and R_b) or postdrought (TOL_a) periods as the Test population in an outgroup-f3 of the form f3(Test, Source; YRI); for the Source, we used another ancient individual from Mexico or ancient representatives of other regions in the Americas (North, Central, and South America) (table S3).

Under a population replacement scenario, we would expect the Toluquillan individuals who lived after the drought to be more closely related to pre-Hispanic individuals with higher proportions of Aridoamerican genetic ancestry (Pericúes and F9_ST_a) than to Toluquillan individuals who lived before the drought (2417Q_TOL_b). Instead, what we observed is that the pre-Hispanic individuals from Sierra Gorda before and after the climate change episode share higher genetic drift between them than with any other pre-Hispanic individual (Fig. 4A), except for the individual 11R_R_b, who shows higher shared genetic drift with an individual from Cañada de la Virgen (fig. S21 and table S14). However, standard errors for 11R_R_b are high because of its low coverage (fig. S21). In the case of the two individuals from Sierra Tarahumara, the highest value of genetic drift is between them (fig. S22). In contrast, outgroup-f3 tests with individuals from Cañada de la Virgen (CdV) as the Test population in f3(CdV, Source; YRI) (Fig. S23) have the highest outgroup-f3 values when a related individual [inferred with a relatedness analysis, (*17*)] from the same site is used as the Source, but the Source with the second-highest value does not belong to Cañada de la Virgen.

We performed D-statistic tests in the form D(SG_b, SG_a; Pop3, YRI), with the individuals from Sierra Gorda (SG) of each period as the Source—first without collapsing them per period and then merging the individuals in their respective group—and Pop3 being any pre-Hispanic or present-day individual or population from Mexico. SG_b represents predrought individuals 2417Q_TOL_b and 11R_R_b, whereas SG_a includes postdrought individuals 2417J_TOL_a and 333B_TOL_a. Notably, 2417Q_TOL_b shared more alleles with California or Pericú individuals than with the postdrought individuals, but this sharing was not observed when only transversions were used (fig. S24 and table S15). When merging the individuals per period, none of the comparison under this model deviates significantly from $D = 0$, meaning that SG_b and SG_a form a clade to the exclusion of other individuals and are more closely related to each other than with any other pre-Hispanic or present-day population. The only exception was when Pericúes were used as Pop3, where $D > 0$ with a z-score = 3.625, which indicates that the predrought individuals are more closely related to Pericúes (fig. S25 and table S16). However, this significant deviation disappears when repeating the analysis with only transversion sites (Fig. 4B and table S17).

We then used f-statistic–based admixture graphs (qpGraph) (*44*) to further evaluate the population continuity scenario in Sierra Gorda by testing the fit of models that grouped individuals from before and after the drought period in the same clade. For the model, we used the 11,500-kyr-old USR1 (*45*) as the surrogate for ancient Beringian, the 10,700-kyr-old Anzick (*46*) individual as the surrogate for ancestral Southern Native American (SNA) (*14*), and the 4200-kyr-old Ancient Southwestern Ontario individual (ASO) (*15*) as the surrogate for Northern Native American (NNA) (*14*). In the analysis, we included the pre-Hispanic individuals with the highest coverage: 2417Q_TOL_b, 2417J_TOL_a, and 333B_TOL_a from Toluquilla, and F9_ST_a as the surrogate for Aridoamerica. The three surrogates can be modeled in the same clade separated from F9_ST_a, with no contribution to this (Fig. 4C).

Finally, we used qpWave (*44*) to assess whether pre- and postdrought Sierra Gorda individuals can share a demographic history without the need of an additional wave of genetic ancestry from Aridoamerica or another population. Indeed, we found that pre- and postdrought Sierra Gorda individuals can be modeled in the same clade without the need of an admixture event ($P > 0.12$). By contrast, an additional genetic source is needed to reflect genetic variation in F9_ST_a, Pericúes, Akimel O' odham, and Maya (table S18).

Admixture model for Central Mexico populations

Given the complex relationships observed for ancient individuals in central Mexico, we tested additional demographic models using qpGraph. The base model that fit the pre-Hispanic populations includes the Anzick (*46*) individual as the ancestral SNA (*14*), present-day Athabascan (*14*) as the representative for NNA (*14*), and present-day Indigenous populations in Mexico (*36*) that represent the northern (Konkaak), central (Nahua from Puebla), and southern (Triqui) regions. Furthermore, this model involves a split of the SNA into S2A and S2B sources. The pre-Hispanic populations from Sierra Gorda and Cañada de la Virgen show different levels of genetic ancestry from SNA and NNA. We found that the population from Sierra Gorda had a higher percentage of the SNA branch from S2B (64%) than the population from Cañada de la Virgen (37%), which had 62% genetic ancestry shared with the NNA branch, indicating a different demographic history for each of these populations

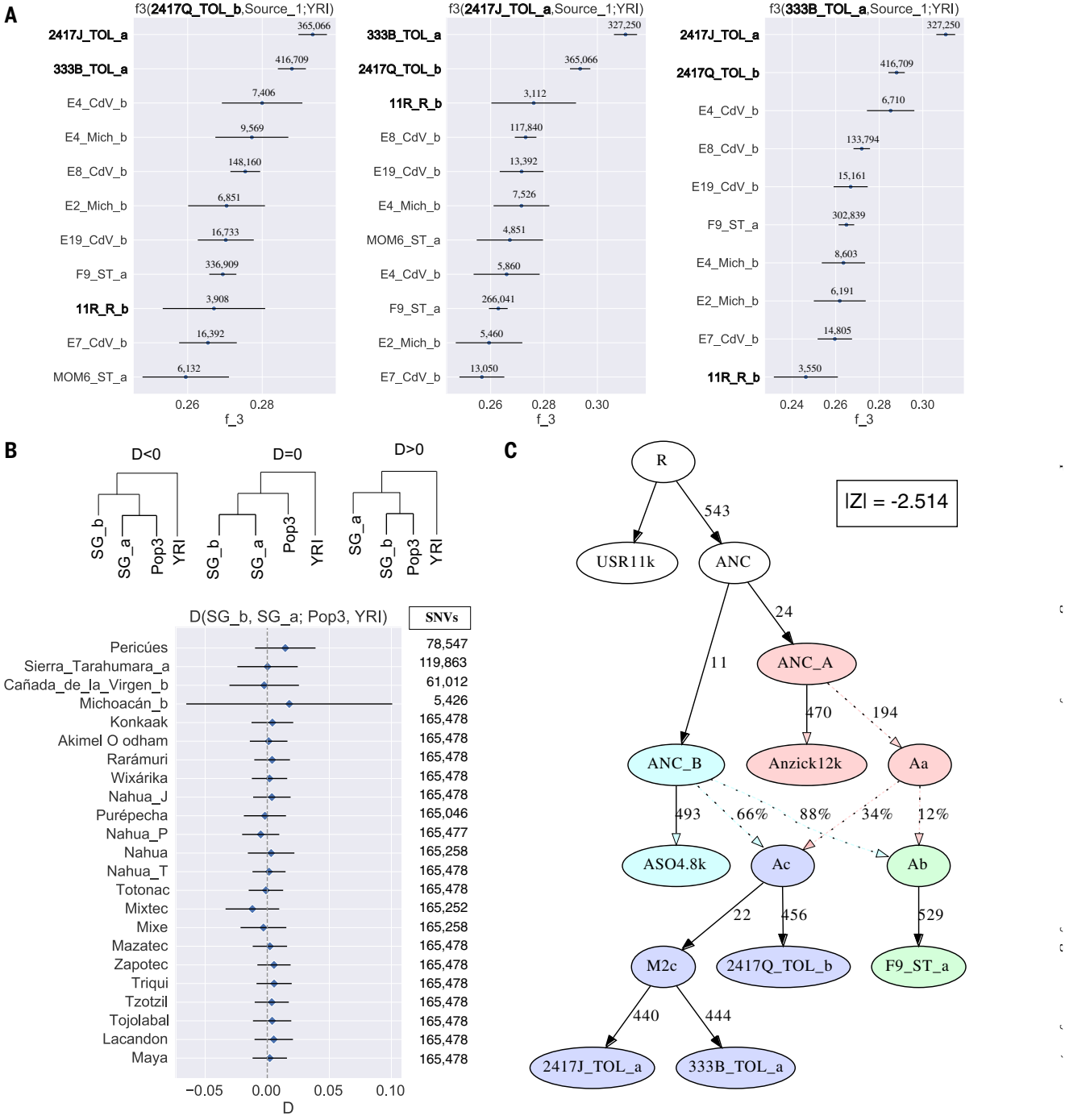


Fig. 4. Population continuity in Sierra Gorda. (**A**) Shared genetic drift and genetic affinities of pre-Hispanic individuals from Toluquilla, 2417Q_TOL_b, 2417J_TOL_a, and 333B_TOL_a, compared with pre-Hispanic individuals from Mexico. Higher values of *f3* indicate higher shared genetic drift. Point estimates and one standard error are shown. (**B**) D statistics in the form D(SG_b, SG_a; Pop3,YRI), including transversions only. SG_b includes 2417Q_TOL_b and 11R_R_b, and SG_a includes 2417J_TOL_b and 333B_TOL_a. Pop3 is any of the other pre-Hispanic or present-day populations from Mexico and is shown in the *y* axis. YRI as an outgroup corresponds to Yoruba from Africa from the 1000 Genomes Project (*37*, *95*). Expected tree topologies according to different *D* values are drawn on the top of the plot. Blue dots indicate no significant deviations from *D* = 0 (*z*-score > 3). The number of SNVs used in each test is shown in the right. (**C**) qpGraph model of individuals from Sierra Gorda from before and after the drought is shown in the same clade, without contribution of F9_ST_a to the postdrought individuals.

despite their relative geographic proximity (fig. S26). This model supports previous studies that report several admixture events between the two branches that have given rise to Central and South American populations (*15*). For the Michoacán population, we obtained terminal branches with no support (zeros); thus, we did not make conclusions with this model for the Michoacán pre-Hispanic population.

Ghost population contribution to pre-Hispanic Mexico

Earlier studies report the contribution of "ghost" genetic ancestry from an unsampled group, designated as UpopA, among the present-day

Mixe from Mexico (*47*). We tested the presence of this UpopA in the pre-Hispanic individuals using combinations of admixture graph models. Notably, Sierra Tarahumara (represented by F9_ST_a) and Cañada de la Virgen (represented by E8_CdV_b) showed a genetic ancestry contribution from a ghost population at 28% and 17%, respectively (fig. S27, A and B). This was consistent with the ghost genetic ancestry, UpopA, that was previously reported in Mixe (fig. S27C), whereas models including individuals from the other archaeological sites were rejected. When we modeled the pre-Hispanic individuals and Mixe together, we found that Sierra Tarahumara and Mixe share the same ghost genetic ancestry, UPopA (with a better *z*-score when Sierra Tarahumara received NA1 genetic ancestry) (fig. S28), whereas Cañada de la Virgen required the contribution of a second ghost genetic ancestry from that found in Mixe (models without this second ghost genetic ancestry are rejected, *z*-score > 3), which we named UPopA2 (fig. S29). Finally, we included Sierra Tarahumara, Cañada de la Virgen, Mixe, and the two ghost populations in the same model and confirmed that F9_ST_a and Mixe share the contribution from UpopA1, whereas Cañada de la Virgen requires the contribution from an additional unsampled group, UpopA2 (Fig. 5) (*18*).

Discussion

We have generated an ancient genome dataset from Mexico to address long-standing questions about population dynamics between Aridoamerica and Mesoamerica. The data allowed us to (i) describe levels of population structure and genetic diversity in Mexico before European colonization, (ii) test a previous hypothesis of population replacement in central Mexico following a drastic climate change between 900–1300 CE, (iii) demonstrate a complex admixture model for central Mexico populations, and (iv) detect a contribution from unsampled population A (UpopA) to some northern and central Mexico populations.

Our findings show a geographical structure in pre-Hispanic individuals that differentiates northern and central populations from Mexico, which is consistent with the northwest-southeast cline observed in the genetic structure of present-day Indigenous populations (*36*). The pre-Hispanic individuals cluster in proximity to present-day Indigenous populations from the same geographical area in the mdPCA and exhibit similar ancestral components based on ADMIXTURE analysis, except for individual MOM6_ST_a (Fig. 3 and figs. S9 and S10). This reflects an overall conservation of the genetic structure of the populations inhabiting the Mexican territory (*36*) for at least 1400 years (which is the date of the most ancient individual in the dataset). This is consistent with demographic models based on present-day Indigenous populations (*43*), (*48*) which propose a northern-southern population split between 4000 and 10,000 years ago followed by multiple waves of admixture events between them (*48*). This geographical structure is also reflected in the maternal lineages (table S5). The spatial distribution of the haplogroups found—namely A, B, C, and D—closely resembles that of present-day Mexico (*31*, *49*).

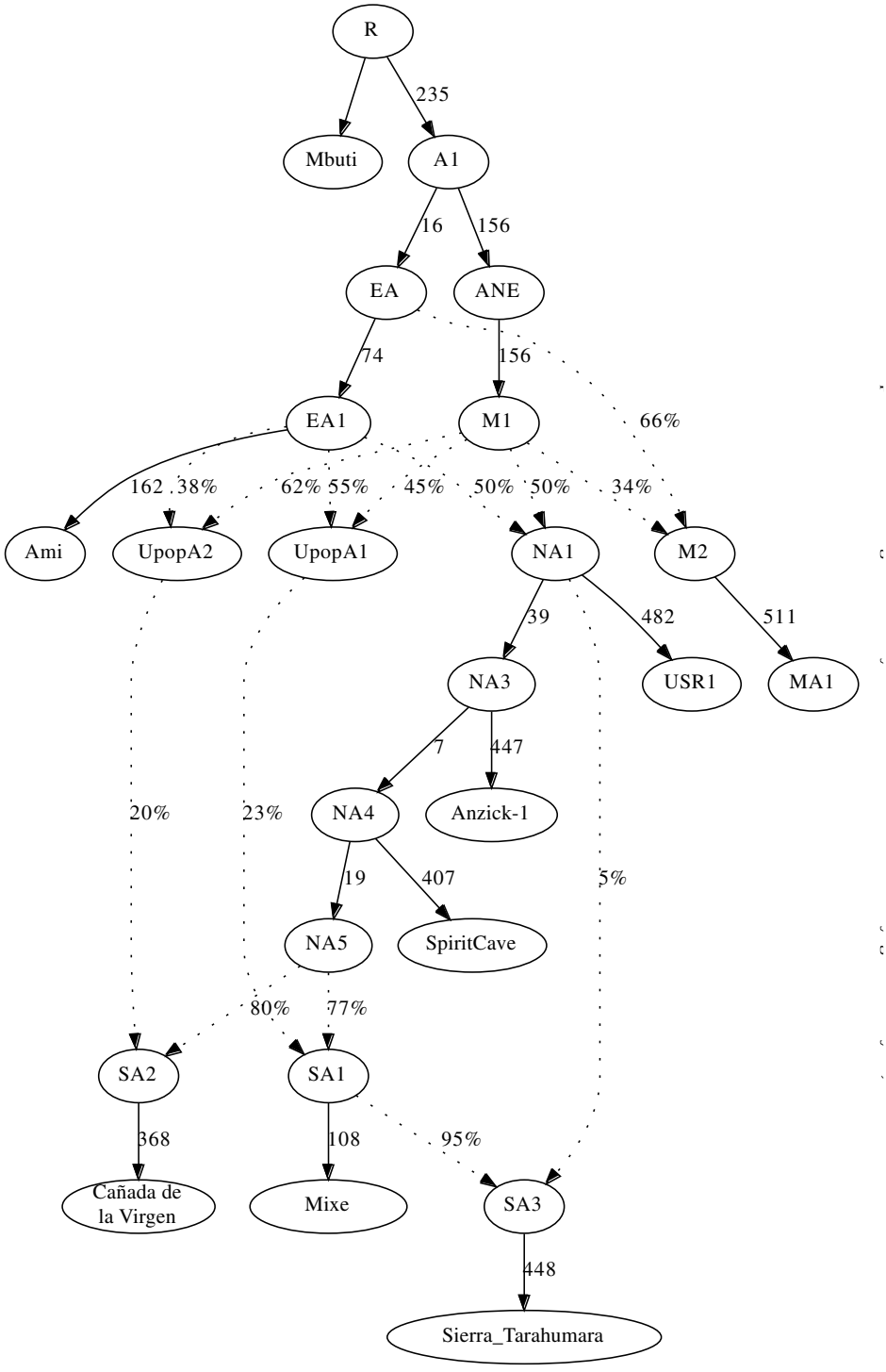


Fig. 5. Ghost ancestries in pre-Hispanic individuals from Mexico. qpGraph model for Sierra Tarahumara and Cañada de la Virgen including present-day Mixe. The tree has no inner 0 branches and a *z*-score of 2.415. The model shows different ghost ancestries, UPopA1 and UPopA2, in Sierra Tarahumara and Cañada de la Virgen individuals, respectively. UPopA1 indicates the previously unsampled population genetic ancestry reported in Mixe, whereas UPopA2 is a previously unidentified genetic ancestry. All *f* statistics are within a standard error of 1.75.

Although we lacked the resolution to make accurate inferences about past female population size based solely on the ancient mitogenomes in the EBSP analyses, we were able to recapitulate a population expansion ~5 kyr B.P. when using modern mitogenomes, as it has been previously observed with whole-genome data from present-day Indigenous people from Mexico (*50*). This expansion is likely related to the domestication and propagation of maize cultivation in Mexico (*51*). Notably, we did not recover a signal of a population bottleneck ~500 yr B.P. after European colonization in the mitogenome data. We speculate that the demographic impact of European colonization on the female population was variable across Mexico and less drastic than in insular and other isolated regions of the Americas (*52*–*54*). The recovery of additional ancient mitogenomes from a wider temporal and spatial range will provide resolution to paint a more accurate picture of the female population history.

When we focused on the genetic diversity at the autosomal level, our results revealed that ancient Aridoamericans and the populations at the northern frontier of Mesoamerica have similar levels of conditional nucleotide diversity (CND) between them and those observed in present-day Indigenous populations from northern (Akimel O'odham) and southern Mexico (Mixe and some Mayan) (fig. S19). However, we caution that we lack genome-wide genetic information of present-day individuals from the exact same site as the pre-Hispanic individuals in central Mexico analyzed here to make a direct temporal comparison. Yet, we observed that the lowest CND and the highest total number of segments in ROH were found in the ancient hunter-gatherer Pericúes and the present-day Akimel O'odham population, both from northern Mexico. In contrast, pre-Hispanic individuals from Sierra Gorda and Cañada de la Virgen and present-day populations from central, southern, and southeastern Mexico had the lowest total number of segments in ROH (fig. S20). Altogether, these results are reflective of the lower population sizes maintained in isolated Aridoamerica populations and the larger population sizes in pre-Hispanic and present-day populations in Mesoamerica. Notably, the pre-Hispanic individuals from this study came from small-sized villages (*55*, *56*) and do not necessarily reflect the demography of other Mesoamerican pre-Hispanic populations from larger and multiethnic pre-Hispanic metropolises (e.g., Teotihuacán, Tenochtitlán, or Palenque) (*57*, *58*), where higher levels of diversity and lower ROH would be expected. Ancient DNA studies on individuals from these sites would be needed to reveal the extent of genetic diversity in these sites before the population collapse inflicted by European colonization.

To directly test the hypothesis of a population replacement at the northern frontier of Mesoamerica by Aridoamerican hunter-gatherers after the 900–1300 CE droughts, we studied Sierra Gorda individuals from the times before and after the climate change episode. Outgroup-f3 analyses showed that individuals from Sierra Gorda pre- and postdrought shared higher genetic drift between them than with any other pre-Hispanic individual. Using D statistics, we observed that pre- and postdrought individuals from Sierra Gorda always form a clade to the exclusion of the rest of pre-Hispanic or present-day populations from Mexico. Then, qpGraphs showed that pre- and postdrought individuals from Sierra Gorda can be modeled in the same clade, which is different from the one formed by F9_ST_a as an Aridoamerican surrogate. Regardless, F9_ST_a might not belong to the population that allegedly replaced the population in Sierra Gorda following the drought; we used F9_ ST_a safely as a surrogate for northern genetic ancestry because previous studies have estimated that the split-time between the northern and southern ancestries occurred ~7200 years ago (*43*). Therefore, even though F9_ST_a likely did not belong to the population that allegedly replaced inhabitants of Sierra Gorda following the drought, we can confidently use this genome as a surrogate for northern genetic ancestry because, by the time of the climate change episode, both ancestries were already well differentiated. Furthermore, with qpWave, we confirmed that the genetic makeup of pre- and postdrought individuals from Sierra Gorda can be explained within the same genetic history without an additional source of genetic variation. Even if this replacement had happened by an unsampled northern population, we would still expect to see signals that significantly differentiated the postdrought individuals with an input from Aridoamerica in the qualitative (mdPCA and ADMIXTURE) and the quantitative (Outgroup-f3, D statistics, qpGrap, and qpWave) analyses, which is not the case. Instead, the evidence points to population continuity in Sierra Gorda after the climate change episode.

A possible explanation for a population continuity despite the droughts is that the favorable climatic conditions at the northern Sierra Gorda maintained higher humidity than in other arid sites of the northern frontier of Mesoamerica, such as Cañada de la Virgen. Because agriculture was the main subsistence strategy in Cañada de la Virgen, the heavy droughts forced its inhabitants to migrate to other regions, which resulted in an abandonment of the site between 1000–1100 CE (*56*). By contrast, the main subsistence strategy in Toluquilla and Ranas was the mining and trade of cinnabar, a valuable mineral of sacred value in pre-Hispanic cultures (*55*, *59*, *60*). We hypothesize that the cinnabar trade and the landscape of the Sierra Gorda allowed the peoples of Toluquilla and Ranas to subsist despite low rainfall conditions during the drought. Notably, we found shorter ROH segments in the predrought pre-Hispanic individual than in the postdrought individuals. This could indicate a reduction in population size after the climate change episode and suggest a possible demographic impact in the same population. Additional assessment of this hypothesis at other sites at the northern frontier of Mesoamerica will help explain the population migrations and dynamics in a wider geographical range.

We were also interested in obtaining insights into the demography of central Mexico populations because previous attempts of demographic modeling have been hampered by the high genetic heterogeneity observed in these populations (*43*, *50*). Our outgroup-f3 and D statistics results show that, although pre-Hispanic individuals from Mesoamerica in central Mexico share higher genetic drift with present-day populations from that region, none of these relationships were statistically significant. Using qpGraph, we found that the pre-Hispanic populations from central Mexico all have different genetic-ancestry sharing with the NNA and SNA branches. This was expected given previous studies on ancient genomes from Central and South America, which report multiple admixture events between these two branches after their split ~15,000 years ago (*15*, *48*). Together, these observations point to a scenario in which populations from central Mexico have not completely diverged from one another, possibly because of extensive gene flow, as expected on the basis of the active commercial exchange between different Mesoamerican populations that has occurred for centuries (*61*). This interaction was mainly through trade routes and alliances between different nations (*61*), as revealed in present-day Indigenous populations who share identity-by-descent (IBD) segments (*36*, *48*). The study of IBD segments in present-day Indigenous populations from Mexico has also evidenced gene flow between Mesoamerican and Aridoamerican populations in pre-Hispanic times (*48*). However, this gene flow occurred less frequently than within Mesoamerica. The individual MOM6_ST_a from Aridoamerica may have been the result of such admixture between Mesoamerican and Aridoamerican ancestors.

Furthermore, the qpGraph admixture models that we explored for the pre-Hispanic populations showed that F9_ST_a and the ancient individuals from Cañada de la Virgen have genetic ancestry from a ghost, or unsampled, population. A contribution from an unsampled population named UpopA was previously identified in present-day Mixe

(47) as well as in present-day northern and central Indigenous populations from Mexico (48). UpopA was estimated to have diverged ~24,700 years ago from Native Americans (47). Notably, although Sierra Tarahumara shares the contribution of UpopA with Mixe, a second ghost population (UpopA2) was needed to model Cañada de la Virgen in the admixture model together with Sierra Tarahumara and Mixe (Fig. 5). This observation reveals a complex population history in the Americas during the late Pleistocene that needs to be further characterized. Additional ancestral DNA studies from the Americas could help identify the source of both ghost ancestries, which seemingly contribute to many present-day Indigenous populations from Mexico.

We find that the pre-Hispanic population structure from over a thousand years ago can still be observed today. Our work, together with previous studies (15, 36, 43, 48, 50), shows that the demographic events that gave rise to Aridoamerican and Mesoamerican populations are more complex than previously thought. Commercial trade routes may have contributed to increased mobility, facilitating gene flow between different populations within and between various cultural areas. Furthermore, we found genetic continuity in the Sierra Gorda region at the northern frontier of Mesoamerica, which suggests that the local population stayed in their homeland despite the long-standing droughts that forced other populations to abandon their cities. The identification of a second ghost genetic ancestry contribution to some pre-Hispanic central Mexico populations reveals a complex past that needs to be characterized through the ethical study of ancient genomes from Mexico. Our study opens the door for further research to address the questions of the unknown genetic past and population dynamics of Mexican pre-Hispanic populations, whose genetic legacy is retained today among Indigenous and admixed populations.

Materials and Methods

Laboratory procedures

DNA extraction and library preparation (before amplification) were performed in the Human Paleogenomics Laboratory, a clean lab facility at the International Laboratory for Human Genome Research, Universidad Nacional Autónoma de México (LIIGH-UNAM). Bones and teeth surfaces were cleaned with a 1% sodium hypochlorite solution, followed by a solution of 75% ethanol. Then, the surface was UV irradiated (256 nm) for 1.5 min on each side using a UVP CL-1000 crosslinker. A Dremel tool was used to remove the outer surfaces of bones and teeth. Bones were cut to get an inner sample of around 100-200 mg for DNA extraction. Teeth were cut at the cementoenamel junction, the roots were wrapped in aluminum foil and pulverized using a hammer. Around 100-200 mg of the root was used for DNA extraction. DNA extraction was performed using the methods described in (62, 63), as described in table S2.

DNA extraction from individual P_CCM_b was performed taking 150 mg of mummified skin. The sample was washed in deionized sterile water and dried. Then, the sample was UV irradiated (256 nm) for 1.5 min using a UVP CL-1000 crosslinker. Epithelium was cut in small pieces and incubated at 50°C for 24 hours in lysis buffer (10mM Tris-HCl, 10mM NaCl, 5mM $CaCl_2$, 2.5 mM EDTA, 1% SDS, 10 mg/ml proteinase K, 10 mg/mL DTT). Followed by centrifugation at 16,100 × g for 5 min and recovery of the supernatant. Then, DNA purification was performed according to the protocol in (63).

Double-stranded libraries for Illumina sequencing were prepared according to (64), using single-indexed adapters with 6–base pair (bp) barcodes or double-indexed with 7-bp barcodes, depending on the sequencing platform, NextSeq550 or NovaSeq, respectively. Libraries were analyzed with qPCR to determine the optimum number of cycles during the indexing PCR step. Barcoded libraries were sequenced for a first screening on the NextSeq550 equipment from Illumina using a 2x75 run at either LANGEBIO's genomics core facility (National Laboratory of Genomics for Biodiversity, Irapuato, Guanajuato) or INMEGEN genomics facilities (National Institute of Genomic Medicine, Mexico City). Depending on the quality of the libraries (% endogenous and % clonality), some were chosen to be subjected to mitochondrial genome capture or whole-genome capture using Daicel Arbor Biosciences (Ann Arbor, MI, USA) commercial kits (table S2). Captured libraries were sequenced to assess complexity and yield with the tool preseq (65) and sequenced to higher depth using the NextSeq 550 (2x75 cycles) at LANGEBIO's core facility or in the NovaSeq at the Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT) with an S1, 2x100 cycles run (table S2). Sequencing runs for Sierra Tarahumara individuals were carried out at the Sci LifeLab Uppsala Sequencing Center and at the Bustamante Lab in the Department of Genetics at Stanford University (table S2).

Sequence data processing

Raw reads were processed using Adapterremoval (66) for trimming Illumina adapter sequences and collapsing of pairs of reads with an overlap of at least 11 bp (–trimns–trimqualities–qualitybase 33–minlength 30–collapse). Collapsed reads with > 30 bp and quality above 33 were retained for downstream processing.

The retained reads were mapped to the human reference genome b37 (hg19), with the mitochondrial sequence replaced by the revised Cambridge reference sequence (rCRS, NC_012920). Mapping was done with bwa 0.7.13, aln algorithm, seed disabled (-l 500), and keeping reads with mapping quality > 25. Clonal duplicates and reads mapping to more than one place in the genome were removed with SAMtools rmdup (67) and eliminating reads with labels 'XT:A:R' and 'XA:Z', respectively. Then, reads were realigned using GATK with RealignerTargetCreator and IndelRealigner (68). To authenticate ancient DNA sequence data, mapDamage2 (69) was used to assess the damage patterns and distribution of reads length, using default parameters (figs. S1 to S4). Alignments to rCRS were separated into new bam files using SAMtools (67) to analyze the mitochondrial genome for haplogroup assignment, haplotype network analysis and Extended Bayesian Skyline Plot (EBSP). For the analysis of autosomal variants, base qualities were rescaled with mapDamage2 (69). Depth of coverage was estimated using ATLAS (70).

mtDNA analysis

The program Schmutzi (71) was used to estimate mtDNA contamination and to obtain consensus sequences. The reads aligned to the Cambridge reference sequence (rCRS, NC_012920) were analyzed with the script contDeam.pl with parameters–library double and the default–uselength, and schmutzi.pl was then run with 8 threads. Haplogrep 2 (72) was used to assign a mitochondrial haplogroup and quality of assignment for each library using the fasta file with the consensus mitochondrial genome generated by Schmutzi (table S1).

For the estimation of the mitochondrial genetic diversity, we aligned the 27 mitochondrial genomes reconstructed in this study along with the eight public mitochondrial genomes from Mexico (table S4), using the BioEdit sequence alignment editor v. 7.0.5.3 (73). The haplotype list was obtained with DNA Sequence Polymorphism software (DNAsp) v. 6.12.03x64 (74). The sites with gaps were not considered. Then, we estimated the nucleotide diversity Pi through the genetic distance method of Tajima and Nei, with the software Arlequin v. 3.5.1.3 using default parameters (75).

Consensus fasta sequences of the ancient mtDNA genomes from this and previous studies were aligned together with previously published modern sequences (14, 15, 18, 25, 32, 53, 76–85) (tables S4 and S6) using MEGA X (86). Multiple sequence alignments of mitochondrial genomes of the same haplogroup (A, B, C, or D) were further visually inspected and edited (i.e., removing common indels, as well as to confirm variants and making sure sequences remained at the same length), to assure an accurate alignment. Common hypervariable sites that are phylogenetically uninformative were excluded from the analysis (309, 315,

515-522 AC indels, 3107, 16182-16183, 16193, and 16519) (*32*). mtDNA sequences were excluded from the alignments based on two criteria: 1) > 200 "N's" as observed in fasta files or 2) Visual evaluation of an excess of C > T mutations that cannot be explained by random mutation. Four of the mtDNA sequences (333B_TOL_a, 333O_TOL_b, E8_CdV_b and, BC30) were excluded from EBSP analysis but not from haplotype network analysis (table S4), as their high number of N's, would have likely resulted in overestimation of the mutation rate and biased inferences in ESBP analyses.

Haplotype median-joining networks were constructed in Popart (*87*) after collapsing aligned full mtDNA sequences into haplotypes with DnaSP v. 6.12.03x64 (*74*). Haplotype networks were constructed with both ancient (from this study and previously published) (n = 25) and present-day (n = 78) mtDNA of individuals in the Mexican territory (tables S4 and S6). All present-day samples were retrieved from the PhyloTree (version 17) database (*32*, *76*). Only sequences that were reported as being of the same sub haplogroup or neighboring sub-haplogroup as our ancient samples were chosen for sequence comparison.

Past female effective population sizes were reconstructed using a Bayesian skyline approach (*88*) in BEAST 2 (*89*). The ancient mtDNA sequences, mentioned above were merged with present-day mtDNA sequences collected from published public databases (tables S4 and S6). Alignments were partitioned into five concatenated regions in the following order (Dloop, Coding, rRNA, and tRNA) as proposed by (*53*). HKY+G was found to be the best substitution model for this arrangement of the data using PartitionFinder 2 software (*90*). Substitution rates were estimated following a strict molecular clock model starting from point estimations as in (*91*). Additionally, we used tip calibrations for ancient samples using dating estimates (table S1). We ran Markov Chain Monte Carlo (MCMC) chains of 100 million steps for each haplogroup alignment with a sampling of parameters every 10,000 generations, discarding the first 10 million steps as burn-in. Extended Bayesian skyline analysis was plotted using SkyViz reported in (*54*). Two independent runs of EBSP were run, showing a similar behavior (see table S7 for the statistics of each run). Figure S6, A to N, corresponds to the first of the two EBSP runs.

Sex assignment

Determination of biological sex was made using the tool reported in (*92*). This approach computes the proportion of reads mapped to the Y chromosome with respect to the reads mapping to the X chromosome (Ry) and Y chromosomes. According to the method, Ry > 0.075 corresponds to XY, while Ry < 0.016 corresponds to XX (table S1).

Contamination estimation based on the X chromosome

To account for the contamination present in the nuclear DNA, we estimated the contamination with the software hapCon (*93*), which is based on detecting polymorphic sites on the X chromosome of male individuals. The estimation of contamination was performed only in three individuals meeting the inclusion criteria (XY assignment and > 0.02x on the X chromosome). Estimates with hapCon are reported in Table S1 and were lower (< 1%) than the ones estimated with the mitochondrial genome (1-2%) (table S2).

Y-chromosome haplogroup inference

Y chromosome genotype calling and haplogroup assignments were made as reported in (*94*). Genotype calling was performed sampling one random base at each site of the Y chromosome covered at least once, with ANGSD (-dohaplocall 1 -doCounts 1 -r chrY: -minMinor 0 -maxMis 4). Then, haplogroup assignments were made using the phylogenetic tree of Y chromosome single nucleotide polymorphisms constructed from the 1000 Genomes Project Phase 3, as in (*94*). The most derived haplogroup was assigned.

Autosomal reference panels

For analysis at a continental level, we constructed a reference panel with modern populations (Supplementary Table S5) that includes genetic information of: Yoruba (YRI), European genetic ancestry (CEU), and Chinese (CHB) from the 1000 Genomes Project Phase 3 (*37*, *95*), and available genome-wide information for Indigenous populations from Mexico previously published by the Simons Genome Diversity Project (SGDP) and the Human Genome Diversity Project (HGDP) (*18*, *22*). These genome-wide data were intersected using plink (*96*) with previously published reference panels of genotype information of Indigenous populations from Mexico (*36*), after masking non-Native American sites as in (*97*) and keeping individuals with > 85% genotype information (mind 0.15) and single-nucleotide variants (SNVs) with a minor allele frequency of 0.01 (maf 0.01), and with a 90% genotyping rate (geno 0.1) using the software plink (*96*). The final reference panel includes genotype information of 576,409 SNVs from 596 individuals from 4 continental populations (YRI, CEU, CHB, and Indigenous from Mexico) (table S8).

For the analysis within Mexico, we used the reference panel including only the Indigenous populations from Mexico (without the other three continental populations: YRI, CEU, and CHB) and keeping the individuals with >90% of Native American genetic ancestry. In total, this reference panel included genetic information of 561,327 SNVs from 268 individuals (table S5).

We also integrated our generated ancient genome-wide data with previously reported low-depth genomes (0.09 – 0.3 x) from three ancient Pericúes from Aridoamerica reported in (*14*, *15*), resulting in a total of 15 pre-Hispanic individuals from Mexico. Our analysis also included ancient genomic data from previous studies for 21 individuals from across the Americas (table S3) (*14*, *15*,*45*, *46*, *98*–*101*).

Pseudo-haploid calls

Pseudo-haploid calls of the ancient genomic data were made for the positions included in the reference panel. This was done by randomly sampling one read (when more than one read covered the site) at each site present in the reference panel and keeping the observed base at that site if it had a minimum base quality of 30. If only one read covered the site, the observed base was kept if it had a base quality above 30. Then, the calls were turned into homozygous genotypes. Similarly, for the genotype data in the reference panel, one allele was randomly sampled at heterozygous sites and turned into a homozygous genotype. For downstream analyses, we kept ancient individuals with 0.01x genome-wide coverage (table S1), all having > 4650 SNVs intersected with the reference panel.

Relatedness inference

Relatedness between ancient individuals from the same site was assessed using READ (*102*). A normalization step with a panel of non-related individuals is required previous to estimating the relationship in the ancient individuals, this step was performed with all Nahuas from the reference panel as they are the biggest group in the reference panel. Then, the normalization value was used to run READ with the ancient individuals. Two individuals from Toluquilla (333C_TOL_a and 333B_TOL_a) were identified as first-degree relatives. Thus, the one with the highest coverage (333B_TOL_a) was used for downstream analyses. Furthermore, individuals E4_CdV_b and E8_CdV_b were identified as second-degree relatives, as well as E7_CdV_b and E19_CdV_b.

ADMIXTURE analysis

For each dataset analyzed with ADMIXTURE, one hundred replicates were run with randomly generated seed values and calculating the cross-validation error with parameter -cv. The run with the best likelihood for each k was plotted using pong software (*103*). A first ADMIXTURE (*33*) analysis was carried out on the genotype reference panel including Indigenous individuals from Mexico and the three continental populations (YRI, CEU, and CHB)

and k=2 through k=9 to identify present-day Indigenous individuals with <90% Native American genetic ancestry. We found k=4 to be the k with the lowest mean cv error (fig. S8A), separating YRI, CEU, CHB, and Native American genetic ancestry. The run of k=4 with the best likelihood was chosen and the Indigenous individuals with <90% Native American genetic ancestry were removed from the panel. A second ADMIXTURE analysis was carried out with only the present-day Indigenous populations with >90% Native American genetic ancestry, from k=2 to k=8, being k=2 the one with the lowest cv error (fig. S8B) and separating populations with northern and southern ancestries. Then, a third ADMIXTURE run was carried out including the present-day Indigenous populations and pre-Hispanic individuals (Fig. 3).

Missing DNA PCA

Missing DNA PCA (mdPCA) is a principal component analysis that corrects for genotypes missing due to genetic ancestry masking or degradation in ancient DNA samples. The correction is performed by comparing genetic distances between all samples. The genetic distance is computed as the average number of pairwise differences (pi as defined in (*104*). The manual of the method and the description of the parameters can be found in the mdPCA GitHub repository https://github.com/AI-sandbox/mdPCA. The dataset included 15 ancient samples (12 from this study and 3 Pericúes previously published) and the reference panel including present-day Indigenous individuals from Mexico. Genotyping data in plink format was converted to an unphased VCF format with Plink version 1.90 beta (*96*). The method used this VCF file as an input with the following parameters: each individual was plotted as an average of both parental haplotypes (AVERAGE_PARENTS=True), weights for each individual were used inversely proportional to the number of samples from the corresponding population (IS_WEIGHTED=True), the simple weighted of covariance PCA without any optimization was used (METHOD=1), no genotype was masked based on local genetic ancestry calls (IS_MASKED=False), and only the three first PC's were calculated (NUM_DIMS=3). Individuals with <0.15x whole genome coverage (19,851 SNVs or less) were further down-weighted to 0.1x to avoid having the principal components be selected based on noise variance stemming from these samples. These included BC30, MOM6_ST_a, 11R_R_b, E4_CdV_b, E7_CdV_b, E19_CdV_b, E2_Mich_b, and E4_Mich_b.

Principal component analysis

We performed principal component analyses (PCA) using smartpca from the software eigensoft v6.0.1 (*34*) and projecting the individuals (lsqproject: YES) on the PCs estimated for the modern populations. The PCA (fig. S7) with only the ancient individuals as well as the present-day Indigenous populations from Mexico, only included present-day individuals masked and with >90% Native American genetic ancestry.

Temporal factor analysis

Temporal aactor analysis (TFA) (*38*) were run with the reference panel that includes only the Indigenous populations from Mexico. The ancient samples included in the TFA consisted exclusively of individuals from Mexican territory from this study with >1x genome-wide coverage: one individual from Sierra Tarahumara (F9_ST_b), and three from Toluquilla (2417Q_TOL_b, 2417J_TOL_a, and 333B_TOL_a). Genotyping data was manipulated with Plink version 1.90 beta (*96*). The dataset was converted from Plink format to Eigenstrat's geno format with the command convertf from Eigensoft version 6.0.1.1 (*34*). Missing genotypes were imputed with the snmf function from the LEA package, using the parameters K=2, entropy= TRUE, and repetitions=5. The runs with the lowest cross-entropy value were considered for posterior steps. Genotypes were corrected according to the autosomal coverage depth of each sample. We assigned the average depth from the HGDP dataset mentioned in the publication: 35X (*22*) (table S8). We assigned the respective coverage read depth for all 11 individuals from the Simons Genome Diversity Project (*18*) (table S8).Modern samples genotyped with microarrays (*36*) did not have a coverage depth value as they represent a different technology without sequencing reads, therefore, we assigned the maximum value from the dataset to all microarray samples, i.e., 56.19X. TFA was applied to these imputed and corrected genotypes with the following parameters: lambda = 5e-1, and K=2. Imputation, coverage correction, TFA, and the manipulation of the genotyping matrix in geno format were performed with R version 4.0.2. Variance explained by lambda did not change considerably across logarithmic values, thus we chose the default lambda value 5e-1. We plotted factor 1 and 2 as negative when required to match geography and facilitate the comparison between TFA plots (fig. S10).

Outgroup f3 and D statistics

To estimate each pre-Hispanic individual's genetic relatedness with a particular pre-Hispanic or present-day Indigenous population, we performed outgroup f3 statistics using ADMIXTOOLS v5.0 (*44*). Outgroup f3 was calculated in the form (Test, Source1; YRI), where the "Test" is the pre-Hispanic individual under analysis, and Source1 being another pre-Hispanic individual or present-day Indigenous population. We used all YRI from the 1000 Genomes Project (*37*, *95*) as an outgroup population. Assuming there was no admixture in the tree, the f3 value is proportional to the shared genetic drift between Test and Source1.

We performed D statistics using ADMIXTOOLS v5.0 (*44*) to identify whether a pre-Hispanic individual had higher genetic affinities with: i) other pre-Hispanic individuals from the same archaeological site than to pre-Hispanic individuals from other sites; or ii) a specific present-day indigenous population than to other present-day population. D values were computed in the form (Pop1, Pop2; Test, YRI). The "Test" refers to the pre-Hispanic individual under analysis, while Pop1 and Pop2 were all possible combinations between pre-Hispanic individuals and present-day Indigenous populations. Under the null hypothesis, Test individual would be equally related to Pop1 and Pop2, and we expect no significant deviations from D=0. Significant D<0 indicates closer relation between Test individual and Pop2. Moreover, a significant D>0 indicates closer relation between Test individual and Pop1. Only absolute values of Z-score > 3 were considered as significant (which corresponds to a p-value of ~ 0.0027).

Population continuity test between F9_ST_a and present-day Rarámuri

According to the PCA, TFA, ADMIXTURE, outgroup-f3 and D statistics, individual F9_ST_a from north Mexico is significantly closer to present-day Rarámuri. Thus, we analyzed whether individual F9 belonged to the present-day Rarámuri ancestral population using the method published in (*39*). Derived alleles in individual F9 and present-day Rarámuri were identified using ancestral alleles' information from (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/pilot_data/technical/reference/ancestral_alignments). Then, allele counts were performed for each site in the present-day Rarámuri (reference population), excluding alleles with frequency 0 or 1. We performed allele counts for the pre-Hispanic individual F9. Parameters alpha and beta necessary for the analysis were estimated using min_samples = 14, which refers to the minimum number of individuals required in the reference panel that have information for each SNV. Under the null hypothesis, it is assumed that reference population and ancient individual belong to a continuous population, rejection of the null hypothesis would mean no population continuity.

qpGraph

We used the qpGraph tool of the ADMIXTOOLS v5.0 software package (*44*) to construct admixture graphs and test which various models of demographic history best fit the data. We applied different models according to the

analysis: (i) For the model of population continuity, we used the 11,500-year-old USR1 (*45*) as the surrogate for ancient Beringian, 10,700 year-old Anzick (*46*) individual as the surrogate for ancestral Southern Native American (SNA) (*14*), and the 4200 year-old Ancient Southwestern Ontario individual (ASO) (*15*) as the surrogate for Northern Native American (NNA) (*14*). (ii) For the admixture model for Central Mexico populations, the base model included the pre-Hispanic populations, Anzick (*46*) individual as the ancestral SNA (*14*), present-day Athabascan as the representative for NNA (*14*), and present-day Indigenous populations in Mexico representing the northern (Konkaak), central (Nahua from Puebla), and southern (Triqui) regions. (iii) For UpopA we used the different models described in main text and (*17*). pre-Hispanic individuals were added to each model in various combinations to test the fit. The following parameters were applied: outpop (NULL), diag (0.0001), hires (YES), blgsize (0.05), and lsqmode (YES). A worst fitting f statistical test of z-score > 3 is set as a threshold for model rejection.

qpWave

We used qpWave version 410 (*44*) to test whether pre- and post-drought Sierra Gorda (SG) individuals could be modeled using either SG individuals or other ancient or modern Mexican indigenous population relative to a set of reference groups (Mbuti, Onge, USR1; Papuan, Anzick, Han, Mixe, MA1, Zapotec, Karitiana, Piapoco) as in (*105*). In order to maximize the power for this test we increased the number of SNVs to 2,224,359k transversion sites ascertained in 1KGP and having a 5% MAF. Present-day human sequencing data from HGDP genomes (*22*) was used to represent Mexican indigenous genetic variation. With these data, we tested whether we could model the genetic variation of pre (2417Q_TOL_b) and post-drought Sierra Gorda (2417J_TOL_a and 333B_TOL_a) individually, using as a source either another Sierra Gorda individual or another Mexican indigenous population. For the latter case we used as references the pre-Hispanic individual F9_ST_a, and Pericúes, along with present-day Akimel O'odham and Mayan. If pre- and post-drought Sierra Gorda individuals arise from different demographic histories to each other, then when analyzed together it would be required to add one or more different sources of genetic variation to explain the data, which would be denoted by a p-value < 0.05. We also performed qpWave analysis with the panel of Indigenous populations from Mexico reported in (*36*) to increase the number of populations, though at the cost of reducing the number of overlapping sites. Results are consistent with those obtained using the HGDP Akimel O'odham and Mayan in that post-drought genetic variation can be explained with pre-drought genetic variation, suggesting genetic continuity. However, using the Moreno-Estrada reference panel resulted in a low number of overlapping SNVs for the qpWave analysis and, consequently, some p-values could not be computed by the program (e.g., when using the pre and post-drought samples as both target (Pop1) and sources (Pop2) (table S18).

Conditional nucleotide diversity

CND, a relative measure of genetic diversity, was estimated by genetic pairwise comparison of the genomic positions that are variable in African individuals (*40*). For these comparisons we grouped the individuals per archeological site. This analysis included 16 pre-Hispanic individuals (Pericúes, Sierra Tarahuamara, Sierra Gorda, Cañada de la Virgen and Patagonia) with at least 10,000 SNVs overlapping between them (tables S11 and S12) and present-day populations from Mexico (*18*, *22*).

This consists of counting the differences between the ascertained alleles present in one pair of individuals from the same population. For this analysis, we used a panel of 1,938,919 transversions SNVs with a minor allele frequency of 0.01 in the YRI population. In all cases, pseudohaploid genotype calls were performed before the comparisons. CND values were calculated for all pairs of individuals for each present-day Indigenous population and for all pairs of pre-Hispanic individuals within the same site. For present-day populations we masked the sites with no Native American genetic ancestry to avoid counting levels of diversity from a different genetic ancestry. Mean and standard error values were estimated 100 times using a window size of 50 Kb even for pairs of ancient individuals. Results are reported in a violin plot, where each dot inside the violin corresponds to the CND value for a unique pair of individuals (fig. S19).

Runs of homozygosity

We used the tool hapROH (*41*) to detect runs of homozygosity in pre-Hispanic individuals. The distribution of ROHs was used as an indirect way of measuring genetic diversity and effective population size. This method detects the ROH's lengths longer than four centimorgans (cM) and up to 300 cM. The sum of the ROH detected gives insights into the relatedness of the parents of each individual and informs about population sizes. Length of ROH and relatedness between parents are positively correlated. While the higher the number of ROH segments, the smaller the population size and the smaller the number of heterozygote sites, which means less genetic diversity. hapROH uses a built-in set of 1240K SNVs from to perform pseudohaploid genotype calls and a reference panel of 5,008 present-day human haplotypes from (*37*). We performed a pseudohaploid variant calling for the 1240K SNVs for each pre-Hispanic genome. hapROH's manual recommends using samples with a minimum of 400K SNVs intersected with the 1240K SNVs. However, we also included the individual E8_CdV_b, which intersected 369,504 SNVs, and the Pericú individual B03, which intersected 370,026 SNVs, to have at least one individual of these sites in the analyses and considering the number of SNVs is not too far from 400K (table S13). Detection of runs of homozygosity was made using the set of 5,008 haplotypes mentioned above.

REFERENCES AND NOTES

1. A. L. Austin, L. L. Luján, *El pasado indígena* (Fondo de Cultura Económica, 2019).
2. B. B. Cornejo, *La frontera septentrional de Mesoamérica* (Instituto Nacional de Antropología e Historia, México, ed. 2, 2000), vol. 1 of *Historia Antigua de México*.
3. P. Kirchhoff, Mesoamerica: its geographic limits, ethnic composition and cultural characteristics. *Heritage of conquest*, pp. 17–30 (1952).
4. M. del C. Solanes, E. Vela, Atlas del México prehispánico. *Arqueol. Mex.* **5**, 5–18 (2000).
5. D. W. Stahle *et al*., Major Mesoamerican droughts of the past millennium. *Geophys. Res. Lett.* **38**, L05703 (2011). doi: 10.1029/2010GL046472
6. M. K. Hughes, H. F. Diaz, Was there a 'medieval warm period', and if so, where and when? *Clim. Change* **26**, 109–142 (1994). doi: 10.1007/BF01092410
7. P. Armillas, Condiciones ambientales y movimientos de pueblos en la frontera septentrional de Mesoamérica. *Homenaje a Fernando Márquez-Miranda*, 62–82 (1964).
8. K. H. Wogau, H. W. Arz, H. N. Böhnel, N. R. Nowaczyk, J. Park, High resolution paleoclimate and paleoenvironmental reconstruction in the Northern Mesoamerican Frontier for Prehistory to Historical times. *Quat. Sci. Rev.* **226**, 106001 (2019). doi: 10.1016/j.quascirev.2019.106001
9. D. J. Kennett *et al*., Development and disintegration of Maya political systems in response to climate change. *Science* **338**, 788–791 (2012). doi: 10.1126/science.1226299; pmid: 23139330
10. J. A. Hoggarth *et al*., The political collapse of Chichén Itzá in climatic and cultural context. *Global Planet. Change* **138**, 25–42 (2016). doi: 10.1016/j.gloplacha.2015.12.007
11. S. L. O'Hara, S. E. Metcalfe, F. A. Street-Perrott, On the arid margin: The relationship between climate, humans and the environment. A review of evidence from the highlands of central Mexico. *Chemosphere* **29**, 965–981 (1994). doi: 10.1016/0045-6535(94)90163-5; pmid: 7953466
12. B. Braniff, Oscilación de la frontera norte mesoamericana: un nuevo ensayo. *Arqueología, segunda época*, pp. 99–114 (1989).
13. A. Gómez Aíza, S. Sánchez Vázquez, La Frontera Cultural Meso-Aridoamericana: Construcción De Imaginarios Nacionalistas En La Historia Mexicana. *Xihmai* **9**, (2014). doi: 10.37646/xihmai.v9i18.244
14. M. Raghavan *et al*., Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* **349**, aab3884–aab3884 (2015). doi: 10.1126/science.aab3884; pmid: 26198033
15. C. L. Scheib *et al*., Ancient human parallel lineages within North America contributed to a coastal expansion. *Science* **360**, 1024–1027 (2018). doi: 10.1126/science.aar6851; pmid: 29853687
16. M. C. Ávila-Arcos, C. de la Fuente Castro, M. A. Nieves-Colón, M. Raghavan, Recommendations for Sustainable Ancient DNA Research in the Global South: Voices From a New Generation of Paleogenomicists. *Front. Genet.* **13**, 880170 (2022). doi: 10.3389/fgene.2022.880170; pmid: 35559028
17. Materials and methods are available as supplementary materials.
18. S. Mallick *et al*., The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016). doi: 10.1038/nature18964; pmid: 27654912

19. I. De La Cruz *et al*., Sex Identification of Children Sacrificed to the Ancient Aztec Rain Gods in Tlatelolco. *Curr. Anthropol.* **49**, 519–526 (2008). doi: 10.1086/587642
20. D. Perez-Benedico *et al*., Mayans: A Y chromosome perspective. *Eur. J. Hum. Genet.* **24**, 1352–1358 (2016). doi: 10.1038/ejhg.2016.18; pmid: 26956252
21. K. Sandoval *et al*., Y-chromosome diversity in Native Mexicans reveals continental transition of genetic structure in the Americas. *Am. J. Phys. Anthropol.* **148**, 395–405 (2012). doi: 10.1002/ajpa.22062; pmid: 22576278
22. A. Bergström *et al*., Insights into human genetic variation and population history from 929 diverse genomes. *Science* **367**, eaay5012 (2020). doi: 10.1126/science.aay5012; pmid: 32193295
23. B. Z. González-Sobrino *et al*., Genetic Diversity and Differentiation in Urban and Indigenous Populations of Mexico: Patterns of Mitochondrial DNA and Y-Chromosome Lineage. *Biodemography Soc. Biol.* **62**, 53–72 (2016).
24. H.-J. Bandelt *et al*., Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: Some caveats. *Ann. Hum. Genet.* **67**, 512–524 (2003). doi: 10.1046/j.1469-1809.2003.00049.x; pmid: 14641239
25. A. Y. Morales-Arce, M. H. Snow, J. H. Kelley, M. Anne Katzenberg, Ancient mitochondrial DNA and ancestry of Paquimé inhabitants, Casas Grandes (A.D. 1200-1450). *Am. J. Phys. Anthropol.* **163**, 616–626 (2017). doi: 10.1002/ajpa.23223; pmid: 28382719
26. J. Mata-Míguez, L. Overholtzer, E. Rodríguez-Alegría, B. M. Kemp, D. A. Bolnick, The genetic impact of Aztec imperialism: Ancient mitochondrial DNA evidence from Xaltocan, Mexico. *Am. J. Phys. Anthropol.* **149**, 504–516 (2012). doi: 10.1002/ajpa.22152; pmid: 23076995
27. A. Y. Morales-Arce *et al*., Ancient mitochondrial DNA and population dynamics in postclassic Central Mexico: Tlatelolco (ad 1325–1520) and Cholula (ad 900–1350). *Archaeol. Anthropol. Sci.* **11**, 3459–3475 (2019). doi: 10.1007/s12520-018-00771-7
28. A. González-Oliver, L. Márquez-Morfín, J. C. Jiménez, A. Torre-Blanco, Founding Amerindian mitochondrial DNA lineages in ancient Maya from Xcaret, Quintana Roo. *Am. J. Phys. Anthropol.* **116**, 230–235 (2001). doi: 10.1002/ajpa.1118; pmid: 11596002
29. A. Y. Morales-Arce *et al*., Successful reconstruction of whole mitochondrial genomes from ancient Central America and Mexico. *Sci. Rep.* **7**, 18100 (2017). doi: 10.1038/s41598-017-18356-0; pmid: 29273718
30. B. M. Kemp *et al*., Evaluating the Farming/Language Dispersal Hypothesis with genetic variation exhibited by populations in the Southwest and Mesoamerica. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 6759–6764 (2010). doi: 10.1073/pnas.0905753107; pmid: 20351276
31. M. Bodner *et al*., The Mitochondrial DNA Landscape of Modern Mexico. *Genes* **12**, 1453 (2021). doi: 10.3390/genes12091453; pmid: 34573435
32. M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, E386–E394 (2009). doi: 10.1002/humu.20921; pmid: 18853457
33. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009). doi: 10.1101/gr.094052.109; pmid: 19648217
34. N. Patterson, A. L. Price, D. Reich, Population structure and eigenanalysis. *PLOS Genet.* **2**, e190 (2006). doi: 10.1371/journal.pgen.0020190; pmid: 17194218
35. D. Reich, A. L. Price, N. Patterson, Principal component analysis of genetic data. *Nat. Genet.* **40**, 491–492 (2008). doi: 10.1038/ng0508-491; pmid: 18443580
36. A. Moreno-Estrada *et al*., The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science* **344**, 1280–1285 (2014). doi: 10.1126/science.1251688; pmid: 24926019
37. A. Auton *et al*., A global reference for human genetic variation. *Nature* **526**, 68–74 (2015). doi: 10.1038/nature15393; pmid: 26432245
38. O. François, F. Jay, Factor analysis of ancient population genomic samples. *Nat. Commun.* **11**, 4661 (2020). doi: 10.1038/s41467-020-18335-6; pmid: 32938925
39. J. G. Schraiber, Assessing the Relationship of Ancient and Modern Populations. *Genetics* **208**, 383–398 (2018). doi: 10.1534/genetics.117.300448; pmid: 29167200
40. P. Skoglund *et al*., Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science* **344**, 747–750 (2014). doi: 10.1126/science.1253448; pmid: 24762536
41. H. Ringbauer, J. Novembre, M. Steinrücken, Parental relatedness through time revealed by runs of homozygosity in ancient DNA. *Nat. Commun.* **12**, 5425 (2021). doi: 10.1038/s41467-021-25289-w; pmid: 34521843
42. D. I. Cruz Dávalos *et al*., Indigenous peoples in eastern Brazil: Insights from 19th century genomes and metagenomes. bioRxiv 448717 [Preprint] (2022) [cited 2023]. doi: 10.1101/2022.01.27.477466
43. M. C. Ávila-Arcos *et al*., Population History and Gene Divergence in Native Mexicans Inferred from 76 Human Exomes. *Mol. Biol. Evol.* **37**, 994–1006 (2020). doi: 10.1093/molbev/msz282; pmid: 31848607
44. N. Patterson *et al*., Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012). doi: 10.1534/genetics.112.145037; pmid: 22960212
45. J. V. Moreno-Mayar *et al*., Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature* **553**, 203–207 (2018). doi: 10.1038/nature25173; pmid: 29323294
46. M. Rasmussen *et al*., The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* **506**, 225–229 (2014). doi: 10.1038/nature13025; pmid: 24522598
47. J. V. Moreno-Mayar *et al*., Early human dispersals within the Americas. *Science* **362**, eaav2621 (2018). doi: 10.1126/science.aav2621; pmid: 30409807
48. H. García-Ortiz *et al*., The genomic landscape of Mexican Indigenous populations brings insights into the peopling of the Americas. *Nat. Commun.* **12**, 5942 (2021). doi: 10.1038/s41467-021-26188-w; pmid: 34642312
49. A. Gorostiza *et al*., Reconstructing the history of Mesoamerican populations through the study of the mitochondrial DNA control region. *PLOS ONE* **7**, e44666 (2012). doi: 10.1371/journal.pone.0044666; pmid: 23028577
50. S. Romero-Hidalgo *et al*., Demographic history and biologically relevant genetic variation of Native Mexicans inferred from whole-genome sequencing. *Nat. Commun.* **8**, 1005 (2017). doi: 10.1038/s41467-017-01194-z; pmid: 29044207
51. J. Ramos-Madrigal *et al*., Genome Sequence of a 5,310-Year-Old Maize Cob Provides Insights into the Early Stages of Maize Domestication. *Curr. Biol.* **26**, 3195–3201 (2016). doi: 10.1016/j.cub.2016.09.036; pmid: 27866890
52. M. A. Nieves-Colón *et al*., Ancient DNA Reconstructs the Genetic Legacies of Precontact Puerto Rico Communities. *Mol. Biol. Evol.* **37**, 611–626 (2020). doi: 10.1093/molbev/msz267; pmid: 31710665
53. B. Llamas *et al*., Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Sci. Adv.* **2**, e1501385 (2016). doi: 10.1126/sciadv.1501385; pmid: 27051878
54. B. D. O'Fallon, L. Fehren-Schmitz, Native Americans experienced a strong population bottleneck coincident with European contact. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20444–20448 (2011). doi: 10.1073/pnas.1112563108; pmid: 22143784
55. A. J. Herrera Muñoz, E. Mejía Campos, Un minero en la Sierra Gorda: caso de contaminación ocupacional multielemental de metales pesados a finales del periodo Clásico. *Arqueología (México, DF)*, 127–135 (2017).
56. E. Nalda, La arqueología de Guanajuato: Trabajos recientes. *Arqueol. Mex.* **16**, 36–43 (2008).
57. M. E. Smith, Mesoamerica's First World City: Teotihuacan in Comparative Context. *Teotihuacan: The World Beyond the City* (2020).
58. B. A. Álvarez-Sandoval, L. R. Manzanilla, M. González-Ruiz, A. Malgosa, R. Montiel, Genetic Evidence Supports the Multiethnic Character of Teopancazco, a Neighborhood Center of Teotihuacan, Mexico (AD 200-600). *PLOS ONE* **10**, e0132371 (2015). doi: 10.1371/journal.pone.0132371; pmid: 26200455
59. E. Campos, J. Campos, A. Muñoz, Mercurialism Determination in Fetuses Bone Remains from Toluquilla, Queretaro, Mexico. *J. Sci. Res. Rep.* **8**, 1–10 (2015). doi: 10.9734/JSRR/2015/18043
60. E. M. Pérez Campos, A. J. Herrera Muñoz, Minas y mineros: presencia de metales en sedimentos y restos humanos al sur de la Sierra Gorda de querétaro en México. *Chungara (Arica)* **45**, 161–176 (2013). doi: 10.4067/S0717-73562013000100008
61. D. A. C. Velàzquez, Expansión Territorial Comercial en Mesoamérica y Mesoamérica Septentrional por medio del Intercambio, *Geografia Ensino & Pesquisa* **19**, 59–68 (2015).
62. N. Rohland, M. Hofreiter, Ancient DNA extraction from bones and teeth. *Nat. Protoc.* **2**, 1756–1762 (2007). doi: 10.1038/nprot.2007.247; pmid: 17641642
63. J. Dabney *et al*., Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15758–15763 (2013). doi: 10.1073/pnas.1314445110; pmid: 24019490
64. M. Meyer, M. Kircher, Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, prot5448 (2010). doi: 10.1101/pdb.prot5448; pmid: 20516186
65. T. Daley, A. D. Smith, Predicting the molecular complexity of sequencing libraries. *Nat. Methods* **10**, 325–327 (2013). doi: 10.1038/nmeth.2375; pmid: 23435259
66. M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016). doi: 10.1186/s13104-016-1900-2; pmid: 26868221
67. H. Li *et al*., The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009). doi: 10.1093/bioinformatics/btp352; pmid: 19505943
68. A. McKenna *et al*., The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010). doi: 10.1101/gr.107524.110; pmid: 20644199
69. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013). doi: 10.1093/bioinformatics/btt193; pmid: 23613487
70. V. Link *et al*., ATLAS: Analysis Tools for Low-depth and Ancient Samples, bioRxiv 105346 [Preprint] (2017) [cited 2023]. doi: 10.1101/105346
71. G. Renaud, V. Slon, A. T. Duggan, J. Kelso, Schmutzi: Estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224 (2015). doi: 10.1186/s13059-015-0776-0; pmid: 26458810
72. H. Weissensteiner *et al*., HaploGrep 2: Mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58-63 (2016). doi: 10.1093/nar/gkw233; pmid: 27084951
73. T. A. Hall, BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT, *Nucleic Acids Symp Ser* **41**, 95–98 (1999).
74. J. Rozas *et al*., DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol. Biol. Evol.* **34**, 3299–3302 (2017). doi: 10.1093/molbev/msx248; pmid: 29029172
75. L. Excoffier, H. E. Lischer, Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–567 (2010). doi: 10.1111/j.1755-0998.2010.02847.x; pmid: 21565059
76. M. van Oven, PhyloTree Build 17: Growing the human mitochondrial DNA tree. *Forensic Sci. International. Genet. Suppl. Ser.* **5**, e392–e394 (2015). doi: 10.1016/j.fsigss.2015.09.155
77. S. Kumar *et al*., Large scale mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American origins. *BMC Evol. Biol.* **11**, 293 (2011). doi: 10.1186/1471-2148-11-293; pmid: 21978175
78. U. A. Perego *et al*., Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr. Biol.* **19**, 1–8 (2009). doi: 10.1016/j.cub.2008.11.058; pmid: 19135370
79. A. Achilli *et al*., Reconciling migration models to the Americas with the variation of North American native mitogenomes. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 14308–14313 (2013). doi: 10.1073/pnas.1306290110; pmid: 23940335
80. A. Achilli *et al*., Mitochondrial DNA backgrounds might modulate diabetes complications rather than T2DM as a whole. *PLOS ONE* **6**, e21029 (2011). doi: 10.1371/journal.pone.0021029; pmid: 21695278
81. U. A. Perego *et al*., The initial peopling of the Americas: A growing number of founding mitochondrial genomes from Beringia. *Genome Res.* **20**, 1174–1179 (2010). doi: 10.1101/gr.109231.110; pmid: 20587512
82. D. M. Behar *et al*., A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum.*

Genet. **90**, 675–684 (2012). doi: 10.1016/j.ajhg.2012.03.002; pmid: 22482806
83. A. Hartmann *et al*., Validation of microarray-based resequencing of 93 worldwide mitochondrial genomes. *Hum. Mutat.* **30**, 115–122 (2009). doi: 10.1002/humu.20816; pmid: 18623076
84. R. S. Just, T. M. Diegoli, J. L. Saunier, J. A. Irwin, T. J. Parsons, Complete mitochondrial genome sequences for 265 African American and U.S. "Hispanic" individuals. *Forensic Sci. Int. Genet.* **2**, e45–e48 (2008). doi: 10.1016/j.fsigen.2007.12.001; pmid: 19083815
85. M. Ingman, H. Kaessmann, S. Pääbo, U. Gyllensten, Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708–713 (2000). doi: 10.1038/35047064; pmid: 11130070
86. S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018). doi: 10.1093/molbev/msy096; pmid: 29722887
87. J. W. Leigh, D. Bryant, POPART: Full-feature software for haplotype network construction. *Methods Ecol. Evol.* **6**, 1110–1116 (2015). doi: 10.1111/2041-210X.12410
88. J. Heled, A. J. Drummond, Bayesian inference of population size history from multiple loci. *BMC Evol. Biol.* **8**, 289 (2008). doi: 10.1186/1471-2148-8-289; pmid: 18947398
89. R. Bouckaert *et al*., BEAST 2: A software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol.* **10**, e1003537 (2014). doi: 10.1371/journal.pcbi.1003537; pmid: 24722319
90. R. Lanfear, P. B. Frandsen, A. M. Wright, T. Senfeld, B. Calcott, PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Mol. Biol. Evol.* **34**, 772–773 (2017). doi: 10.1093/molbev/msw260; pmid: 28013191
91. P. Soares *et al*., Correcting for purifying selection: An improved human mitochondrial molecular clock. *Am. J. Hum. Genet.* **84**, 740–759 (2009). doi: 10.1016/j.ajhg.2009.05.001; pmid: 19500773
92. P. Skoglund, J. Storå, A. Götherström, M. Jakobsson, Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* **40**, 4477–4482 (2013). doi: 10.1016/j.jas.2013.07.004
93. Y. Huang, H. Ringbauer, hapCon: Estimating contamination of ancient genomes by copying from reference haplotypes. *Bioinformatics* **38**, 3768–3777 (2022). doi: 10.1093/bioinformatics/btac390; pmid: 35695771
94. The 1000 Genomes Project Consortium, D. G. Poznik, et al., Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nature Genetics* **48**, 593–599 (2016).
95. G. R. Abecasis *et al*., An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012). doi: 10.1038/nature11632; pmid: 23128226
96. S. Purcell *et al*., PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007). doi: 10.1086/519795; pmid: 17701901
97. B. K. Maples, S. Gravel, E. E. Kenny, C. D. Bustamante, RFMix: A discriminative modeling approach for rapid and robust local-ancestry inference. *Am. J. Hum. Genet.* **93**, 278–288 (2013). doi: 10.1016/j.ajhg.2013.06.020; pmid: 23910464
98. M. Rasmussen *et al*., Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762 (2010). doi: 10.1038/nature08835; pmid: 20148029
99. P. Flegontov *et al*., Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. *Nature* **570**, 236–240 (2019). doi: 10.1038/s41586-019-1251-y; pmid: 31168094
100. C. Posth *et al*., Reconstructing the Deep Population History of Central and South America. *Cell* **175**, 1185–1197.e22 (2018). doi: 10.1016/j.cell.2018.10.027; pmid: 30415837
101. C. de la Fuente *et al*., Genomic insights into the origin and diversification of late maritime hunter-gatherers from the Chilean Patagonia. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E4006–E4012 (2018). doi: 10.1073/pnas.1715688115; pmid: 29632188
102. J. M. Monroy Kuhn, M. Jakobsson, T. Günther, Estimating genetic kin relationships in prehistoric populations. *PLOS ONE* **13**, e0195491 (2018). doi: 10.1371/journal.pone.0195491; pmid: 29684051
103. A. A. Behr, K. Z. Liu, G. Liu-Fang, P. Nakka, S. Ramachandran, pong: Fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics* **32**, 2817–2823 (2016). doi: 10.1093/bioinformatics/btw327; pmid: 27283948
104. M. Nei, *Molecular Evolutionary Genetics* (Columbia University Press, New York, 1987).
105. K. Nägele *et al*., Genomic insights into the early peopling of the Caribbean. *Science* **369**, 456–460 (2020). doi: 10.1126/science.aba8697; pmid: 32499399

ACKNOWLEDGMENTS

We greatly acknowledge the IT support of L. A. A. Bautista, A. de León Cuevas, C. S. F. Bautista, and J. G. Sotelo of the Laboratorio Nacional de Visualización Científica Avanzada (LAVIS UNAM); A. C. Carbajal and C. U. Díaz from LIIGH-UNAM for wet-lab technical support during the project; A. Sockell for technical support at the Bustamante Lab; R. M. Duarte for his early contribution in the Cañada de la Virgen project; P. S. Fernández for help in the collection of samples from Cañada de la Virgen; L. Correa-Mendoza for technical support with qpWave; D. Ortega-Del Vecchyo and A. G. Oliver for enriching comments throughout the project on earlier versions of the manuscripts; and Y. J. Islas for her help in the summary figure. **Funding:** V.V.-I. is a doctoral student from the Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM), and has received the CONACyT fellowship (número de becaria 282188). M.C.A.-A. received support from UNAM-PAPIIT (project nos. IA201219 and IA203821), the Wellcome Trust Seed Award in science (grant no. 208934/Z/17/Z), and the International Centre for Genetic Engineering and Biotechnology (project no. CRP/MEX10-03). M.C.A.-A., E.H.-S., and F.J. received support from the Human Frontiers Science Program (grant no. RGY0075/201). R.F. was supported by a grant from Fondo Europeo de Desarrollo Regional (FEDER)/Ministerio de Ciencia e Innovación—Agencia Estatal de Investigación (PALEUCOL; PGC2018-094101-A-I00) and from CajaCanarias and La Caixa Banking foundations (GENPAC; 2018PATRI16). Data generation for the Sierra Tarahumara individuals was funded with a Trainee Research Grant to M.C.A.-A. by Stanford Center for Computational, Evolutionary, and Human Genomics. E.H.-S. received support from NIH R35GM128946 and the Alfred P. Sloan Foundation. Data generation for the Cañada de la Virgen individuals was supported by The Mexican Biobank Project (grants FONCICYT/50/2016 and MR/N028937/1 from CONACYT and Newton Fund, respectively) awarded to A.M.-E. This work has been partially supported by the government of Guanajuato (grant no. 093/2016 CINVESTAV) awarded to K.S. Data generation for the Aridoamerican individuals was supported in part by a Leakey Foundation grant to F.A.V. Additionally, M.L. received Swedish Research Council grant 2020-04789. **Author contributions:** M.C.A.-A. and V.V.-I. conceived and designed the study with input from E.M.P.C. M.S.-V., M.B.-L., B.M., R.F., K.S., and M.A.N.-C. performed the wetlab work. V.V.-I. processed all sequence data. V.V.-I. carried out most analyses with help from, A.I.-G., M.L., J.M.T., J.E.R.-R., D.A.V.-R., and F.S.-Q. E.M.P.C., A.H.-M., K.S., G.Z.G.M., E.G.-M., R.A.-H., and C.V. contributed samples, provided the archaeological context, and discussed results. F.A.V., A.M.-E., F.J., A.G.I., E.H.-S., J.V.M.-M., and F.S.-Q. suggested analyses, discussed results, and provided feedback throughout the project. V.V.-I. and M.C.A.-A. wrote and edited the manuscript with input from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data are available from the European Nucleotide Archive under project accession PRJEB51440. **License information:** Copyright © 2023 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.add6142
Supplementary Text
Figs. S1 to S33
References (*106*–*125*)

Submitted 25 June 2022; accepted 17 March 2023
10.1126/science.add6142

RESEARCH ARTICLE

RADIO ASTRONOMY

Magnetic field reversal in the turbulent environment around a repeating fast radio burst

Reshma Anna-Thomas[1,2]*†, Liam Connor[3,4]†, Shi Dai[5,6,7]†, Yi Feng[8]†, Sarah Burke-Spolaor[1,2]†, Paz Beniamini[9,10], Yuan-Pei Yang[11], Yong-Kun Zhang[6], Kshitij Aggarwal[1,2], Casey J. Law[3,4], Di Li[6,8,12,13]*, Chenhui Niu[6], Shami Chatterjee[14,15], Marilyn Cruces[16], Ran Duan[6], Miroslav D. Filipovic[5], George Hobbs[7], Ryan S. Lynch[17], Chenchen Miao[6], Jiarui Niu[6], Stella K. Ocker[14,15], Chao-Wei Tsai[6,12,18], Pei Wang[6], Mengyao Xue[6], Ju-Mei Yao[19], Wenfei Yu[20], Bing Zhang[21,22], Lei Zhang[6], Shiqiang Zhu[8], Weiwei Zhu[6,18]

Fast radio bursts (FRBs) are brief, intense flashes of radio waves from unidentified extragalactic sources. Polarized FRBs originate in highly magnetized environments. We report observations of the repeating FRB 20190520B spanning 17 months, which show that the FRB's Faraday rotation is highly variable and twice changes sign. The FRB also depolarizes below radio frequencies of about 1 to 3 gigahertz. We interpret these properties as being due to changes in the parallel component of the magnetic field integrated along the line of sight, including reversing direction of the field. This could result from propagation through a turbulent magnetized screen of plasma, located 10^{-5} to 100 parsecs from the FRB source. This is consistent with the bursts passing through the stellar wind of a binary companion of the FRB source.

Fast radio bursts (FRBs) are millisecond flashes of radio waves (*1*) from distant galaxies. Their emission mechanism, sources, and local environments are not well understood. The magnetization and density of astrophysical plasmas along the line of sight (LOS) between an FRB source and Earth modify the FRB's polarization properties and can provide constraints on the source's local environment. The observed properties of FRBs have provided insight into their potential origins, such as the identification of highly magnetized and variable plasma environments around FRBs (*2*–*4*), detection of FRB-like bursts from a nearby magnetar (*5*, *6*), and finding temporal structure in FRB bursts similar to that of radio emission from the Crab Pulsar (*7*).

The repeating FRB 20190520B has been localized to a dwarf galaxy at redshift $z = 0.241 \pm 0.001$ (*8*). FRB 20190520B has a sustained high repetition rate [at 1.2 GHz, above a fluence of 9.3 mJy ms (where mJy is the millijansky, equal to 10^{-29} W m^{-2} Hz^{-1})] of $R_{1.2\mathrm{GHz}} = 4.5^{+1.9}_{-1.5}$ hour^{-1}. It is colocated with a compact persistent radio source (PRS), and the host galaxy makes a large contribution to the electron column density along the LOS. The integrated electron column density $n_e(l)$ along the LOS distance l is quantified by the dispersion measure (DM), which is defined as $\mathrm{DM} \equiv \int n_e(l)\mathrm{d}l$. For FRB 20190520B, the DM is larger than the expected contributions from the foreground Milky Way and the intergalactic medium (IGM), which dominate the DMs of most other FRBs (*9*). The rest-frame contribution of the host galaxy to the DM of FRB 20190520B has been estimated as $\mathrm{DM_{host}} = 902^{+88}_{-128}$ pc cm^{-3} (*8*, *10*) (where pc is the parsec, equal to 3.08×10^{16} m). This is about an order of magnitude larger than is typical for FRB hosts, implying a large number of free electrons in the local environment of the FRB source (*11*).

FRB 20190520B has several similarities with the repeating FRB 20121102A (*12*, *13*): Both are located in dwarf galaxies with low abundances of heavy elements, and both are associated with a PRS. Faraday rotation is a change in the polarization direction of a linearly polarized radio wave as it propagates through a magnetized plasma (*14*). The amount of Faraday rotation integrated along the LOS is quantified by the rotation measure (RM), given by $\mathrm{RM} = \left[e^3/\left(2\pi m_e^2 c^4\right)\right]\int n_e(l)B_\parallel(l)\mathrm{d}l$, where $B_\parallel$ is the component of the magnetic field parallel to the LOS, e is the charge of an electron, c is the speed of light and m_e is the mass of an electron. FRB 20121102A has a high RM (*2*), but the RM of FRB 20190520B has not previously been determined.

Bursts observed from FRB 20190520B

We conducted polarimetric observations of FRB 20190520B (located at right ascension 16^h 02^m $04^s.266$, declination +11°17′17″.33, J2000 equinox) using the 100-m Robert C. Byrd Green Bank Telescope (GBT) and the 64-m Parkes Radio Telescope (also known as Murriyang). The GBT observations used the 1.1 to 1.8 GHz (L-band) receiver on 17 September 2020 and the 4 to 8 GHz (C-band) receiver on 14 September 2020 (hereafter epoch 1), 19, 23, 27, and 31 March 2021 (epoch 2), and 4 November 2021 (epoch 3), in coordinated universal time (UTC). The observations totaled 6.5 hours at L-band and ~20 hours at C-band (*15*). The Parkes observations were performed fortnightly from 5 April 2021 to 23 January 2022 by using the Ultra-Wideband Low (UWL) receiver, which covers 704 to 4032 MHz (*16*). We performed a threshold-based search for bursts in these data, including a search of frequency subbands to find spectrally narrow-band bursts (*15*). In total, we found nine bursts in the GBT L-band data, 16 bursts in the GBT C-band data, and 113 bursts in the Parkes data (between ~1600 and 4032 MHz), with each having a signal-to-noise ratio (S/N) ≥ 10. We performed polarimetric and flux calibration on each detected burst data segment (*15*).

The burst profiles and spectrograms of 12 bursts are shown in fig. S2, and the properties of all the 13 bursts with polarization detections are listed in table S2; DM and polarization upper limits for all other bursts are reported in table S3. We designate the Parkes bursts with polarization detections as P1 to P8, sorted by their modified Julian dates (MJDs), and the GBT C-band bursts with polarization detection as C1, C2.1, C2.2, C3.1, and C3.2, where the first number indicates the epoch of the GBT observation and the second indicates the order of the burst arrival time. The detected bursts are spectrally narrow banded, as has been seen for other repeating FRBs (*17*–*19*), and some pulses have a time-resolved substructure. The average detected burst bandwidths are 350 MHz in the GBT L-band, 850 MHz in the GBT C-band, and 750 MHz in the Parkes UWL data.

[1]Department of Physics and Astronomy, West Virginia University, Morgantown, WV 26506, USA. [2]Center for Gravitational Waves and Cosmology, West Virginia University, Morgantown, WV 26506, USA. [3]Cahill Center for Astronomy and Astrophysics, California Institute of Technology, Pasadena, CA 91125, USA. [4]Owens Valley Radio Observatory, California Institute of Technology, Big Pine, CA 93513, USA. [5]School of Science, Western Sydney University, Locked Bag 1797, Penrith, NSW 2751, Australia. [6]National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China. [7]Australia Telescope National Facility, Commonwealth Scientific and Industrial Research Organisation–Space and Astronomy, Epping, NSW 1710, Australia. [8]Zhejiang Lab, Hangzhou, Zhejiang 311121, China. [9]Department of Natural Sciences, Open University of Israel, Ra'anana 43107, Israel. [10]Astrophysics Research Center of the Open University, The Open University of Israel, Ra'anana 43537, Israel. [11]South-Western Institute for Astronomy Research, Yunnan University, Kunming 650500, Yunnan, China. [12]University of Chinese Academy of Sciences, Beijing 100049, China. [13]National Astronomical Observatories, Chinese Academy of Sciences–University of KwaZulu-Natal Computational Astrophysics Centre, University of KwaZulu-Natal, Durban 4000, South Africa. [14]Department of Astronomy, Cornell University, Ithaca, NY 14853, USA. [15]Cornell Center for Astrophysics and Planetary Science, Cornell University, Ithaca, NY 14853, USA. [16]Max-Planck Institute for Radio Astronomy, D-53121 Bonn, Germany. [17]Green Bank Observatory, Green Bank, WV 24401, USA. [18]Institute for Frontiers in Astronomy and Astrophysics, Beijing Normal University, Beijing 102206, China. [19]Xinjiang Astronomical Observatory, Chinese Academy of Sciences, Urumqi, Xinjiang 830011, China. [20]Shanghai Astronomical Observatory, Chinese Academy of Sciences, 80 Nandan Road, Shanghai 200030, China. [21]Nevada Center for Astrophysics, Las Vegas, NV 89154, USA. [22]Department of Physics and Astronomy, University of Nevada, Las Vegas, NV 89154, USA.
*Corresponding author. Email: rat0022@mix.wvu.edu (R.A.-T.); dili@nao.cas.cn (D.L.) †These authors contributed equally to this work.

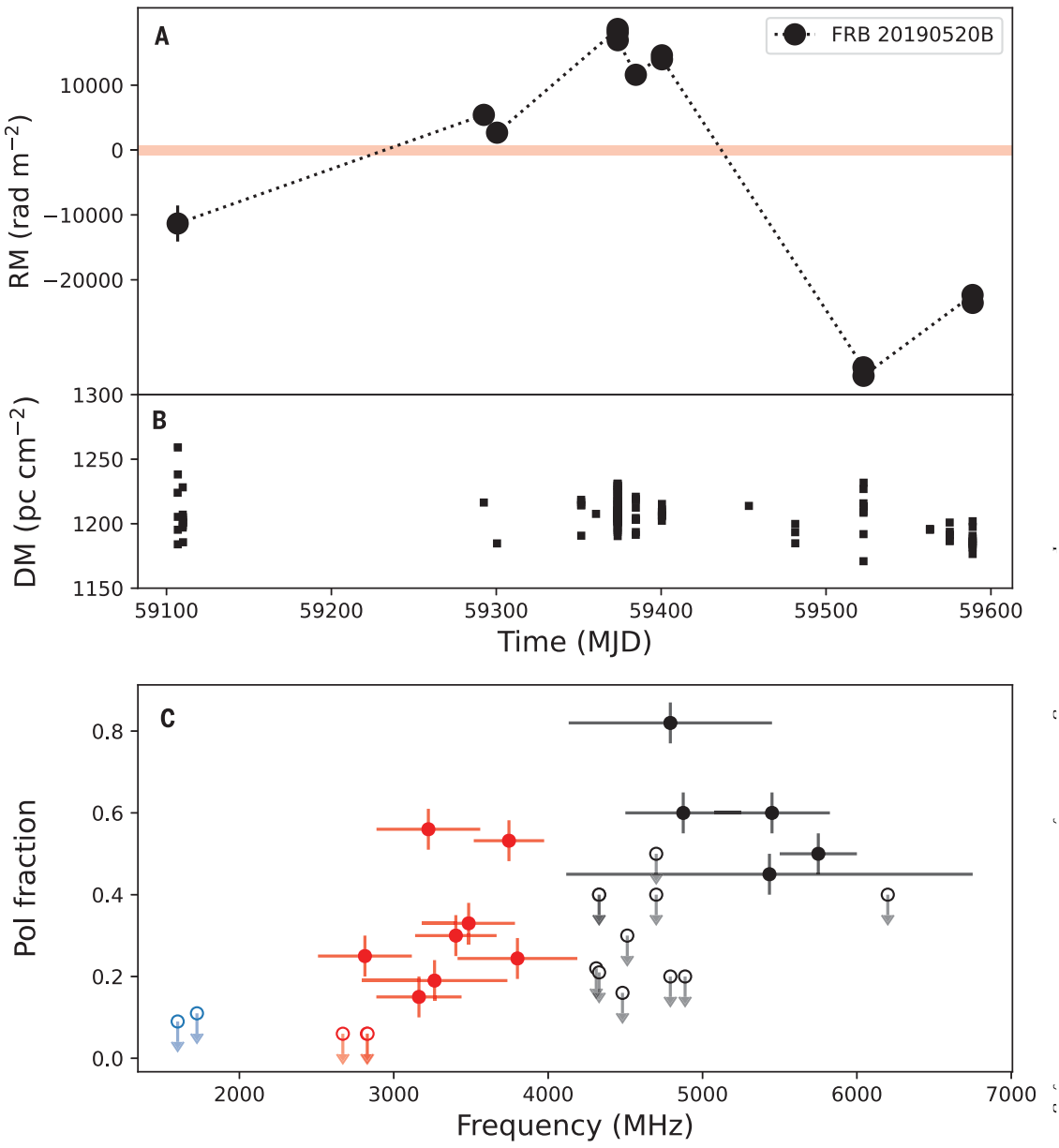


Fig. 1. Observed time- and frequency-dependent properties of FRB 20190520B. (**A**) The rest-frame RM measured for FRB 20190520B. The pink shaded region indicates 0 rad m^{-2}. (**B**) The DM variation of FRB 20190520B over a 17-month period. In (A) and (B), error bars indicate 1σ uncertainties but are mostly smaller than the symbols. (**C**) The polarization fraction of FRB 20190520B as a function of observed frequency, with points indicating the GBT L-band (blue), GBT C-band (black), and Parkes UWL (red) data. Solid symbols are detections, and open symbols with downward-pointing arrows are 3σ upper limits. Only the most constraining upper limits are shown at each frequency for clarity. Numeric values for all bursts are listed in table S3.

We analyzed the Stokes parameters of the radio waves to determine the linear polarization fractions and RMs of each burst (*15*). We found that the polarization of FRB 20190520B is time- and frequency-dependent and that the RM is also variable (Fig. 1). The RM detections, which we determined by maximizing the linear polarization fraction (*15*), are shown in Fig. 2. After removing the Faraday rotation, five GBT C-band bursts and eight UWL bursts have fractional linear polarization (L/I) (where L is the linear polarization and I is the total intensity) in the range of 15 to 80%. Only one burst (C2.1) had detectable fractional circular polarization, with V/I (where V is the circular polarization) of −42 ± 7%. No L-band bursts had detectable polarization, with the brightest burst having an upper limit of $L/I \leq 9\%$. We estimate that the L-band data had sufficient frequency resolution to recover RMs up to 10^5 rad m^{-2}, so this nondetection is not due to instrumental depolarization (*15*). The spectrogram and the profile of three bursts that demonstrate depolarization toward low frequency are shown in Fig. 3. The low linear polarization in the L-band is therefore an intrinsic property of the FRB, as is the variable polarization in the C-band and across the UWL band. These properties could be due to the emission mechanism at the source, propagation of the bursts through an intervening plasma screen, or a combination of the two effects.

For the 13 bursts with detected polarization, we measured RMs ranging from -2.4×10^4 rad m^{-2} in the observer's frame (corresponding to -3.6×10^4 rad m^{-2} in the source frame) to $+1.3 \times 10^4$ rad m^{-2} in the observer's frame ($+2.0 \times 10^4$ rad m^{-2} in the source frame). The foreground Milky Way contribution to the RM in this direction is expected to be small, −9.84 ± 12.64 rad m^{-2} (*20*). We observed RM variability on week-long timescales of around ±300 rad m^{-2} day^{-1} and similar average variability on 6-month timescales. The largest inter-epoch rate of RM change, $\mathrm{dRM/d}t = -444$ rad m^{-2} day^{-1}, occurred between 8 June and 19 June 2021.

Comparison with other sources

The range of RM we measured for FRB 20190520B spans—in a single target—the full range of negative and positive RMs observed for all pulsars in the Milky Way (Fig. 4). Although the maximum absolute RM value for FRB 20190520B is three times lower than that of FRB 20121102A, their peak-to-peak variations are of similar size. The fractional change in RM for FRB 20190520B is much larger, but this implies

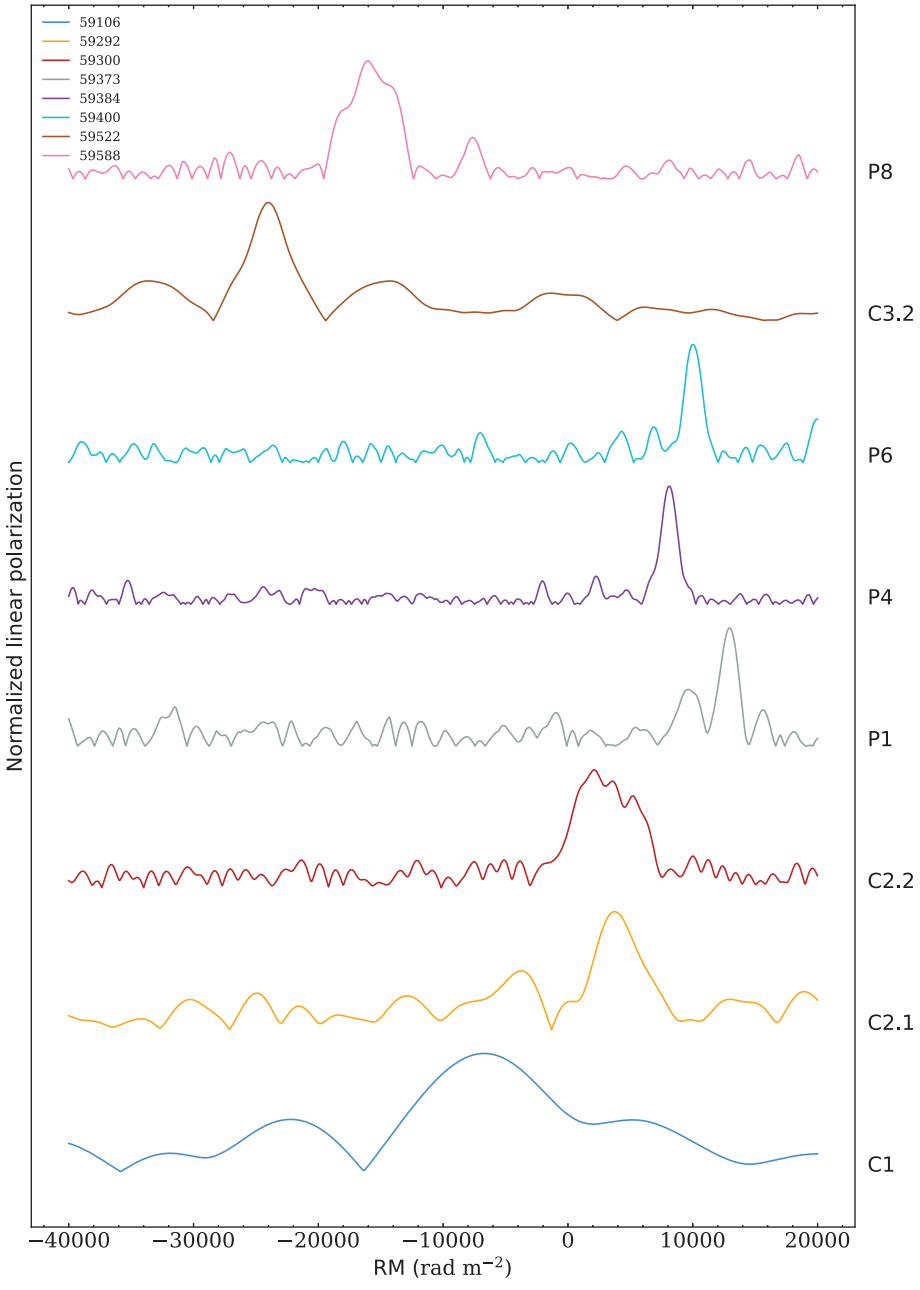


Fig. 2. Variability of RM in bursts from FRB 20190520B. The normalized linear polarization fraction is shown as a function of RM, for eight bursts. The maximum value of linear polarization determines the RM at the time of each burst. Each trace shows the measurement from a different burst (on MJDs listed in the legend), sorted by date, with arbitrary offsets. Only one burst per day is shown, for clarity. The burst notation on the right axis is discussed in the text.

that both FRBs are located in magnetized plasma environments with similar properties. However, the RM of FRB 20190520B crosses zero, requiring an additional environmental change to allow sign reversals. The frequency-dependent polarization fraction change for FRB 20190520B indicates multipath depolarization, whereas the fractional RM change and the sign reversals require changes in the LOS magnetic field—either because of rapid reorientation of a bulk material or integrated changes due to turbulence.

The Galactic source with the most similar RM variability to FRB 20190520B is PSR B1259–63 (Fig. 4), a pulsar that is periodically eclipsed by its binary star companion (*21*). Both FRB 20190520B and PSR B1259–63 have RM variations that cross zero. This indicates a reversal in the integrated LOS magnetic field orientation, in addition to variation in the electron density or magnetic field strength. The time-dependent variations of RM (in both sign and magnitude) in PSR B1259–63 have been attributed to passage of the radio pulses through a clumpy decretion disk around its binary star companion during the closest part of their orbit, causing depolarization (*22*). The pulsar's DM rose by around 20 pc cm^{-3} over a few days just before the eclipse part of its orbit.

We also compared FRB 20190520B with other FRBs, pulsars, and a magnetar that exhibit RM variations (supplementary text).

Dispersion measure

Whatever the cause of the RM variations, if $\int_l n_e dl$ along the LOS contributes to the RM magnitude changes for FRB 20190520B, there should be an accompanying DM change. We therefore sought to quantify the DM vaiability, although this is complicated by the "sad trombone" effect (*23*), in which FRB pulse structures sweep downward in frequency as a function of time, as is visible for some bursts of FRB 20190520B shown in fig. S3. We therefore inspected the brightest pulses (detection S/N > 20) using a DM measurement technique that maximizes the structure in the burst (*15*), in addition to visual inspection of pulses with substructures. We measured a ΔDM range of 10 to 15 pc cm^{-3}, although the associated uncertainties prevent us from determining whether these vary in a systematic way (*15*). Assuming that the changes in DM and RM arise from the same medium, we calculated the average LOS magnetic field as $B_{\parallel} \simeq 1.2 \left[\frac{\Delta \mathrm{RM}/(10^4 \text{ rad m}^{-2})}{\Delta \mathrm{DM}/(10 \text{ pc cm}^{-3})}\right]$ mG $\simeq$ 3 to 6 mG. This calculation assumes a fully turbulent medium, for which the ΔDM of a plasma screen is approximately equal to the average DM contributed by that screen. For comparison, for FRB 20121102A the lower limit range on the average magnetic field is 0.6 to 2.4 mG (*2*).

Interpretation of the magnetized environment

We interpret FRB 20190520B's properties as being due to propagation of the radio waves through a dense, turbulent magnetized region, as discussed in previous work (*24*). We consider two alternative scenarios—large-scale magnetic reorientations of a discrete object along the LOS, or an origin in a larger region of the host galaxy (such as the LOS passing through a spiral arm)—to be implausible because of the large variations and rapid timescales involved. We do not explore potential explanations intrinsic to the FRB emission mechanism; although those are possible, the observed properties can be produced by a propagation model, which we regard as a simpler explanation (supplementary text). In our preferred scenario, the polarization, RM, and DM are imparted by the burst passing through a region dominated by bulk magnetization, with that region being made up of subeddies or filamentary structures that cause time variability in the integrated LOS magnetic field or electron density.

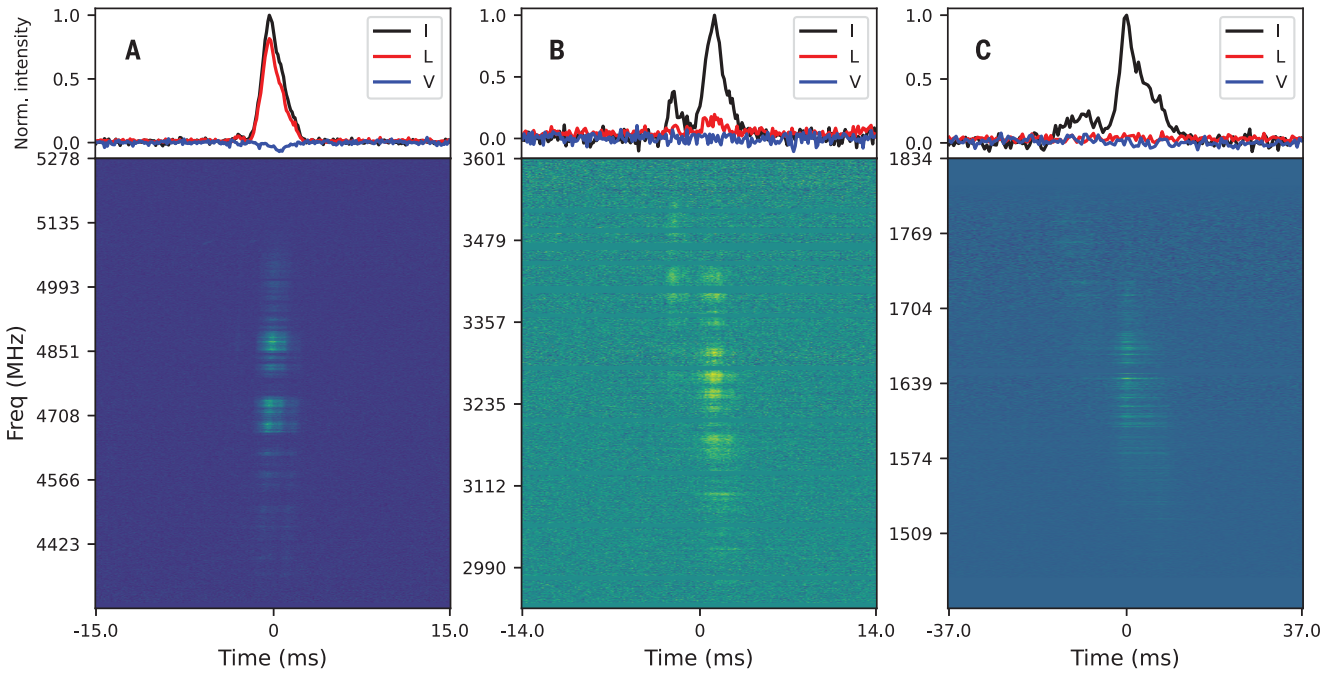
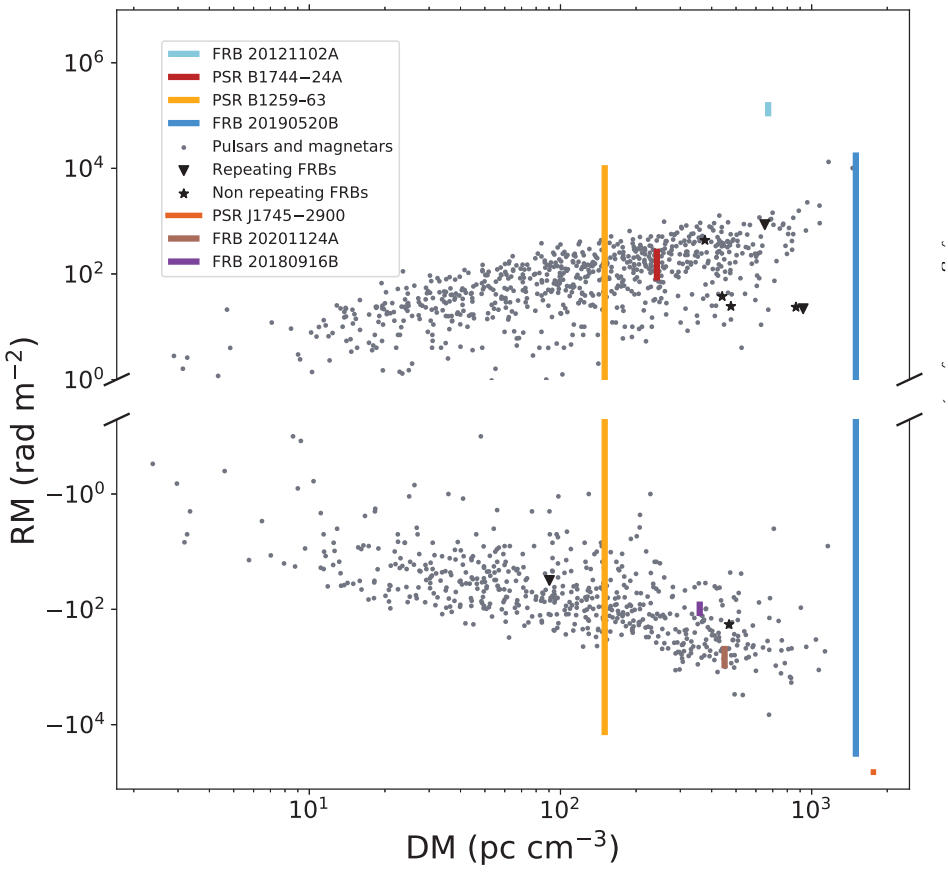


Fig. 3. Depolarization toward low frequencies. Spectrograms and normalized polarization profiles are shown for three bursts observed at different frequencies. (**A**) The GBT C-Band (4.8 GHz) burst that has the highest linear polarization percentage we observed. (**B**) An example Parkes UWL (3.2 GHz) burst, which shows a small linear polarization fraction. (**C**) An example GBT L-Band (1.6 GHz) burst, which shows no detectable linear polarization. In (A) to (C), the black, red, and blue lines indicate the Stokes I, linear polarization L, and Stokes V measurements, respectively.

Fig. 4. Comparison of FRB 20190520B to the RMs and DMs of pulsars, magnetars, and FRBs. Thick vertical lines are sources that show RM variations and extend across the full range of RMs observed for each. Gray circles are pulsars and magnetars (*32*), and stars and downward triangles are non-repeating and repeating FRBs, respectively (*33*, *34*), without RM variations. FRB 20190520B (blue bar) and PSR B1259–63 (yellow bar) have larger fractional variability than that of other FRBs or pulsars, and both cross zero RM. Data sources for specific targets are FRB 20121102A, (*35*); PSR B1744–24A, (*27*); PSR B1259–63, (*22*, *26*, *36*); FRB 20190520B, this work; PSR J1745–2900, (*37*); FRB 20201124A, (*3*); and FRB 20180916B, (*4*).



Multipath depolarization and large RM variability can be produced by turbulent dynamic and dense magneto-ionic plasma environments (*24*). However, the magnitude of those effects—and therefore whether they can be detected in observations—depends on the physical parameters of the plasma screen. The observable effects are determined by the characteristic size of turbulent eddy regions; the screen depth; and the relative velocities of the FRB source, bulk screen, and internal eddies. Using a simple model for propagation through such a turbulent plasma screen (*15*), our measurements of FRB 20190520B constrain the screen to locations between 10^{-5} and 100 pc from the FRB source, and the screen's free electron density to between 5×10^6 and 2 cm^{-3}, respectively. These results would cover a broader range if the LOS DM variation is substantially lower than we infer (*15*). A previous model of FRB 20121102A (*2*), assuming electron temperature $T_e = 10^6$ K, found that its foreground plasma screen has an electron density $n_e = 10^2$ cm^{-3} and thickness $L_{RM} = 1$ pc, which is similar to the upper limit on the size of the associated PRS. Higher electron temperatures could increase the thickness to $L_{RM} = 100$ pc (*2*).

We considered several scenarios that could explain such environmental turbulence local to an FRB source. Informed by the similar properties of PSR B1259–63, we explored a binary model in which the LOS to FRB 20190520B passes close to the surface of a companion star, so the radio propagation is affected by the magnetization and turbulence of its stellar wind. In this scenario, the star provides the bulk magnetization that causes the broad range in RM, while the turbulent stellar wind causes the rapid RM swings (*15*). The distance to the plasma screen is then approximately the separation between the two objects. Taking $d \approx 10^{-5}$ pc, the required electron density is $n_e \approx 5 \times 10^6$ cm^{-3}, which could be provided by a stellar mass loss rate of $\dot{M} \approx 3 \times 10^{-8} M_\odot$ year^{-1} (where $M_\odot$ is the solar mass, equal to 2×10^{30} kg) and wind velocity $v_w \sim 10^3$ km s^{-1} (*15*). These values are consistent with those observed for massive stars (*25*). As a check, we applied the same model to the PSR B1259–63 system (supplementary text) and derived properties that are consistent with previous studies (*21*).

If the binary wind scenario is correct, we expect there to be an underlying periodicity to the observed properties (a rise and fall of the RM variability envelope, DM variability range, and average DM), equal to the orbital period. During times when the LOS to the FRB source does not pass close to the star, we expect the RM value to stop varying and settle at an ambient value, reflecting the bulk magnetization of the host galaxy. This effect has been observed for the pulsars PSR B1259–63 and PSR B1744–24A (*26*, *27*) and other FRBs with RM variability (*3*, *4*).

We also considered alternative models of a shocked neutron star wind, or an FRB source close to an intermediate mass black hole, but we regard both scenarios as unlikely (supplementary text).

Implications for other sources

Although the RM variations and host DM of FRB 20190520B are unusual, they do not necessarily indicate that it is an unusual FRB source. FRB 20190520B appears to have a burst rate and burst substructures that vary with time, which is consistent with other repeating FRBs, and has similar energy scales to those of other FRBs (*8*, *28*). It is possible that repeating FRBs have a common source type but vary in local conditions (such as in binary orbital period, eccentricity, phase, or inclination).

Previous studies have proposed a connection between the emission cycles seen in some repeating FRBs and a binary orbit (*29*–*31*). If our model is correct, the observed properties of the FRB constrain the binary configuration and some properties (n_e, magnetic field strength) of the stellar wind.

REFERENCES AND NOTES

1. D. R. Lorimer, M. Bailes, M. A. McLaughlin, D. J. Narkevic, F. Crawford, *Science* **318**, 777–780 (2007).
2. D. Michilli *et al*., *Nature* **553**, 182–185 (2018).
3. H. Xu *et al*., *Nature* **609**, 685–688 (2022).
4. R. Mckinven *et al*., A large scale magneto-ionic fluctuation in the local environment of periodic fast radio burst source, FRB 20180916B. arXiv:2205.09221 [astro-ph.HE] (2022).
5. CHIME/FRB Collaboration, *Nature* **587**, 54–58 (2020).
6. C. D. Bochenek *et al*., *Nature* **587**, 59–62 (2020).
7. K. Nimmo *et al*., *Nat. Astron.* **6**, 393–401 (2022).
8. C. H. Niu *et al*., *Nature* **606**, 873–877 (2022).
9. J. P. Macquart *et al*., *Nature* **581**, 391–395 (2020).
10. S. Koch Ocker *et al*., *Astrophys. J.* **931**, 87 (2022).
11. J. P. Macquart *et al*., *Advancing Astrophysics with the Square Kilometre Array (AASKA14)* (Sissa Medialab, 2015), p. 55.
12. S. Chatterjee *et al*., *Nature* **541**, 58–61 (2017).
13. S. P. Tendulkar *et al*., *Astrophys. J. Lett.* **834**, L7 (2017).
14. M. Faraday, *Philos. Trans. R. Soc. London Ser. I* **136**, 1 (1846).
15. Materials and methods are available as supplementary materials.
16. G. Hobbs *et al*., *Publ. Astron. Soc. Aust.* **37**, e012 (2020).
17. P. Kumar *et al*., *Mon. Not. R. Astron. Soc.* **500**, 2525–2531 (2021).
18. V. Gajjar *et al*., *Astrophys. J.* **863**, 2 (2018).
19. Z. Pleunis *et al*., *Astrophys. J.* **923**, 1 (2021).
20. N. Oppermann *et al*., *Astron. Astrophys.* **542**, A93 (2012).
21. S. Johnston *et al*., *Astrophys. J.* **387**, L37 (1992).
22. S. Johnston *et al*., *Mon. Not. R. Astron. Soc.* **279**, 1026–1036 (1996).
23. J. W. T. Hessels *et al*., *Astrophys. J. Lett.* **876**, L23 (2019).
24. P. Beniamini, P. Kumar, R. Narayan, *Mon. Not. R. Astron. Soc.* **510**, 4654–4668 (2022).
25. N. Langer, *Annu. Rev. Astron. Astrophys.* **50**, 107–164 (2012).
26. S. Johnston, L. Ball, N. Wang, R. N. Manchester, *Mon. Not. R. Astron. Soc.* **358**, 1069–1075 (2005).
27. D. Li, A. Bilous, S. Ransom, R. Main, Y.-P. Yang, A highly variable magnetized environment in a pulsar binary resembling fast radio bursts. arXiv:2205.07917 [astro-ph.HE] (2022).
28. D. Li *et al*., *Nature* **598**, 267–271 (2021).
29. M. V. Barkov, S. B. Popov, *Mon. Not. R. Astron. Soc.* **515**, 4217–4228 (2022).
30. K. Ioka, B. Zhang, *Astrophys. J. Lett.* **893**, L26 (2020).
31. S. P. Tendulkar *et al*., *Astrophys. J. Lett.* **908**, L12 (2021).
32. G. Hobbs, R. Manchester, A. Teoh, M. Hobbs, *Young Neutron Stars and Their Environments*, F. Camilo, B. M. Gaensler, Eds. (International Astronomical Union, 2004), vol. 218, p. 139; http://www.atnf.csiro.au/research/pulsar/psrcat.
33. E. Petroff *et al*., *Publ. Astron. Soc. Aust.* **33**, e045 (2016).
34. CHIME/FRB Collaboration, *Astrophys. J.* **257** (suppl.), 59 (2021).
35. A. Plavin *et al*., *Mon. Not. R. Astron. Soc.* **511**, 6033–6041 (2022).
36. T. W. Connors, S. Johnston, R. N. Manchester, D. McConnell, *Mon. Not. R. Astron. Soc.* **336**, 1201–1208 (2002).
37. G. Desvignes *et al*., *Astrophys. J. Lett.* **852**, L12 (2018).
38. D. Li, Extreme magneto-ionic environment of actively repeating FRB 20190520B, version 1 (Science Data Bank, 2022); https://doi.org/10.11922/sciencedb.o00069.00007.
39. R. Anna-Thomas, L. Connor, S. Burke-Spolaor, Magnetic field reversal in the turbulent environment around a repeating fast radio burst. *Zenodo* (2022); doi:10.5281/zenodo.7339930.
40. R. A. Thomas, ReshmaAnnaThomas/FRB20190520B: FRB 20190520B RM variations. Code to reproduce GBT results (2023); doi:10.5281/zenodo.7700849.
41. Y. K. Zhang, SukiYume/RMS: Rotation Measure Fitting (v1.0), *Zenodo* (2023).

ACKNOWLEDGMENTS

The Parkes (Murriyang) Radio Telescope is part of the Australia Telescope National Facility, which is funded by the Commonwealth of Australia for operation as a National Facility managed by CSIRO. The Green Bank Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities. **Funding:** D.L. and Y.F. are supported by NSFC grants 11988101, 12203045, and 11725313; by the National Key R&D Program of China 2017YFA0402600; and by Key Research Project of Zhejiang Lab 2021PE0AC03. R.A.-T., S.B.-S., and K.A. acknowledge support from NSF grant AAG-1714897. S.D. is the recipient of an Australian Research Council Discovery Early Career Award (DE210101738) funded by the Australian Government. Y.-P.Y. is supported by NSFC grant 12003028. C.J.L. acknowledges support from NSF grant 2022546. P.B. was supported by a grant (2020747) from the United States–Israel Binational Science Foundation (BSF), Jerusalem, Israel. S.B.-S. is a CIFAR Azrieli Global Scholar in the Gravity and the Extreme Universe program. W.Z. is supported by National SKA Program of China 2020SKA0120200 and the NSFC 12041303 and 11873067. P.W. is supported by NSFC grant U2031117, the Youth Innovation Promotion Association CAS (2021055), the CAS Project for Young Scientists in Basic Reasearch (grant YSBR-006), and the Cultivation Project for FAST Scientific Payoff and Research Achievement of CAMS-CAS. J.-M.Y. is supported by the National Science Foundation of Xinjiang Uygur Autonomous Region (grant 2022D01D85) and the CAS Project for Young Scientists in Basic Research (grant YSBR-063). L.Z. is supported by an ACAMAR Postdoctoral Fellowship and the National Natural Science Foundation of China (grant 12103069). M.C., D.L., and W.Z. acknowledge support from the CAS-MPG LEGACY project. C.-W.T. is supported by NSFC grant 12041302. W.Y. is supported by NSFC grant U1838203. The GBT Epoch 2 observations were carried out through a time procurement agreement funded by a NSFC grant 11988101. **Author contributions:** R.A.-T., S.B.-S., C.J.L., Y.F., W.Y., and D.L. implemented the GBT observation campaign. S.D. and D.L. implemented the Parkes (Murriyang) observation campaign. R.A.-T., K.A., R.S.L., and S.B.-S. searched the GBT data for bursts; R.A.-T., L.C., and K.A. carried out analysis of their properties. S.D. searched the Parkes data for bursts and analyzed the burst properties. R.A.-T., L.C., Y.F., and Y.-K.Z. conducted the polarization analysis and visualization. R.A.-T., S.B.-S., L.C., D.L., Y.F., and S.D. led the interpretation of the results and writing of the manuscript. P.B., S.B.-S., L.C., and R.A.-T. produced the turblence model. Y.-P.Y. and B.Z. contributed theoretical investigation of the physical implications. All authors contributed to the analysis or interpretation of the data and to the final version of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The data for each of the Parkes bursts are available at Science Data Bank (*38*), and data for the GBT bursts are available at Zenodo (*39*). The software and notebooks we used for the polarization analysis are available at https://github.com/ReshmaAnnaThomas/FRB20190520B and https://github.com/SukiYume/RMS and archived at Zenodo (*40*, *41*). Our derived properties of the bursts are listed in table S2 for the polarized bursts and in table S3 for bursts with no detected polarization.

License information: Copyright © 2023 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.abo6526
Materials and Methods
Supplementary Text
Figs. S1 to S3
Tables S1 to S3
References (*42*–*81*)

Submitted 18 February 2022; resubmitted 25 August 2022
Accepted 13 March 2023
10.1126/science.abo6526

PALEOECOLOGY

Human adaptation to diverse biomes over the past 3 million years

Elke Zeller[1,2]*, Axel Timmermann[1,3], Kyung-Sook Yun[1,3], Pasquale Raia[4], Karl Stein[1,3], Jiaoyang Ruan[1,3]

To investigate the role of vegetation and ecosystem diversity on hominin adaptation and migration, we identify past human habitat preferences over time using a transient 3-million-year earth system-biome model simulation and an extensive hominin fossil and archaeological database. Our analysis shows that early African hominins predominantly lived in open environments such as grassland and dry shrubland. Migrating into Eurasia, hominins adapted to a broader range of biomes over time. By linking the location and age of hominin sites with corresponding simulated regional biomes, we also find that our ancestors actively selected for spatially diverse environments. The quantitative results lead to a new diversity hypothesis: *Homo* species, in particular *Homo sapiens*, were specially equipped to adapt to landscape mosaics.

Homo sapiens is the only surviving hominin species today, and it is largely unknown whether this is because our species was uniquely successful at adapting to the strongly fluctuating Pleistocene environmental conditions (*1*), because we outcompeted other species (*1*–*3*) through increased birth-rates or cognitive or social abilities, or because we outlived other *Homo* species by chance (*4*). To address *Homo* species' adaptation to environmental change (*5*–*8*) and extremes (*9*), we need to understand the connection between hominins and their ecological environment. This includes hominins' preferences for specific vegetation types, their tolerance to changes in biotic and climatic conditions, how such preferences and tolerances evolved, and their use of, adaptation to, and specialization toward different types of ecosystems.

Identifying the time-evolving habitat preferences of hominins using observational data alone is virtually impossible. With few exceptions, information on past vegetation and ecosystem changes near archaeological sites is sparse. Pollen data, leaf wax, and isotope reconstructions from terrestrial, lacustrine, or marine archives (*10*–*14*), which can provide valuable information on past vegetation and biome changes, are too limited in space and time to obtain a complete large-scale picture of how *Homo* species exploited environmental variation and responded to fluctuations in climate and available vegetation type. To overcome this limitation, various climate model simulations have been used to quantify the effect of orbital-scale changes in temperature, precipitation, and net primary productivity (NPP) on human migration and habitats (*4*, *15*–*17*).

[1]Center for Climate Physics, Institute for Basic Science, Busan, Republic of Korea. [2]Department of Climate System, Pusan National University, Busan, Republic of Korea. [3]Pusan National University, Busan, Republic of Korea. [4]DiSTAR, Napoli Università di Napoli Federico II, Monte Sant'Angelo, Italy.
*Corresponding author. Email: elkezeller@pusan.ac.kr

We examine the evolving linkage between hominins and different regional (mega-) biome types, from tropical forest to savanna to tundra and deserts, which depend on specific combinations of climatic factors, soil properties, and atmospheric CO_2 concentrations. We aim to determine the preferred environmental conditions of the different hominin species, ascertain whether their biome preferences changed in time and diversity over the Pleistocene, and identify potential regional environmental drivers of hominin adaptation. Studying biome preferences gives us a new perspective on how hominins have adapted to landscape changes, and how they exploited their habitats and food resources.

To this end, and to provide deeper insights into the adaptation patterns of the earliest members of our genus *Homo*, we extended our previously conducted quasi-transient 2-million years (Myr) Community Earth System Model (CESM version 1.2) simulation by another 1 Myr (*17*, *18*). Monthly precipitation, temperature, and percent sunshine data from the 3-Myr simulation—which experiences changes in orbital insolation (*19*), greenhouse gases and, icesheet extent (*20*)—were subsequently used to force the offline equilibrium BIOME4 vegetation model (*21*) at 0.5° resolution (*22*). BIOME4 translates climate conditions and estimates of CO_2 forcing (*20*, *23*) into 28 biomes, which are clustered into 11 megabiomes (table S2). The proxy-validated global 3 Myr climate and vegetation model simulations (figs. S1 to S4) (*22*) were then used to extract biome and NPP data for 3232 fossil and archaeological sites (*17*), compiled for 6 different *Homo* species (*H. habilis*, *H. ergaster*, *H. erectus*, *H. heidelbergensis*, *H. neanderthalensis*, and *H. sapiens)* (Fig. 1). The hominin database used here attributes well-dated fossil remains and identifiable archaeological artifacts to different hominin species. This species attribution is naturally prone to uncertainties because (i) there is no universally accepted paleo-anthropological species definition and (ii) different hom[illegible] species or even hybrids could have used [illegible]ilar tools (*17*, *24*). These uncertainties have been discussed previously, and their effect on calculations of large-scale hominin habitat preferences has been shown to be relatively minor (*17*).

Climate and vegetation changes over the past 3 Myr

To better understand how Pleistocene climate variability affected biomes and NPP, we first elucidate the response of the climate system and global vegetation to Milanković cycles and long-term shifts in atmospheric CO_2 concentrations (fig. S1, A and B). The simulated glacial/interglacial global temperature variations (fig. S1C)—primarily caused by the prescribed orbital-scale CO_2 variations in our model—have amplitudes of 7 to 8°C over the Late Pleistocene, are in good agreement with a transient simulation conducted previously with the CLIMBER intermediate complexity earth system model (*20*), and are slightly weaker than proxy-based paleotemperature reconstructions (fig. S1C) (*25*) for at least the past ~0.8 Myr (8 to 9°C). The global mean temperature variations are linked to global mean precipitation changes through the Clausius-Clapeyron relation, thus having a direct and indirect impact on global mean NPP in the offline forced BIOME4 model (fig. S1D).

Even though globally aggregated vegetation characteristics are controlled mainly by global mean temperature variations, the relationship between orbital-scale forcings (Milanković cycles, greenhouse gases, and icesheet effects) and vegetation is much more complex on regional scales (figs. S5 and S6). For instance, we find strong eccentricity-modulated precessional variability in northern Africa's expansion and contraction of grasslands and desert areas (fig. S5E). For low values of the precession index (northern summer perihelion; fig. S1A), the amplitude of the Northern hemisphere seasonal cycle of insolation increases, which leads to increased Northern Hemispheric summer rainfall, a northward expansion of grasslands, and a contraction of the Sahara desert, in qualitative agreement with paleo proxy data and other model simulations (*26*, *27*). This process has been suggested to play a key role in creating green corridors, which may have supported multiple transcontinental migration events of archaic humans (*15*, *28*). Migration from Africa into Eurasia exposed archaic humans such as *H. erectus*, *H. heidelbergensis*, and early *H. sapiens* to new environmental conditions and ecosystems.

According to our BIOME4 simulation, eastern African tropical forests gradually turned into grasslands during the early- to mid-Pleistocene (fig. S5E), which had clear implications for hominin migration and evolution. In contrast to the idea that this trend was caused

by Pleistocene aridification (*29*, *30*), we find only a weak negative rainfall trend over eastern Africa (fig. S2C), which indicates a more complex response of vegetation in this area to a variety of climate and biogeochemical drivers including CO_2 fertilization (fig. S7 and table S3). The CO_2 fertilization effect would affect C4 plants (e.g., savanna grasses) less than C3 plants because the latter exhibit a less efficient photosynthetic pathway (*31*–*33*). The simulated gradual vegetation changes, forced by orbital forcing and CO_2 changes, are important in understanding how archaic humans have adapted to their environment.

On a continental scale, our simulation reveals considerable changes in biome types (figs. S5 and S6). In Africa a gradual decrease of savanna and dry woodland in northern Africa and an expansion of grassland and dry shrublands in southern Africa occurred. Superimposed on the long-term trend are orbital variations on timescales of 20 thousand years (kyr) and ~100 kyr, which intensified after the Mid-Pleistocene Transition (MPT) ~1 million years ago (Ma) as a result of the increased amplitude of temperature and rainfall variability (fig. S1, C and H). In Asia the steady planetary cooling due to decreasing CO_2 concentrations and the expansion of the Northern Hemisphere ice sheets caused an expansion of tundra and grassland areas during glacial periods and a contraction of boreal and tropical forests. In Europe a massive increase in the latitudinal range of tundra, boreal forests, and grassland—at the expense of temperate forests—occurred following the MPT. Overall, the increasing frequency of cold climates throughout the Pleistocene led to an expansion of open biomes. This trend could have facilitated the migration of archaic humans, requiring adaptation to a broader range of environments.

Preferred biome types of hominins

Although early hominin species during the late Pliocene and early Pleistocene (~3 to 2 Ma) might have benefited from open grasslands to exploit more accessible migration routes, the diversity of biomes experienced by later species might have given them the upper edge in terms of range expansion. To demonstrate this point, we extracted the biome types for the geographic coordinates and age estimates provided in the hominin database (*17*, *34*). The resulting species-stratified biome histograms (Fig. 2) for the fossil and archaeological sites document that the habitats of both *H. habilis* and *H. ergaster* were predominately located in savanna and grassland areas (78 and 73%, respectively; table S1), in agreement with regional paleoclimate evidence (*35*). In contrast with the earlier African species, *H. erectus*—who left Africa no later than 1.8 Ma—selected for a much more diverse suite of habitats, with 58% of the sites associated with temperate or tropical forest and 38% with grassland and savanna regions. More recent species such as *H. heidelbergensis* and *H. neanderthalensis* were able to adapt to colder habitats, with 8 and 11.3% of their occurrences falling in the boreal forest biome, respectively. This post-MPT adaptation to cold environments is associated with the northward expansion of these species (*36*) (Fig. 1). *H. sapiens* managed to settle in extreme environments, including previously less occupied biome types such as tundra and desert (Fig. 2), revealing *H. sapiens* as a "generalist specialist" (*9*). This ability to occupy even extreme habitats suggests the acquisition of advanced, unparalleled cognitive abilities (*36*), which is consistent with reported changes in brain shape (*8*), more sophisticated technologies, and the transition from Middle to Upper Paleolithic toolmaking traditions around 0.05 to 0.03 Ma (*3*).

Overall, our hominin/biome occurrence analysis (Fig. 2) suggests a bimodal structure for habitat openness preferences. Early African hominins and Neanderthals appear to have preferred either open or closed habitats, whereas other species favored mixed conditions. Our analysis suggests that *H. habilis* and *H. ergaster* mostly lived in savanna/dry-woodland and grassland/dry-shrubland. *H. neanderthalensis* became a temperate forest dweller, preferring closed environments (*37*, *38*). According to our results, Neanderthals exhibited only limited capability to adapt to cold-climate open biomes such as tundra. This finding, along with the simulated Late Pleistocene biome variability (figs. S5 and S6), also suggests that Neanderthals went through a massive contraction of their habitat during glacial maxima, forcing them to settle in the warmer and more forested Mediterranean region, in agreement with recent studies (*39*, *40*). This is also seen when comparing the spatial density of specimens found in deep glacial periods with those from interglacial periods (fig. S8, A and B). Our data suggest that in contrast to early African *Homo* and *H. neanderthalensis*, *H. erectus* and *H. heidelbergensis* were able to settle in both open and closed biomes. Diversifying their habitats translated into their much wider expansion across Eurasia. Subsampling of *H. sapiens* data (fig. S9, M to P) by matching

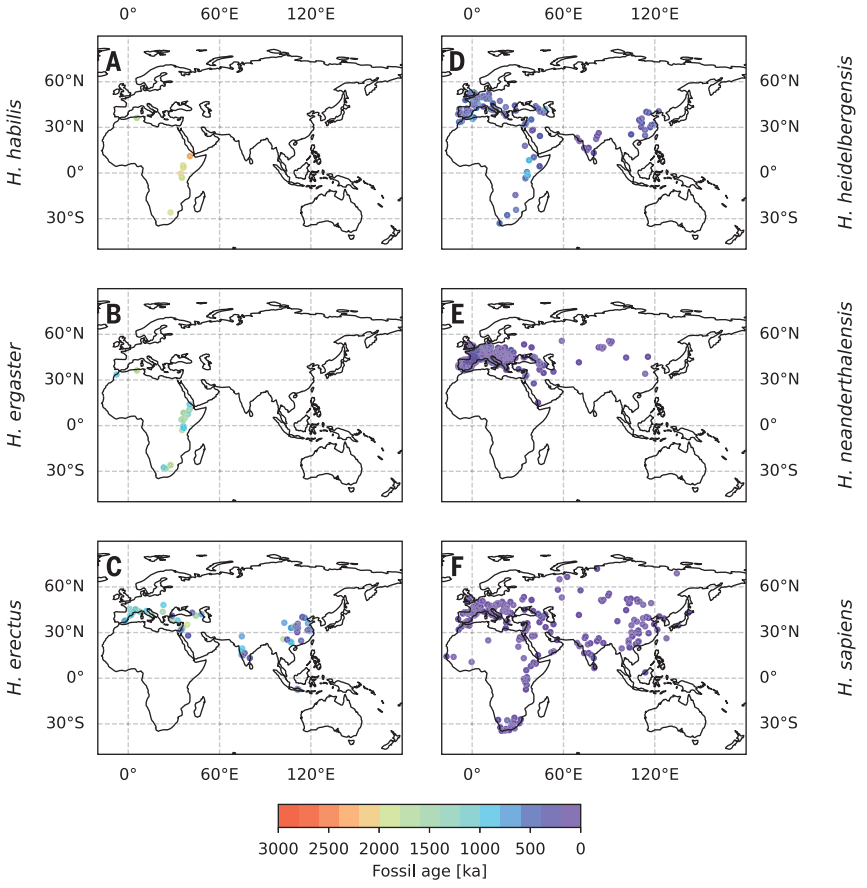


Fig. 1. Locations of hominin species. Anthropological and archaeological sites used in this study, corresponding to *H. habilis* (**A**), *H. ergaster* (**B**), *H. erectus* (**C**), *H. heidelbergensis* (**D**), *H. neanderthalensis* (**E**), and *H. sapiens* (**F**). Age estimates in ka (1000 years ago) are indicated in colored shading.

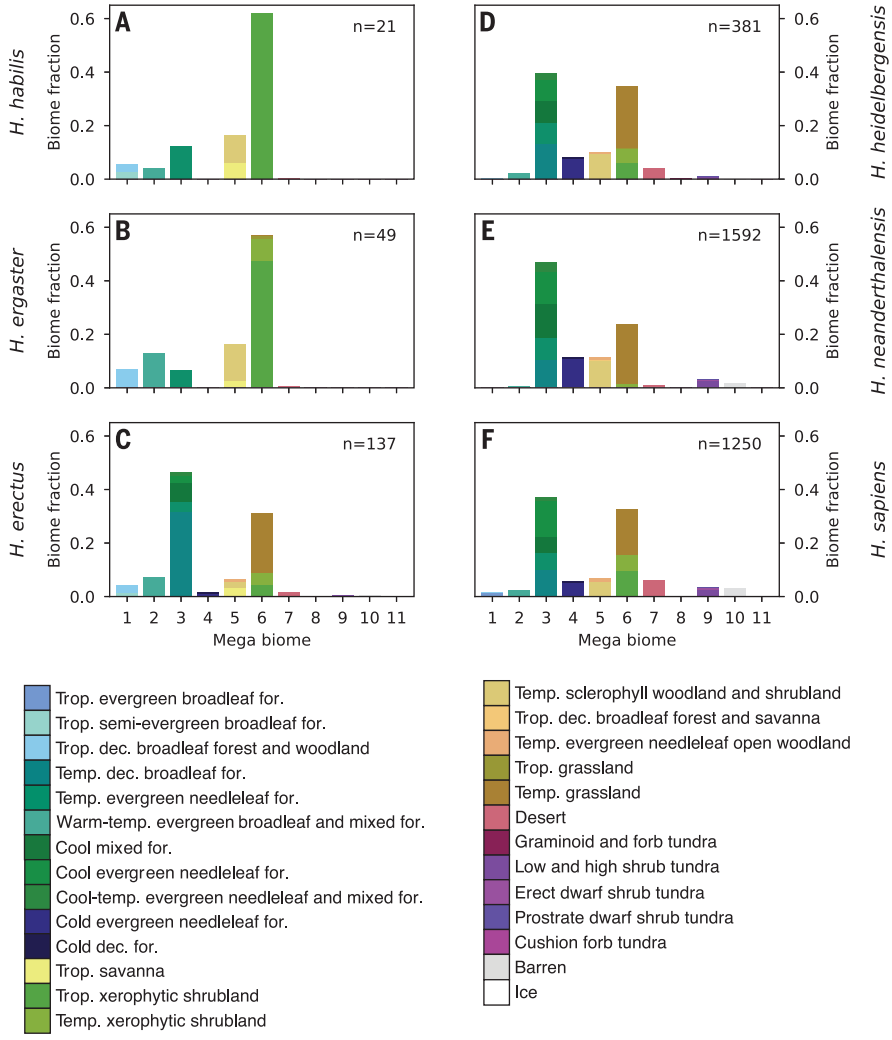


Fig. 2. Biome preferences by hominin species. Biome preferences were calculated by taking an age-weighted average of the biome occurrences at the grid point closest to the hominin specimen: *H. habilis* (**A**), *H. ergaster* (**B**), *H. erectus* (**C**), *H. heidelbergensis* (**D**), *H. neanderthalensis* (**E**), and *H. sapiens* (**F**). Fossil occurrences with uncertain species attributions are included in their respective species groupings (*22*). The number in the biome preferences represent megabiomes: 1, Tropical forest; 2, Warm-temperate forest; 3, Temperate forest; 4, Boreal forest; 5, Savanna and dry woodland; 6, Grassland and dry scrubland; 7, Desert; 8, Dry tundra; 9, Tundra; 10, Barren; 11, Ice. Abbreviations in the color bar are as follows: trop, tropical; temp, temperate; dec, deciduous; for, forest.

the number of observations of older species continues to show the wide range of occupied biomes. Therefore, the variety in occupied biomes is not due to larger sample size but rather is a clear demonstration of *H. sapiens'* ecological plasticity.

The documented diversification of biomes occupied by each hominin species over time (Figs. 2 and 3) could be a byproduct of range expansion into different areas or the consequence of gradual acclimatization to changes in climate and available vegetation types, or a combination thereof. The former would represent an adaptation to different biomes driven by migration, whereas the variability selection hypothesis, which highlights the role of regional adaptation to temporal environmental variability (*6*, *7*, *41*), could achieve the latter. To test for the influence of the temporal evolution of biomes on the presence or absence of hominins in different areas, we randomized the hominin presence data in time through bootstrap resampling (*22*), keeping the specimen locations the same, and repeated our biome preference analysis. If hominins adapted regionally to varying biomes without migrating, the original and time-shuffled analysis should yield the same results. For *H. neanderthalensis* and *H. sapiens* we find significant differences in specific biome preferences (P = 0.03 and P = 10^{-5}, respectively, using Fisher's exact test) (*22*) (fig. S9, E, F, K, and L). Therefore, the exact temporal biome trajectory might have played a role in determining where members of these species lived. The large age uncertainty around the fossil occurrences of older species, which often spans a full glacial cycle, and the corresponding reduced temporal localization could translate into insignificant differences relative to the time-shuffled biome trajectory. Repeating the time shuffling with *H. sapiens* data that was subsampled to match the other hominin group samples sizes, while maintaining the narrower age uncertainties, showed more significant results (fig. S9, M to T), indicating that the nonsignificant results for early species were likely due to the large age uncertainties.

To determine the temporal evolution of hominin biome preferences on a continental scale, we further calculated the time-evolving ratios of the individual biomes in Africa, Europe, and Asia (Fig. 3) at the hominin sites. The main biome preference in Africa is the open environment (grassland and savanna). Nonetheless, we also find occasional episodes of tropical and temperate forest occurrences. Given that the mean biome types across Europe (59% temperate and 14% boreal forest) are quite distinct from those in Africa (49% of the land area is found occupied by savanna/dry woodland or grassland/dry shrubland), we also expect a different time evolution of regional habitat preferences. From 1.8 Ma onward, a gradual trend in biome preferences from open grassland to temperate forests occurred. In certain areas, biome preferences were further modulated by glacial cycle variability during the post-MPT period (fig. S10). More specifically, we find that after 600 ka, the fraction of European temperate and warm-temperate forest dwellers correlates with regional temperatures (fig. S10, C and D), suggesting that hominins migrated in response to the climate and vegetation changes they experienced. For other time periods and regions, the correlation is less robust, which could be explained by increasing age uncertainties, smaller sample size, and/or reduced climate-induced mobility of earlier hominins.

We further examined biome preferences for overlapping species along possible ancestor-to-descendant lines (*22*). Our analysis reveals significant biome differences only for the African *H. heidelbergensis* and *H. sapiens* pair (fig. S11), consistent with the notion of a major shift in the resource exploitation mode, potentially driven by speciation dynamics (*42*).

Humans as diversity seekers

To further characterize the environmental preferences of hominins, we derive a biome diversity metric (*22*) that characterizes the number of different biome types nearest to a hominin site. If hominins preferentially lived in regions with a variety of biomes, then the biome

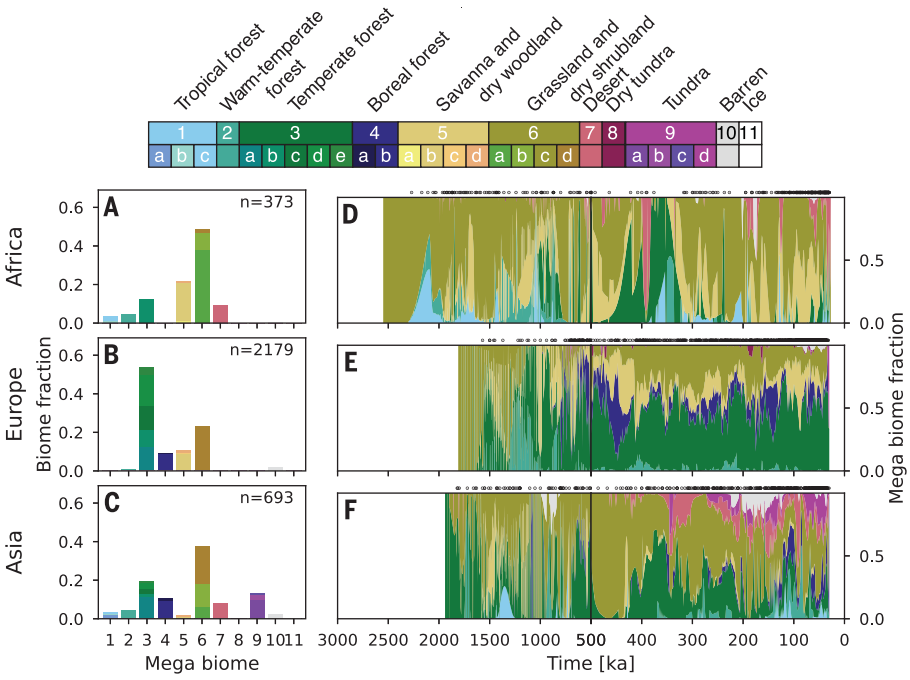


Fig. 3. Hominin biome preferences by continent. Biome preferences for hominin specimens found in Africa (**A**), Europe (**B**), and Asia (**C**) and corresponding biome preferences over time (**D** to **F**). Continents are defined by geopolitical borders and biome preferences were calculated by taking an age-weighted average of the biome occurrence at the grid point closest to the specimen. The biome preferences over time were calculated by taking the age-weighted average, stacking this for all the fossils at each time point and normalizing the stacked values to 1 (*22*). Dots above the biome preferences over time represent the average ages of the findings.

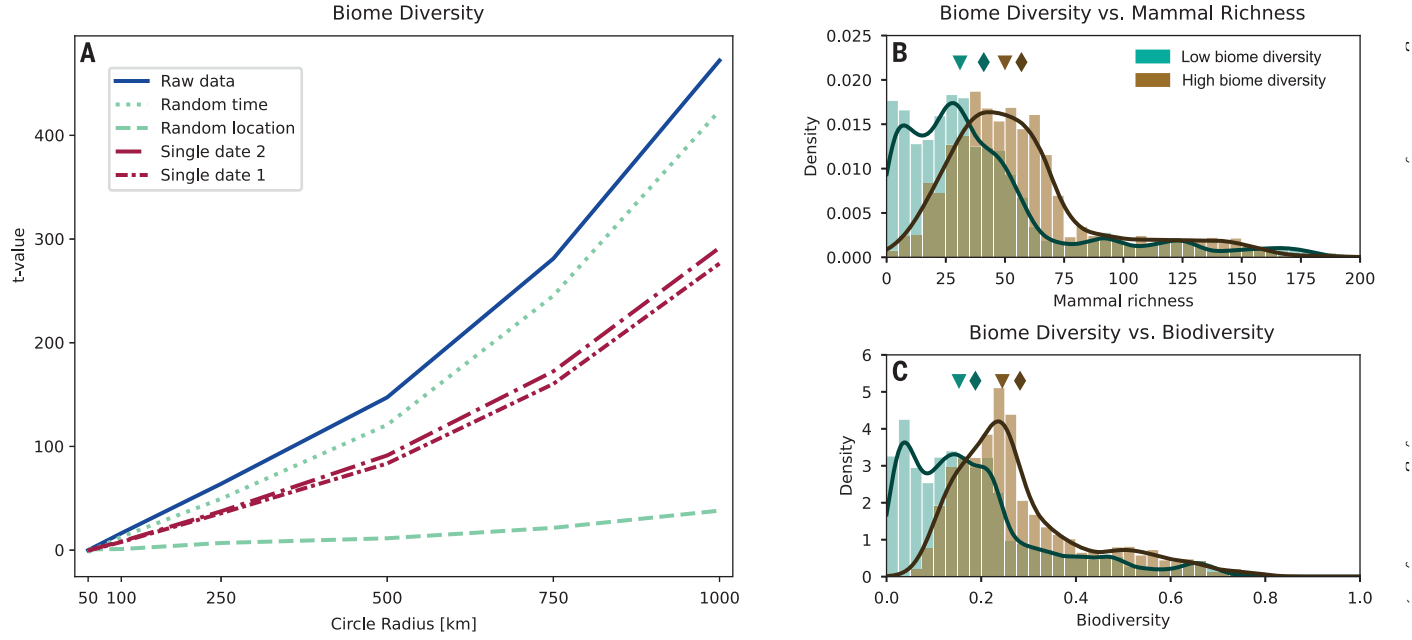


Fig. 4. Environmental preferences of early hominins. (**A**) Student's *t*-value of differences in regional biome diversity at hominin sites and grid points in surrounding circles with increasing radii (fig. S14). Higher *t*-values indicate larger deviations from zero and stronger evidence for hominin selection for higher spatial biome diversity values. The *t*-values are calculated for the original data, spatially shuffled data (between 50°N, 142°E, 45°S, and 17°W), and temporally shuffled values to test for the effect of spatial and temporal preferences (random location, random time). We also tested datasets with different hominin age uncertainties and species attributions (single date 1 and single date 2) (*22*) (fig. S13). Histograms of present-day mammal richness (**B**) (*49*) and biodiversity index computed from mammal, bird, and vascular plant richness (**C**) (*43*) for corresponding points with low biome diversity (blue) and high biome diversity (brown). For the low (1) and high (4) biome diversity cases the normalized histograms (bars and kernel density estimate in solid line) shift from low to high mammal richness and biodiversity in terms of mean value (diamond) and median value (triangle).

diversity metric should be higher closer to hominin sites than within the larger surrounding areas (fig. S12 for illustration). To test whether this hypothesis holds, the value of biome diversity at hominin sites was subtracted from the surrounding grid points within circles of increasing radii (fig. S13) and those differences were weighted by the age uncertainty (fig. S14) (*22*). Subsequently, the differences for all sites were aggregated and the Student's *t*-value was used to measure the deviation of the mean of the subsequent distribution from zero. Comparison of the hominin site values to those in an increasing circle (Fig. 4A) reveals increasing *t*-values,

which implies that hominins on average preferred regions with higher spatial biome diversity. Repeating our *t*-value analysis with randomly selected hominin locations in Africa and Eurasia, between 50°N, 142°E, 45°S, and 17°W, we find major differences with respect to the original data, thereby confirming that hominins did not choose their whereabouts at random. Shuffling the data in time, we see smaller differences with respect to the original data, which suggests that the exact temporal evolution played less of a role in selection compared with spatial biome diversity. Our analyses (movies S1 to S3, Fig. 4, and figs. S13 and S15) clearly illustrate that hominins lived in areas with multiple biome types in close proximity, suggesting that hominins preferred landscape mosaics rather than uniform environments. We repeated the *t*-value analysis for NPP (figs. S15, C and D, and S13, G to L) and found similar behavior: hominins preferred regions with locally high NPP.

Furthermore, using modern biodiversity estimates previously derived by combining mammal, bird, and vascular plant richness data (*43*), we find a connection between biome diversity and this particular measure of biodiversity (Fig. 4, B and C, and fig. S16). Taken together, our findings suggest that human species were actively seeking out more abundant (figs. S15, C and D, and S13, G to L) and diverse food resources (Fig. 4A) in regions with higher ecosystem diversity (Fig. 4C). Using a relationship between densities of historical hunter-gatherer populations (*44*) and our calculated preindustrial biome diversity (fig. S16, I to L) as a proxy for the past, we propose that diverse landscapes may have also sustained larger hominin populations, serving as hotspots for cultural innovation and playing an important role in genetic diversification (*44*).

The derived inclination of humans toward landscape mosaics and their ability to adapt to regionally diverse vegetation conditions has not been shown before at continental scales, leading us to propose the diversity selection hypothesis: *Homo* species, and *H. sapiens* in particular, were specially equipped to exploit heterogeneous habitat conditions. Such heterogeneity, which correlates to high plant and vertebrate richness in current ecosystems (Fig. 4, B and C), may have conferred hominin resilience to environmental perturbations by providing a wider and more stable resource base (*44*). Here we suggest that behavioral and cultural plasticity may have allowed humans to exploit habitat diversity and diverse food resources.

Several single-site studies have alluded to the fact that hominins may have favored heterogeneous habitats (*45*–*47*), leading to the formulation of the microhabitat variability hypothesis (*48*). Combined with our large-scale results, we conclude that Pleistocene hominins selected for increasing regional biome heterogeneity. We acknowledge that temporal variability selection (*6*) may have been present in specific regions and time periods. However, our analysis does not reveal any statistically robust continental-scale linkage between where hominins lived and temporal changes in orbitally driven vegetation variance.

Our data-based diversity selection hypothesis may also add context to our more recent human history. According to our analysis, our genus *Homo* has adapted over the Pleistocene and migrated to areas with higher landscape diversity. Utilizing resources from various biomes provided a resilient and successful strategy over hundreds of millennia. However, during the Anthropocene, our species has caused a massive decline in global ecosystem diversity due to land use practices, gradually shifting away from integrated agricultural practices and toward monocultures. Modern humans have clearly taken an unprecedented path away from our ancestors' resilience and diversity-based strategies.

REFERENCES AND NOTES

1. A. Timmermann, *Quat. Sci. Rev.* **238**, 106331 (2020).
2. W. E. Banks *et al*., *PLOS ONE* **3**, e3972 (2008).
3. C. S. Henshilwood *et al*., *Science* **295**, 1278–1280 (2002).
4. A. R. Vahdati, J. D. Weissmann, A. Timmermann, M. S. P. de Leon, C. P. E. Zollikofer, *Quat. Sci. Rev.* **221**, 105867 (2019).
5. E. S. Vrba, *Am. J. Sci.* **293**, 418–452 (1993).
6. R. Potts, in *Yearbook of Physical Anthropology*, C. Ruff, Ed. (1998), vol. 41, pp. 93-136.
7. M. A. Maslin, S. Shultz, M. H. Trauth, *Philos. Trans. R. Soc. B.* **370**, 20140064 (2015).
8. S. Neubauer, J. J. Hublin, P. Gunz, *Sci. Adv.* **4**, eaao5961 (2018).
9. P. Roberts, B. A. Stewart, *Nat. Hum. Behav.* **2**, 542–550 (2018).
10. R. Patalano, P. Roberts, N. Boivin, M. D. Petraglia, J. Mercader, *Evol. Anthropol.* **30**, 385–398 (2021).
11. S. J. Feakins, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **374**, 62–71 (2013).
12. T. Litt, N. Pickarski, G. Heumann, M. Stockhecke, P. Tzedakis, *Quat. Sci. Rev.* **104**, 30–41 (2014).
13. K. H. Roucoux, P. C. Tzedakis, L. de Abreu, N. J. Shackleton, *Earth Planet. Sci. Lett.* **249**, 307–325 (2006).
14. R. L. Quinn, C. J. Lepre, *J. Hum. Evol.* **140**, 102549 (2020).
15. A. Timmermann, T. Friedrich, *Nature* **538**, 92–95 (2016).
16. A. Eriksson *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16089–16094 (2012).
17. A. Timmermann *et al*., *Nature* **604**, 495–501 (2022).
18. J. W. Hurrell *et al*., *Bull. Am. Meteorol. Soc.* **94**, 1339–1360 (2013).
19. A. Berger, *Quat. Res.* **9**, 139–167 (1978).
20. M. Willeit, A. Ganopolski, R. Calov, V. Brovkin, *Sci. Adv.* **5**, eaav7337 (2019).
21. J. O. Kaplan *et al*., *J. Geophys. Res.* **108**, D19 (2003).
22. Materials and methods are available as supplementary materials.
23. D. Lüthi *et al*., *Nature* **453**, 379–382 (2008).
24. M. D. Petraglia, *Nature* **604**, 430–432 (2022).
25. T. Friedrich, A. Timmermann, M. Tigchelaar, O. Elison Timm, A. Ganopolski, *Sci. Adv.* **2**, e1501923 (2016).
26. J. C. Larrasoaña, A. P. Roberts, E. J. Rohling, *PLOS ONE* **8**, e76514 (2013).
27. C. L. Blanchet *et al*., *Nat. Geosci.* **14**, 97–103 (2021).
28. J. E. Tierney, P. B. deMenocal, P. D. Zander, *Geology* **45**, 1023–1026 (2017).
29. R. B. Owen *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 11174–11179 (2018).
30. P. B. deMenocal, *Science* **331**, 540–542 (2011).
31. S. P. Harrison, C. I. Prentice, *Glob. Change Biol.* **9**, 983–1004 (2003).
32. A. Brooks, G. D. Farquhar, *Planta* **165**, 397–406 (1985).
33. C. P. Osborne, L. Sack, *Philos. Trans. R. Soc. B.* **367**, 583–600 (2012).
34. P. Raia *et al*., *One Earth* **3**, 1–11 (2020).
35. K. E. Reed, *J. Hum. Evol.* **32**, 289–322 (1997).
36. A. Mondanaro *et al*., *iScience* **23**, 101693 (2020).
37. J. S. Carrion *et al*., *Glob. Ecol. Biogeogr.* **12**, 119–129 (2003).
38. J. S. Carrion, C. Lalueza-Fox, J. Stewart, *Quat. Sci. Rev.* **217**, 1–6 (2019).
39. M. Melchionna *et al*., *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **496**, 146–154 (2018).
40. B. M. Benito *et al*., *J. Biogeogr.* **44**, 51–61 (2017).
41. M. Grove, *J. Hum. Evol.* **61**, 306–319 (2011).
42. J. L. Cantalapiedra, M. Hernández Fernández, J. Morales, *PLOS ONE* **6**, e28749 (2011).
43. M. Tallavaara, J. T. Eronen, M. Luoto, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 1232–1237 (2018).
44. M. J. Hamilton, B. T. Milne, R. S. Walker, J. H. Brown, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 4765–4769 (2007).
45. H. M. Liddy, S. J. Feakins, J. E. Tierney, *Earth Planet. Sci. Lett.* **449**, 430–438 (2016).
46. R. Lupien *et al*., *Quat. Sci. Rev.* **186**, 225–235 (2018).
47. J. Mercader *et al*., *Nat. Commun.* **12**, 3 (2021).
48. R. Patalano *et al*., *Front. Earth Sci.* **9**, 787669 (2021) .
49. International Union for Conservation of Nature (IUCN), Center for International Earth Science Information Network (CIESIN), Columbia University, Gridded Species Distribution: Global Mammal Richness Grids (NASA Socioeconomic Data and Applications Center, 2015); http://dx.doi.org/10.7927/H4N014G5.
50. A. Timmermann *et al*., 3ma transient climate simulation, ICCP Dataserver (2022); https://doi.org/10.22741/iccp.20230001
51. E. Zeller *et al*., 3ma Biome4, ICCP Dataserver (2023); https://doi.org/10.22741/iccp.20230002.
52. E. Zeller *et al*., 2023 Science, ICCP Dataserver (2023); https://doi.org/10.22741/iccp.20230003.

ACKNOWLEDGMENTS

The CESM1.2 and BIOME4 simulations were conducted on the ICCP/IBS supercomputer Aleph, a Cray XC50-LC system. **Funding:** This work was funded by the following: Institute for Basic Science (IBS-R028-D1) (to E.Z., A.T., K.-S.Y., K.S., and J.R.). **Author contributions:** Conceptualization: E.Z. and A.T. Data curation: E.Z. Formal analysis: E.Z. and A.T. Funding acquisition: A.T. Investigation: E.Z. and A.T. Methodology: E.Z., A.T., K.-S.Y., P.R., K.S., and J.R. Resources: A.T. Software: E.Z. Validation: E.Z., A.T., and K.S. Visualization: E.Z. Project administration: E.Z. and A.T. Supervision: A.T. Writing – original draft: E.Z. and A.T. Writing – review and editing: E.Z., A.T., K.-S.Y., P.R., K.S., and J.R. **Competing interests:** Authors declare that they have no competing interests. **Data and materials availability:** The climate data (*50*) and biome data (*51*) reported in this manuscript can be found at the ICCP dataserver. Complete source code for data processing, as well as Jupyter notebooks with all analyses and results, can be found at the ICCP dataserver (*52*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abq1288
Materials and Methods
Figs. S1 to S16
Tables S1 to S3
References (*53*–*84*)
Movies S1 to S3

Submitted 20 March 2022; resubmitted 3 January 2023
Accepted 17 March 2023
10.1126/science.abq1288

ELECTROCHEMISTRY

La- and Mn-doped cobalt spinel oxygen evolution catalyst for proton exchange membrane electrolysis

Lina Chong[1], Guoping Gao[2], Jianguo Wen[3], Haixia Li[2], Haiping Xu[1], Zach Green[4], Joshua D. Sugar[5], A. Jeremy Kropf[1], Wenqian Xu[6], Xiao-Min Lin[3], Hui Xu[4], Lin-Wang Wang[2], Di-Jia Liu[1,7]*

Discovery of earth-abundant electrocatalysts to replace iridium for the oxygen evolution reaction (OER) in a proton exchange membrane water electrolyzer (PEMWE) represents a critical step in reducing the cost for green hydrogen production. We report a nanofibrous cobalt spinel catalyst codoped with lanthanum (La) and manganese (Mn) prepared from a zeolitic imidazolate framework embedded in electrospun polymer fiber. The catalyst demonstrated a low overpotential of 353 millivolts at 10 milliamperes per square centimeter and a low degradation for OER over 360 hours in acidic electrolyte. A PEMWE containing this catalyst at the anode demonstrated a current density of 2000 milliamperes per square centimeter at 2.47 volts (Nafion 115 membrane) or 4000 milliamperes per square centimeter at 3.00 volts (Nafion 212 membrane) and low degradation in an accelerated stress test.

Low-temperature water electrolysis can rapidly produce environmentally sustainable, or green, hydrogen and is a prospective means of storing energy from renewable but intermittent power sources, such as wind and solar, in future clean energy infrastructure (*1*–*4*). Commercial systems use either liquid alkaline electrolyte or proton exchange membrane electrolyte (*1*). Compared with the alkaline counterpart, a proton exchange membrane water electrolyzer (PEMWE) offers the advantages of higher current density, higher H_2 purity, lower resistance losses, and more compact design, rendering it a preferred technology where high efficiency and small footprint are essential (*1*, *2*). Working under the acidic and oxidative environment, however, adds substantial challenges to the catalyst activity and stability profile. This is particularly the case for the anode catalyst because of the high overpotential for oxygen evolution reaction (OER) (*3*). At present, OER catalysts for PEMWE are primarily restricted to the platinum group metal (PGM) materials, such as IrO_x (*1*). Their high cost and limited reserve, however, pose substantial barriers to the widespread implementation of PEMWE. Low-cost transition metals and their oxides are known to be active toward OER in alkaline electrolyte (*4*–*6*), but their demonstration in acidic electrolyte is very limited (*1*, *7*, *8*).

Cobalt molecular and oxide compounds have emerged as promising OER catalysts for water splitting in recent years (*9*). Kanan and Nocera systematically investigated water oxidation using a catalyst deposited from Co^{2+} solution in pH 7 phosphate buffer (*10*). Gerken *et al.* performed a comprehensive mechanistic study of cobalt-catalyzed water oxidation from homogeneous to heterogeneous phases in electrolyte over a broad pH range of 0 to 14 (*11*). Those studies have provided profound understanding of electrocatalytic oxygen evolution by cobalt. More recently, thin film spinel-type Co_3O_4 was found to be active toward OER and stable at low overpotential in acid (*12*). The activity and stability of Co oxide were improved substantially when modified with iron, manganese, antimonite, and PbO_x (*13*–*15*). In addition to activity and stability, the inherent conductivity represents an essential requirement for efficient OER electrocatalysis to overcome the insulating properties of most transition metal oxides in their crystalline form. Conventional carbon supports, used to facilitate the electron conductivity, are not stable against oxidation to CO_2 at the PEMWE anode under OER operating potential. For example, a Co-polyoxometalate composited with carbon paste achieved a lower OER onset potential than IrO_2 in 1 M sulfuric acid but decayed quickly, presumably as a result of oxidative corrosion of the carbon (*16*). Self-conductive oxide catalysts can also enhance the active -site volumetric density without being diluted by a secondary nonactive support. A recently developed self-healing OER catalyst has demonstrated excellent activity and durability in acidic electrolyte (*14*). The approach, however, requires the presence of metal precursors in the aqueous electrolyte, hindering integration into a PEMWE. Most of the aforementioned studies were carried out either in half-cells or aqueous electrolyzers, where the demands for OER catalyst stability and conductivity are different from those in a PEMWE. For example, the dissolved transition metal concentration must be minimized to avoid poisoning [illegible] proton exchange membrane in the PEM[illegible]. Effective OER for PEMWE requires optimal interfacial properties, microporosity, and surface catalytic activity (*1*), all of which need to be validated in the operating electrolyzer.

In this work, we present a cobalt spinel–based OER catalyst derived from a zeolitic methyl-imidazolate framework (Co-ZIF) and processed by electrospinning. The catalyst demonstrated excellent OER activity benefiting from its high specific surface area, porous interconnected nanonetwork structure, and high conductivity.

Design of an acid-stable cobalt OER catalyst

Our design concept of an efficient cobalt spinel–based OER catalyst for PEMWE anodes is based on the following rationale: To enhance OER activity in acid, an oversized and more stable second element can be selectively introduced to the cobalt oxide surface to generate strain, oxygen vacancy (V_o), and acid tolerance (*17*); to improve the oxide electronic conductivity, a third element with similar charge and dimension to cobalt may be incorporated inside the lattice to bridge the Fermi bandgap through *d* orbital partial occupation of the third element induced by its d-electron delocalization. Advancing further from the rotating disk electrode (RDE) or half-cell study to a membrane electrode assembly (MEA) demonstration, the catalyst should have a high porosity and surface area easily accessible to the reactant (H_2O). Meanwhile, the electrode layer should be effective in transporting H_2O and releasing O_2 without blocking the water-catalyst interface. Furthermore, the oxide catalyst should be self-conductive without the need for another conductive support, such as carbon, that is unstable under high OER operating potential and current density. Finally, the metal oxide should be stable against oxidative and acidic (pH 2 to 4) corrosion in the PEMWE environment.

We selected Co-ZIF as the precursor for the catalyst preparation because of its high intrinsic porosity and reticular structure. It has recently been used to prepare a nanoplate oxide with excellent OER activity tested in strong alkaline electrolyte (1 M KOH) using the RDE method (*18*). Our catalyst preparation through low-temperature oxidation partially retained the porosity and morphology of Co-ZIFs after their conversion into interconnected hollow metal oxide particles, providing an excellent platform for enhanced charge and mass transfer. Among different elements that we screened, we selected lanthanum (La^{3+}) as the second element because of its much larger radius compared with that of Co^{2+} (1.06 Å versus 0.72 Å) along with its strong affinity to bind –OH groups at the surface of cobalt oxide. We also added manganese ions (Mn^{2+}) of similar

[1]Chemical Science and Engineering Division, Argonne National Laboratory, Lemont, IL 60439, USA. [2]Material Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [3]Center for Nanoscale Materials, Argonne National Laboratory, Lemont, IL 60439, USA. [4]Giner Inc., Auburndale, MA 02466, USA. [5]Sandia National Laboratory, Livermore, CA 94550, USA. [6]X-ray Science Division, Advanced Photon Source, Argonne National Laboratory, Lemont, IL 60439, USA. [7]Pritzker School of Molecular Engineering, University of Chicago, Chicago, IL 60637, USA.
*Corresponding author. Email: djliu@anl.gov

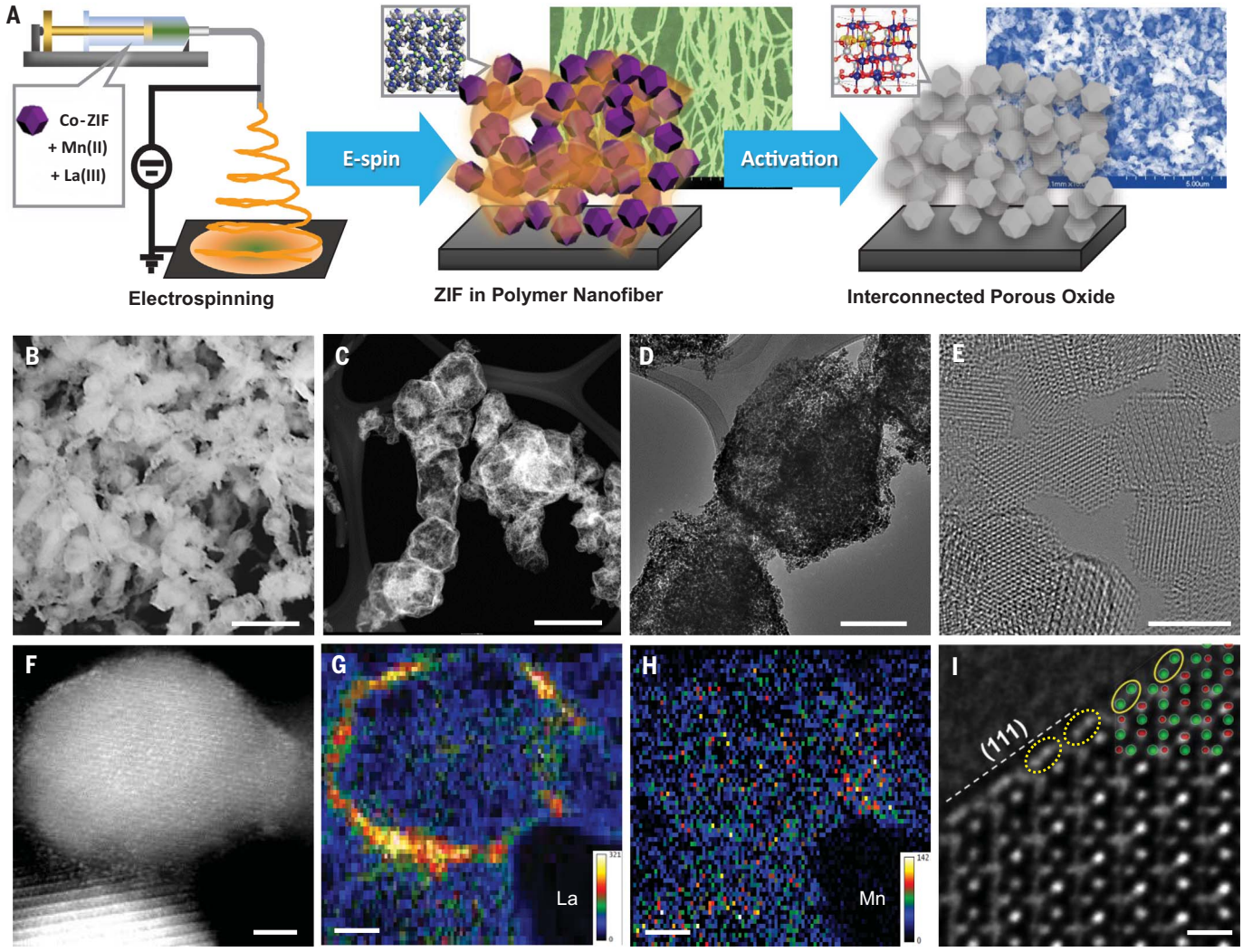


Fig. 1. Synthesis, morphology, and structure of LMCF. (**A**) Schematics of LMCF synthesis including formation of Co-ZIF embedded PAN polymer fiber by electrospinning and thermal oxidative activation to produce interconnected porous cobalt oxide particles after removing all the organics. The background SEM images show a cross-linked, ZIF-containing fiber nanonetwork produced by electrospinning and the interconnected porous oxide after the activation. (**B**) SEM image (scale bar, 1 μm). (**C**) HAADF-STEM image (scale bar, 500 nm). (**D**) TEM image (scale bar, 200 nm). (**E**) HRTEM image (scale bar, 5 nm). (**F** to **H**) STEM image and the corresponding La and Mn distributions (scale bar, 2 nm). The color bars show the element counts. The maximal counts are 321 for La and 142 for Mn. (**I**) HRTEM image (scale bar, 0.5 nm). The green dots represent atomic columns of tetrahedral (T) and octahedral (O) oxygens, and the red dots represent the cobalt atomic columns simulated based on bulk phase inside lattice. The dotted yellow ellipses (on surface) show a different orientation compared with the solid yellow ellipses from the simulation (in bulk), suggesting a shift of oxygen position as a result of lattice relaxation.

radius to Co^{2+} (0.8 Å versus 0.72 Å) during the Co-ZIF synthesis, which would be oxidized to Mn^{3+} (0.72 Å) during the oxidative conversion and uniformly distributed inside the cobalt spinel lattice to promote conductivity (via bandgap) and OER activity (via affinity OH or H group). The catalyst synthesis scheme is shown in Fig. 1A. Briefly, La- and Mn-doped Co-ZIF, LaMn@Co-ZIF, was prepared in solution. Powder x-ray diffraction (XRD) combined with scanning electron microscopy (SEM)–energy dispersive x-ray spectroscopy (EDX) elemental mapping confirmed the successful incorporation of atomic La and Mn into the structure and cavity of the Co-ZIF (fig. S1). The LaMn@Co-ZIF was then suspended in a polyacrylonitrile (PAN) polymer slurry (LaMn@Co-ZIF/PAN) (*19*), which was subsequently electrospun into a fibrous mat. The nanofiber embedded with individual LaMn@Co-ZIF was activated in flowing air at 360°C for 6 hours to remove the organic components, forming La- and Mn-codoped porous cobalt spinel fibers (denoted LMCF). Thermogravimetric analysis (TGA) confirmed the removal of carbon and nitrogen in LMCF after oxidative activation (fig. S2). The SEM image of LMCF shows an interconnected nanofibrous network morphology with ample macropores in between the entwined nanofibers (Fig. 1B). High-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) revealed that individual LMCF particles retained the original Co-ZIF's rhombic dodecahedral shape, aligned and fused together in strings after oxidation (Fig. 1C). The ZIF-shaped LMCF particle is highly porous with a hollow structure composed of nanopores, as shown by transmission electron microscopy (TEM) (Fig. 1D). Each particle is composed of aggregates of Co_3O_4 nanocrystallites, with an average size of ~3.5 nm, determined by aberration-corrected high-resolution transmission electron microscopy (HRTEM) (Fig. 1E and fig. S3). Nitrogen adsorption measurement of LMCF at 77 K

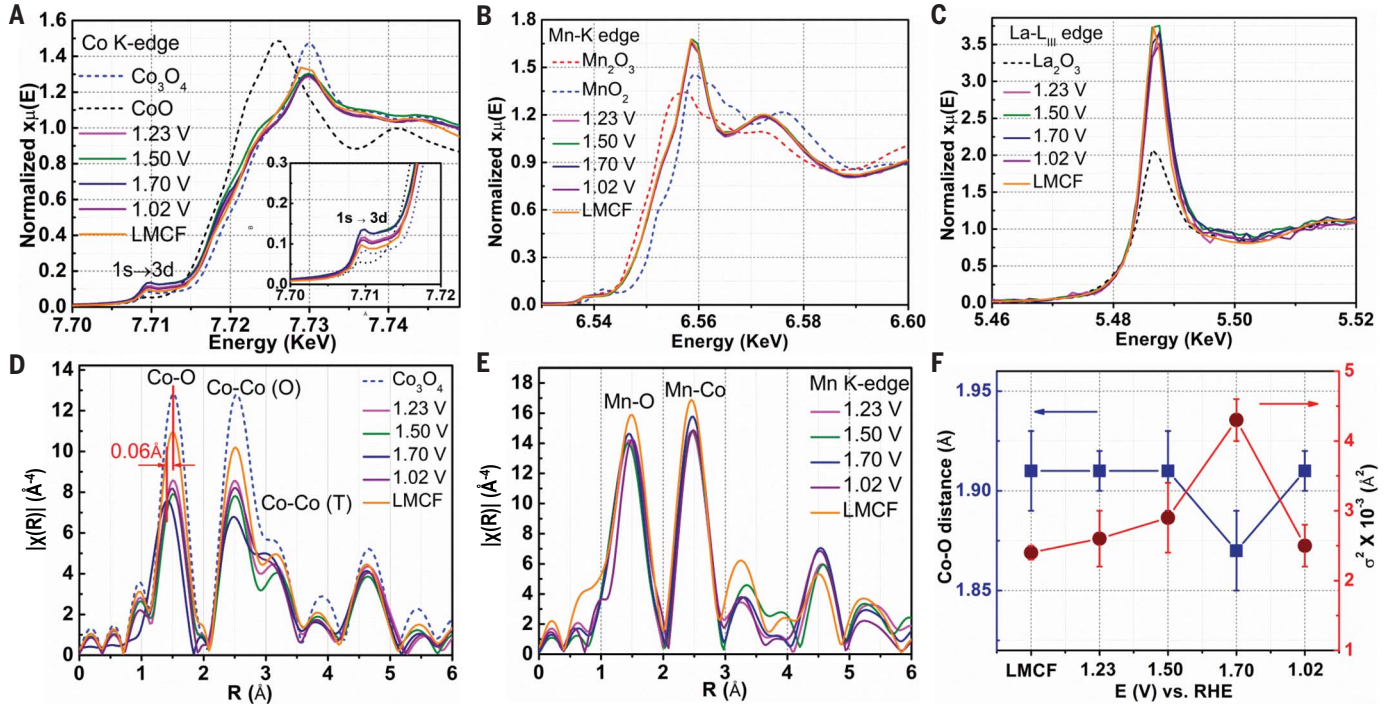


Fig. 2. XAS study of LMCF. (**A** to **C**) Fluorescence XANES spectra collected at Co K-edge (A) (inset: enlarged pre-edge 1s → 3d transition), Mn K-edge (B), and La L_{III}-edge (C) under ex situ (LMCF) and operando conditions at different potentials. Co_3O_4, CoO, Mn_2O_3, MnO_2, and La_2O_3 are used as the references. (**D** and **E**) R-space EXAFS spectra at Co K-edge (D) and Mn K-edge (E) of the same samples. (**F**) Co-O bond distances (blue line) and DWF (red line) surround Co derived from EXAFS data taken at different cell potentials. Error bars represent the uncertainty of Fourier transformation of the experimental EXAFS data.

provided a Brunauer–Emmett–Teller (BET) specific surface area (SSA) of 197 $m^2\ g^{-1}$ and a pore volume of 0.463 $cm^3\ g^{-1}$ (fig. S4). The high porosity of individual LMCF particle combined with nanofibrous morphology improves accessibility of reactants to the catalytic sites and facilitates water-oxygen mass transport in and out of the catalyst structure—an essential attribute for high-OER current density. XRD and Raman spectroscopy confirmed that the catalyst exhibited a spinel Co_3O_4 structure of slightly expanded lattice and higher Co^{2+}/Co^{3+} ratio (fig. S5C and table S1). STEM images showed that the individual crystal surface is dominated by (111) facets, and electron energy-loss spectroscopy (EELS) elemental mapping revealed La localization on the surface, with Mn distributed mainly in the bulk (Fig. 1, F to H, and fig. S6). Low-magnification EDX elemental mapping disclosed a uniform distribution of Co, Mn, La, and O in LMCF, and inductively coupled plasma optical emission spectrometry (ICP-OES) determined an atomic ratio of Co:Mn:La of 80:12:8 (fig. S7 and table S3). HRTEM imaging revealed that the LMCF lattice surface was terminated by oxygen atoms in a relaxed state, with positions shifted from those inside the crystallite (*20*) due to V_o (Fig. 1I and fig. S8), an important attribute in lowering the OER activation energy and stabilizing the intermediate during the reaction (*17*, *21*). We also measured LMCF conductivity using the four-probe van der Pauw method for comparison with IrO_x and commercial Co_3O_4. The LMCF conductivity was found to be 8.6 times as high as that of commercial Co_3O_4 and about two-thirds that of IrO_x (fig. S9).

The electronic configuration and the coordination structure of Co, La, and Mn in LMCF were investigated using x-ray photoelectron spectroscopy (XPS) (fig. S10A and table S4) and synchrotron x-ray absorption spectroscopy (XAS). The high-resolution O 1s XPS spectrum confirmed the presence of high V_o concentration in LMCF (fig. S11A). The Co 2p XPS spectrum revealed a higher ratio of Co^{2+}:Co^{3+} in LMCF compared with that in Co_3O_4 (fig. S11B). The x-ray absorption near-edge structure (XANES) spectrum at the Co K-edge (LMCF) shows a very similar spectral pattern to that of Co_3O_4 with a slightly red-shifted absorption energy and a decreased white line (WL) intensity (Fig. 2A), indicating a lower average oxidation state and a smaller O coordination number (CN) to cobalt in LMCF, which agrees well with the Raman and XPS results. We also observed an enhanced 1s → 3d transition peak intensity, which reveals that the cobalt in LMCF is in a less centrosymmetric coordination environment than that in Co_3O_4, suggesting a distorted Co oxide lattice by V_o. Compared with Co_3O_4, the LMCF extended x-ray absorption fine structure (EXAFS) shows lower peak intensities and CNs corresponding to Co-O and Co-Co shells (Fig. 2D, fig. S12, and table S5), which further supports a less developed lattice with a high concentration of V_o from smaller particle size and higher Co^{2+} (tetrahedral O-coordinated) fraction in LMCF. XANES at the Mn K-edge indicates that the average manganese oxidation state is between +3 and +4 (Fig. 2B). Both K-space and R-space spectra extracted from EXAFS exhibit significant differences from that in Mn oxide references (Mn_3O_4, Mn_2O_3, and MnO_2) (fig. S13, B and C). R-space fitting determined the CNs of Mn to O and Mn to Co to be 5.5 ± 0.4 and 7.2 ± 0.3, respectively (table S5). Most noticeably, the first and second shell radii and CNs are close to those of Co-O and Co-$Co_{(oh)}$ in Co_3O_4 (Fig. 2E and tables S5 and S6) instead of Mn-O and Mn-Mn paths in Mn references, including Mn_3O_4, Mn_2O_3, and MnO_2 (fig. S13B). These observations provide convincing evidence that the Mn substitutes for the Co^{3+} at the edge-sharing octahedral site and is embedded inside the cobalt oxide matrix in LMCF, in agreement with the HAADF-STEM result. The Mn^{3+} in the lattice is known to enhance OER activity of the oxide in acid (*22*). The XANES spectrum of La in LMCF shows significantly higher WL intensity than that of the La_2O_3 reference (Fig. 2C). Given that the WL intensity for oxides is generally proportional to the number of coordinated oxygens, R-space fitting determined the CN

of La to oxygen in LMCF to be 8.3, which is in between that of lanthanum oxide (CN = 6) and hydroxide (CN = 9) (fig. S14).

Electrocatalytic activity in solution

We first evaluated the OER catalytic activity of LMCF using the RDE method in 0.1 M $HClO_4$ electrolyte (pH = 1). To better understand the effects of the second and third elemental doping, we also prepared Co_3O_4 fiber (CF) and La-doped Co_3O_4 fiber (LCF) using a similar Co-ZIF–electrospinning method. Commercial Co_3O_4 and Ir black were also studied as benchmarks. Figure 3A shows a progressive improvement of OER activity measured by cyclic voltammogram (CV) through the addition of Mn and La in CF. The performance of LMCF also significantly exceeds that of commercial Co_3O_4 and approaches that of Ir black. Figure 3B presents the mass activity (MA) Tafel plot for LMCF together with the benchmark samples. LMCF again exhibits a high intrinsic catalytic activity. For example, LMCF catalyst shows a bulk MA of 126 ± 20 A g^{-1} at an overpotential of 370 mV, which is higher than that of commercial Ir black (table S7). This is possibly because of a higher gravimetric catalytic site density of LMCF due to its lower molecular weight compared with IrO_x. LMCF evaluated by linear sweep

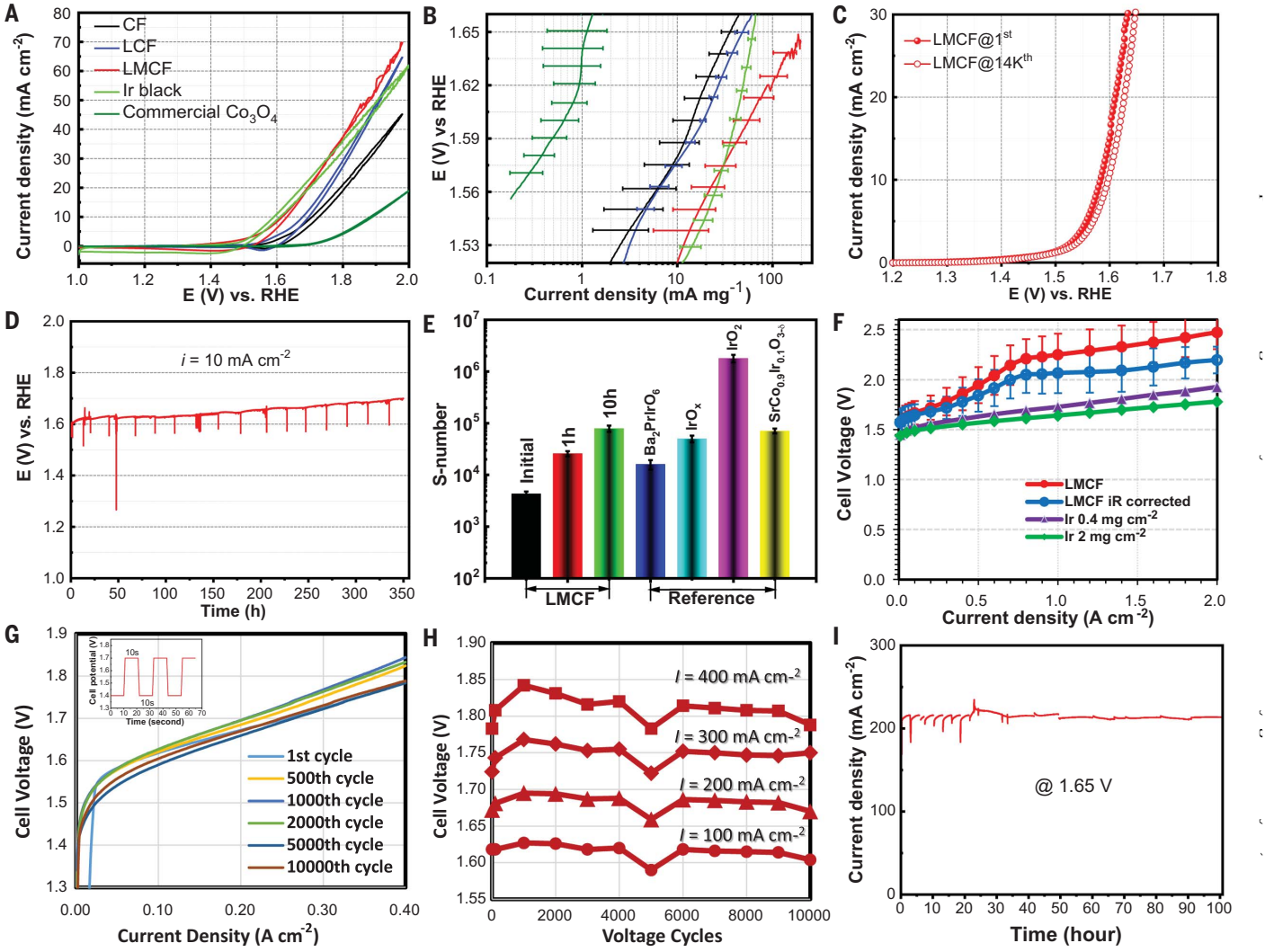


Fig. 3. Electrocatalytic performance of LMCF. (**A**) CVs of LMCF, LCF, CF, Ir black, and commercial Co_3O_4 in O_2-saturated 0.1 M $HClO_4$ (PGM-free catalyst loadings = ~260 ± 30 μg cm^{-2}, Ir black loading = ~230 ± 30 μg cm^{-2}). (**B**) Tafel plots of LMCF, LCF, CF, Ir black, and commercial Co_3O_4, with error bars of one standard deviation over four experimental replicates. (**C**) LSVs of LMCF measured by RDE before and after 14,000 voltage cycles in O_2-saturated 0.1 M $HClO_4$ (with 95% *iR* correction). (**D**) Chronopotentiometric response at 10 mA cm^{-2} with LMCF catalyst loading of 0.9 mg cm^{-2} over 353-hour test. (**E**) *S* number calculated for LMCF after different hours onstream compared with selected benchmark Ir-based catalysts. (**F**) Current-voltage polarizations (*iR* corrected and uncorrected) of the PEMWE cell with LMCF anodic catalyst compared with that of Ir black catalysts with different loadings at 80°C. The polarization plot for LMCF represents an average of three MEA measurements with one standard deviation. (**G**) Current-voltage polarizations of the PEMWE after selected cycle numbers during a multiple-voltage cycling AST. The inset shows the stepwise voltage swing between 1.4 V and 1.7 V with 10-s dwell time at each potential. (**H**) PEMWE cell voltage measured at different current densities after selected voltage cycles during the AST. (**I**) Potentiostatic measurement of PEMWE at the cell potential of 1.65 V for LMCF anodic catalyst over 100 hours. Test conditions: anode LMCF catalyst loading, 1 to 2 mg cm^{-2}; cathode Pt loading, 0.4 mg_{Pt} cm^{-2}; 60° or 80°C cell temperature, unless otherwise specified; 5-cm^2 active electrode area; Nafion 115 membrane for (F), (G), and (H), and Nafion 212 membrane for (I); DI water at flow rate of 10 ml min^{-1}.

voltammetry (LSV) (Fig. 3C, with 95% *iR* correction) shows an onset potential of 1.28 V measured at 0.46 mA cm^{-2} and an overpotential of 353 ± 30 mV at 10 mA cm^{-2}. The LMCF catalyst activity was also measured in 0.5 M H_2SO_4 (pH = 0), and the overpotential was reduced to 335 ± 30 mV at 10 mA cm^{-2} (fig. S15). These results placed LMCF among the most active PGM-free catalysts reported in aqueous acid (*23*) (table S2). We estimated the electrochemical surface areas (ECSA) of the catalysts by measuring the double-layer capacitance from the CV curves in the non-Faradaic region (fig. S16) and produced Tafel plots of the ECSA-based specific current densities (fig. S17). The intrinsic activities of LMCF were further assessed based on turnover frequencies (TOFs) at different overpotentials (320 mV, 370 mV, and 650 mV), which are among the highest when compared with representative PGM-free and PGM OER catalysts tested in various acidic media (table S8). For example, the TOF of LMCF is calculated to be 0.079 ± 0.011 s^{-1} at an overpotential of 370 mV based on the total loading mass, which increases to 0.87 ± 0.09 s^{-1} when calculated based on ECSA (table S7). We also analyzed O_2 produced during OER over LMCF in a H cell using in situ gas chromatography (GC) and calculated the Faradaic efficiencies (FEs) for the oxygen produced at the current densities of 10 mA cm^{-2}, 20 mA cm^{-2}, 30 mA cm^{-2}, and 50 mA cm^{-2} (fig. S18, A to C). An average FE of 99.3 ± 5% was obtained, indicating that the oxygen formation through the four-electron transfer during water splitting is the only electrochemical reaction over LMCF. Ir was measured as reference, for which the FE was 97.0 ± 5% (fig. S18C).

LMCF was subjected to an accelerated aging test through voltage cycling between 1.4 V and 2.0 V versus reversible hydrogen electrode (RHE) by RDE in 0.1 M $HClO_4$ electrolyte. A mere ~20-mV potential loss at 10 mA cm^{-2} was observed after 14,000 CV cycles (Fig. 3C). Such stability was found to be better than that of commercial Co_3O_4 or Ir black at comparable catalyst weight loadings (fig. S19). The morphology, composition, and electronic states of the LMCF after the voltage cycling were found to be nearly unchanged from the pristine state (fig. S20, A to E). Similarly, no appreciable changes in the XANES spectra of Co, Mn, and La were observed after the voltage cycling (fig. S20, F to H). We further tested LMCF galvanostatic stability by holding the electrode current density at 10 mA cm^{-2} for an extended operation period, following a test protocol (*23*, *24*) (Fig. 3D). The change of electrode potential in acidic electrolyte was monitored over 353 hours, and a slow degradation at an average rate of 0.28 mV $hour^{-1}$ was observed. Transient potential dips in Fig. 3D were the result of pauses of the measurement while the electrolyte was refreshed. We also checked the metal contents in the electrolyte from acid leaching by performing ICP-OES under OER electrocatalysis. A mild loss of Co, Mn, and La ions by ~1.9 wt %, ~2.8 wt %, and ~1.1 wt % over their stoichiometric loadings in the LMCF catalyst, respectively, in a period of 80 hours was observed during the chronopotentiometry at 10 mA cm^{-2} (fig. S21A). After an initial jump at the first hour, the metal dissolution rate decreased markedly afterward. Meanwhile, the catalytic activity remained nearly the same, which suggests that the dissolved metals could be attributed mainly to some loosely bound oxides at the surface. We also calculated the stability number (*S* number) for LMCF based on the amount of Co dissolved in the electrolyte at different testing times using the framework proposed by Geiger *et al.* (*25*) as well as the activity-stability factor (ASF) by Kim *et al.* (*26*). These values are compared with some of the Ir-based benchmarks in the literature, and the LMCF stability was found to be comparable to some less-stable Ir materials (Fig. 3E and fig. S21B).

Electrocatalytic activity in a PEMWE

The ultimate test of an OER catalyst is its performance in the operating PEMWE. Key properties, such as porosity, stability, and conductivity, become more important for the catalyst performance under the electrolyzer working conditions in comparison with the less-strenuous RDE measurements. The LMCF was assembled into the anode of a PEMWE single cell and tested using deionized (DI) water as the feed. Figure 3F shows composite current-voltage polarization curves derived from three measurements in the PEMWE. The electrolyzer reached a current density of 2000 mA cm^{-2} at a voltage of 2.47 V (2.20 V after cell *iR* correction), which could be further reduced to 2.30 V by switching the membrane from Nafion 115 to Nafion 212, and reached a current density of 4000 ± 200 mA cm^{-2} at 3.0 V (fig. S22). In comparison, PEMWEs with Ir loading at 0.4 mg_{Ir} cm^{-2} and 2.0 mg_{Ir} cm^{-2} at the anode displayed cell voltages of 1.93 V and 1.78 V at 2000 mA cm^{-2} before *iR* correction, respectively. Comparisons with other reported anodic catalysts are given in table S2 (*14*–*16*).

The MEA with LCMF was also subjected to accelerated stress tests (ASTs) using voltage cycling, galvanostatic, and potentiostatic methods. In the voltage cycling test, the PEMWE cell voltage was swept between 1.4 V and 1.7 V with a 10-s dwell time at each voltage (square-wave) up to 10,000 cycles. Current-voltage polarization was recorded periodically after designated voltage cycles. Figure 3G shows the polarization curves recorded after the selected voltage cycles. Figure 3H demonstrates the *iR*-corrected PEMWE cell voltages at four different current densities (100 mA cm^{-2}, 200 mA cm^{-2}, 300 mA cm^{-2}, and 400 mA cm^{-2}) after each designated cycle number taken from the tests in Fig. 3G. Only small fluctuations in the cell voltage were observed between the first and the 10,000th cycles, suggesting excellent catalyst stability under such cycling conditions. After the voltage cycling, we conducted a galvanostatic test on the same MEA at incrementally increased current densities of 50 mA cm^{-2}, 100 mA cm^{-2}, 200 mA cm^{-2}, and 300 mA cm^{-2} over a 90-hour time span (fig. S23). The cell voltages remained nearly constant after each incremental increase of the cell current until the last 10 hours, when an increasing cell voltage was observed after the current density was raised to 300 mA cm^{-2}. We also conducted a separate potentiostatic test over another PEMWE with LMCF anodic catalyst at a cell potential of 1.65 V for 100 hours (Fig. 3I). A constant cell current density of ~210 mA cm^{-2} was observed, and the anodic effluent from the PEMWE was analyzed by ICP-MS. The cobalt dissolution gradually leveled off during the first 10 hours and became negligible thereafter (fig. S24). The *S* number calculated based on the dissolution rate was found to be two orders of magnitude higher than that obtained from RDE (fig. S25A). The difference may be ascribed to the higher acidity of the electrolyte used in the RDE (pH = 1) relative to that at the MEA anode layer (pH = 2 to 4). A similar phenomenon was found in the study of RuO_2 catalyzed water oxidation (*25*). The lifetime determined by the *S* number suggests excellent durability of LMCF operated at the PEMWE (fig. S25B). The morphology and the surface structure of the catalyst were preserved after AST by RDE as well as by combined voltage cycling and galvanostatic tests in the PEMWE (fig. S26).

Active -site analysis

To understand the nature of the active site and the impact of the second and third metal doping, we performed in situ XAS of LMCF in an O_2-saturated electrolysis cell (0.1 M $HClO_4$) at the Co, Mn K-edge, and La-L_{III} edge. Figure 2A shows the XANES spectra at the Co K-edge under different OER operating potentials. At the onset potential of 1.23 V when OER has yet to commence, XANES already shows a decrease in WL intensity and an increase in pre-edge 1s → 3d peak intensity compared with the spectrum of as-prepared LMCF, accompanied by a decrease of CN of the first Co-O shell (Fig. 2D, fig. S12, and table S5). This signals oxygen loss and structural change as a result of interaction with the acidic electrolyte under the electric field, altering cobalt's centrosymmetric coordination (symmetry breakdown) and electronic structure (*27*). The CN is significantly lower than that in Co_3O_4, indicating a higher fraction of tetrahedral coordinated Co^{2+} combined with higher concentration of V_o, which serves

as the nucleophilic site in promoting O–O bond formation (*28*). As the cell potential increases to initiate OER, the 1s → 3d transition becomes more intense (Fig. 2A, inset). The average valence state of cobalt in LMCF was maintained at a lower value than that of Co_3O_4 (+2.67) under all the test potentials from +2.51 at 1.02 V to +2.32 at 1.70 V (*29*) (fig. S27). Simultaneously, Fourier transformation of EXAFS spectra shows the reduction of Co-O and Co-Co shell intensities (Fig. 2D). At a potential of 1.7 V when the OER reaction proceeds much more rapidly, XAS analysis shows shortening of the Co-O bond length (Fig. 2F and table S5), indicating a high degree of covalency contraction, which positively affects the catalytic activity of the nanoparticles (*28*). A similar phenomenon was observed in an MnO_2 film, where the Mn-O bond was shortened concomitant with accumulation of Mn^{3+} and V_o under applied potential in acid (*22*). Shortened Co-O bonds during OER were also found in the amorphous cobalt oxide OER catalyst (CoCat) (*30*), except that Co was in the +4 state, distinct from our catalyst. Accompanying the rapid acceleration of OER at 1.7 V, a clear increase of the Debye–Waller factor (DWF) σ^2 was also observed. DWF pertains to the motion of the coordinated atoms. In this case, it indicates increased Co-O vibration as the adsorbed H_2O is converted to O_2, possibly involving the shuffling lattice oxygen in LMCF, as was recently observed in a perovskite OER catalyst (*31*). The participation of surface and lattice oxygen creates anisotropic displacement of the ligation to cobalt, causing an increase of 1s → 3d transition intensity. The Co XANES and EXAFS spectra were nearly fully restored to their original intensities at the onset voltage (1.23 V) when the cell potential was returned to 1.02 V, suggesting the reversibility of the active -site restructuring in LMCF. By contrast, XANES spectra at the Mn K-edge and La L_{III}-edges show little changes under different cell potentials (Fig. 2, B and C). Compared with Co, the EXAFS spectrum of Mn only showed a minor reduction of shell peak intensity at 1.23 V after the catalyst was immersed in the electrolyte. No apparent changes in shell structure and DWF were observed for Mn during OER (Fig. 2E and fig. S13). XANES and EXAFS analyses indicate that Mn and La do not participate in the electrocatalysis directly. Rather, their presence modifies the structure and activity of the cobalt site. To this end, we also investigated the in situ XAS at different potentials over the fibrous cobalt oxide catalyst prepared by the same method but in the absence of Mn and La (CF). The changes in XANES and EXAFS are significantly less dominant than those found in LMCF (fig. S28).

Comparison with theory

The experimentally observed atomic and electronic configurations in LMCF correspond well with density functional theory (DFT)

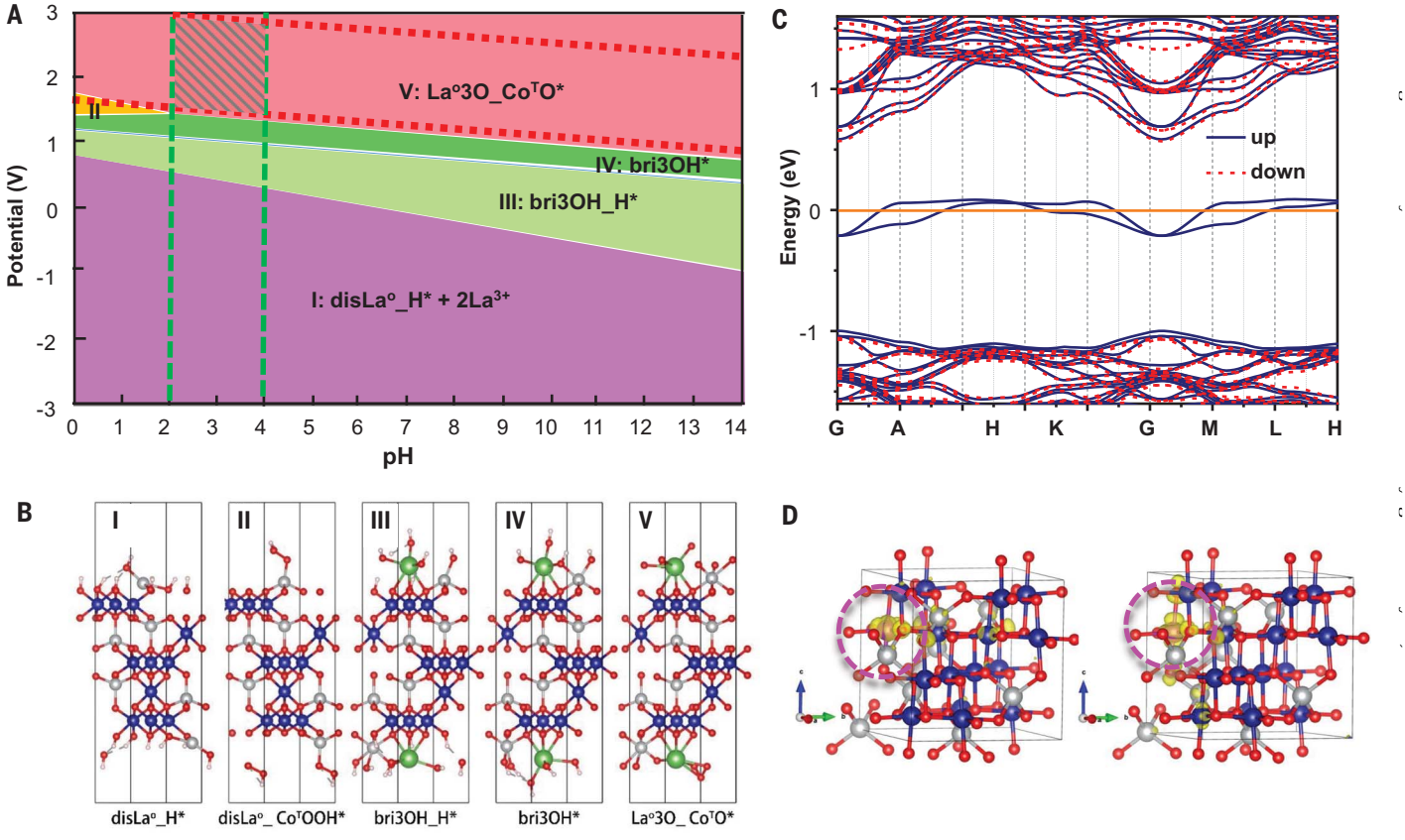


Fig. 4. Computational Pourbaix diagram and Fermi band structure of LMCF. (**A**) Surface Pourbaix diagram for the La-doped Co_3O_4 (111) facet obtained from the DFT + U calculations. (**B**) Possible intermediate state configurations. Blue, gray, red, and white balls denote the octahedral Co^{3+}, tetrahedral Co^{2+}, oxygen, and hydrogen ions. An asterisk denotes the pure surface; H* and _H* denote the configurations in which the surface oxygen atoms are covered by H; Co^O denotes the octahedral coordinated cobalt, and Co^T denotes the tetrahedral coordinated cobalt; the nO/nOH/nOOH refer to the numbers of O/OH/OOH groups covering over each Co atom; configurations denoted "disCo/La" are generated after the Co/La dissolution; and the prefix "bri-" means that the covered groups act as a bridge connecting La^O and Co^T. The potential is relative to the standard hydrogen electrode (SHE). (**C**) Calculated Fermi band structure of LMCF by replacing Co^{3+} with Mn^{3+} in the Co_3O_4 lattice. (**D**) Charge density distribution at the Fermi level upon Mn substitution in LMCF. Yellow refers to the charge density contour. Blue, gray, and red balls indicate octahedral Co^{3+}, tetrahedral Co^{2+}, and oxygen, respectively. Mn ions are behind yellow contour and circled by violet dotted line.

calculations. For example, our calculation reveals that in the bulk, in terms of total energy, both Mn and La preferentially replace Co in octahedral rather than tetrahedral sites, with stability differences of 0.25 eV and 1.28 eV, respectively. This is in agreement with the above XAS observation of the increased Co^{2+} percentage due to the Mn and La doping. The stronger preference for La is because of its most stable +3 oxidation state, whereas Mn could be present in either a +2 or +3 state. Our calculation also shows that Mn preferentially remains in the bulk of the Co_3O_4, whereas La is extruded to the surface because of its large size. We compared the relaxed atomic structures of Mn and La at the surface layers and in the bulk of Co_3O_4 (fig. S29). For Mn, the surface layer structure is merely 0.1 eV higher than that in the bulk, suggesting that it can displace Co anywhere in the system. For La, the energy is 3.09 eV lower at the surface than in the bulk. This agrees well with the STEM measurement of Mn and La distributions, as shown in Fig. 1, G and H, affirming their roles as dopants in enhancing oxide conductivity and surface defect or oxygen vacancies for better OER activity according to our design concept (*32*).

For cobalt spinel–based OER catalysts operable in a PEMWE, the most imposing challenge is stability in the acidic media. To better understand the enhanced acid tolerance of LMCF under electrolytic condition, we calculated the surface Pourbaix diagram of the La-decorated Co_3O_4 (111) facet (Fig. 4A) because it represents the dominant facet in the LMCF observed by STEM (Fig. 1, F, G, and I, and fig. S6, D and E). The Pourbaix diagram consists of five regions—I: disLa^o_H* + $2La^{3+}$, II: disLa^o_Co^TOOH* + $2La^{3+}$, III: bri3OH_H*, IV: bri3OH*, and V: La^o3O_Co^TO*—with their intermediate state configuration shown in Fig. 4B. Among these five phases, phases I and II contain ionic La^{3+} and therefore are unstable and soluble, whereas phases III, IV, and V contain the surfaces of Co^{2+}, Co^{3+}, and La^{3+} bridged by OH*, OH_H*, and O* and exhibit relative stability against Co dissolution. Under low pH and low or negative potential, the surface La in the La-doped Co_3O_4 (111) facet dissolves to form La^{3+} (phase I). At potentials between 0.86 V and 1.1 V (RHE), the cation on the surface is oxidized and stabilized by OH*, and the oxygen on the surface is covered by H* to form the stable bri3OH_H* structure (phase III). As the potential further increases to ≥1.23 V (RHE), the surface is protected by OH* to form a bri3OH* stable structure (phase IV). Under high pH and high potential, the surface of the La-doped Co_3O_4 (111) facet tends to be oxidized by water, and the surface cations La and Co are protected by O* by forming a stable structure La^o3O_Co^TO* (phase V).

The calculated Pourbaix diagram demonstrates that the stability of the LMCF catalyst is defined by the combination of cell potential and pH. It also agrees well with our experimental observations. For example, under no electric field (potential = 0) at pH < 7, the catalyst is in phase I and is thermodynamically unstable in the acidic solution. During our RDE study in 0.1 M $HClO_4$ (pH = 1) at cell potentials between 1.5 V and 2 V, only a very small fraction of tests at low cell potential will overlap with phase II, which could explain why the dissolution rate was higher in the RDE test. In our PEMWE test, the actual cell voltage runs from 1.5 V to 3.0 V (between the two red dashed lines) within the anodic pH between 2 and 4 (defined by the two vertical green dashed lines) in Fig. 4A. Therefore, the actual PEMWE anode operates in a window within these two boundaries, as marked by the area covered by the green diagonal stripes. This operating window falls within phase V, La^o3O_Co^TO*, which is stable, thus offering a preliminary explanation of the stability of the LMCF catalyst in the PEMWE.

For comparison, we also calculated the Pourbaix diagram of the pure Co_3O_4 (111) facet and confirmed its weaker stability compared with the La-doped surface (fig. S30). We furthermore calculated Pourbaix diagrams of the LMCF (110) facet (fig. S31) and (100) facet (fig. S32). Both facets offer good stability under PEMWE operating conditions according to the calculation. Our study revealed the critical role of surface La in stabilizing multiple Co_3O_4 facets against corrosion under working PEMWE condition.

Another critical issue in electrocatalysis is the inherent electron conductivity of the oxide itself. Our calculation of LMCF shows that substituting a low concentration of Co^{3+} ions with uniformly distributed Mn^{3+} ions in the Co_3O_4 lattice induces two partially occupied defect states in the midbandgap (Fig. 4, C and D). The Mn-induced electron wave function overlaps significantly with neighboring Co ions, causing obvious dispersion and hence good electron mobility. This provides a direct enhancement of bulk-based electron conductivity, which, combined with the connectivity of the nanofibrous oxide network, offers an overall high conductivity value for LMCF. The improved electronic conductivity was further confirmed experimentally by the four-probe van der Pauw method (fig. S9).

Outlook

This study offers prospective directions and design insight for the future development of PGM-free OER catalysts for hydrogen production using PEMWE technology. For example, the catalytic activity enhancement can be further explored by increasing the surface functional group density through elemental doping, primary size control, and morphology innovation. The PEMWE durability can be improved by removing the electrochemically unattached oxide, therefore limiting metal ion dissolution because of the lack of electro-potential stabilization. Fundamental understanding of the OER mechanism with respect to mononuclear versus binuclear reaction intermediates and catalytic pathways will help to guide the precursor and catalyst designs for lower overpotential and better acid tolerance (*33*, *34*). These improvements offer paths to the next-generation, PGM-free OER catalysts as viable replacements for precious metals, such as iridium.

REFERENCES AND NOTES

1. K. Ayers *et al*., *Annu. Rev. Chem. Biomol. Eng.* **10**, 219–239 (2019).
2. M. Carmo, D. L. Fritz, J. Mergel, D. Stolten, *Int. J. Hydrogen Energy* **38**, 4901–4934 (2013).
3. M. T. M. Koper, *J. Electroanal. Chem.* **660**, 254–260 (2011).
4. R. D. L. Smith *et al*., *Science* **340**, 60–63 (2013).
5. B. Zhang *et al*., *Science* **352**, 333–337 (2016).
6. Z.-F. Huang *et al*., *Nat. Energy* **4**, 329–338 (2019).
7. T. Reier, H. N. Nong, D. Teschner, R. Schlögl, P. Strasser, *Adv. Energy Mater.* **7**, 1601275 (2017).
8. Z. W. Seh *et al*., *Science* **355**, eaad4998 (2017).
9. V. Artero, M. Chavarot-Kerlidou, M. Fontecave, *Angew. Chem. Int. Ed.* **50**, 7238–7266 (2011).
10. M. W. Kanan, D. G. Nocera, *Science* **321**, 1072–1075 (2008).
11. J. B. Gerken *et al*., *J. Am. Chem. Soc.* **133**, 14431–14442 (2011).
12. J. S. Mondschein *et al*., *Chem. Mater.* **29**, 950–957 (2017).
13. B. Rodríguez-García *et al*., *Sustain. Energy Fuels* **2**, 589–597 (2018).
14. M. Chatti *et al*., *Nat. Catal.* **2**, 457–465 (2019).
15. I. A. Moreno-Hernandez *et al*., *Energy Environ. Sci.* **10**, 2103–2108 (2017).
16. M. Blasco-Ahicart, J. Soriano-López, J. J. Carbó, J. M. Poblet, J. R. Galan-Mascaros, *Nat. Chem.* **10**, 24–30 (2018).
17. J. Bao *et al*., *Angew. Chem. Int. Ed.* **54**, 7399–7404 (2015).
18. S. L. Zhang *et al*., *Adv. Mater.* **32**, 2002235 (2020).
19. J. Shui, C. Chen, L. Grabstanowicz, D. Zhao, D. J. Liu, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10629–10634 (2015).
20. W. Wang, A. Janotti, C. G. Van de Walle, *J. Mater. Chem. C* **4**, 6641–6648 (2016).
21. Y. Wang *et al*., *Adv. Energy Mater.* **4**, 1400696 (2014).
22. Z. Morgan Chan *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, E5261–E5268 (2018).
23. L. C. Seitz *et al*., *Science* **353**, 1011–1014 (2016).
24. C. C. McCrory *et al*., *J. Am. Chem. Soc.* **137**, 4347–4357 (2015).
25. S. Geiger *et al*., *Nat. Catal.* **1**, 508–515 (2018).
26. Y. T. Kim *et al*., *Nat. Commun.* **8**, 1449 (2017).
27. C. H. van Oversteeg, H. Q. Doan, F. M. de Groot, T. Cuk, *Chem. Soc. Rev.* **46**, 102–125 (2017).
28. H. N. Nong *et al*., *Nat. Catal.* **1**, 841–851 (2018).
29. H. Dau, P. Liebisch, M. Haumann, *Anal. Bioanal. Chem.* **376**, 562–583 (2003).
30. M. Risch *et al*., *Energy Environ. Sci.* **8**, 661–674 (2015).
31. A. Grimaud *et al*., *Nat. Chem.* **9**, 457–465 (2017).
32. O. Diaz-Morales *et al*., *Nat. Commun.* **7**, 12363 (2016).
33. I. C. Man *et al*., *ChemCatChem* **3**, 1159–1165 (2011).
34. M. Busch, *Curr. Opin. Electrochem.* **9**, 278–284 (2018).
35. W. Xu *et al*., La and Mn-doped cobalt spinel oxygen evolution catalyst for proton exchange membrane electrolysis, dataset, Dryad (2023); https://doi.org/10.5061/dryad.76hdr7t1v.

ACKNOWLEDGMENTS

We thank J. Bareno, C. Yang, B. Fisher, M. S. Ferrandon, D. J. Myers, D. Abraham, and J. Wang of Argonne National Laboratory and S. Kabir of Giner Inc. for experimental assistance and comments on the manuscript. **Funding:** This work is supported by the US Department of Energy (DOE), Office of Energy Efficiency and Renewable Energy, Hydrogen and Fuel Cell Technologies Office (D. Peterson, project manager), and by Laboratory Directed Research and Development (LDRD) funding of Argonne National Laboratory, provided by the Director, Office of Science, of the US DOE under contract no. DEAC02-06CH11357 through a Maria Goeppert Mayer Fellowship to L.C. Work performed at the Center for Nanoscale Materials and Advanced Photon Source, both US

DOE Office of Science User Facilities, was supported by the US DOE, Office of Basic Energy Sciences, under contract no. DE-AC02-06CH11357. The work at Lawrence Berkeley National Laboratory was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy of the US DOE under the Hydrogen Generation program. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the US DOE's National Nuclear Security Administration under contract no. DE-NA0003525. This paper describes objective technical results and analysis. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States government or any agency thereof; neither the United States government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, nor usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. **Author contributions:** D.-J.L. designed and supervised the experiment. L.C. designed and synthesized the catalyst and conducted electrochemical measurement and data analysis. L.C., Ha.X., D.-J.L., Z.G., and Hu.X. conducted MEA preparation and PEMWE measurements. J.W. and J.D.S. performed electron microscopy imaging and analysis. L.C., Ha.X., A.J.K., W.X., X.-M.L., and D.-J.L. conducted spectroscopic and catalyst structural investigation and data analysis. G.G., H.L., and L.-W.W. performed DFT calculation. L.C., L.-W.W., and D.-J.L. wrote the manuscript. **Competing interests:** A US patent (USP 11,033,888) on the nanofibrous catalyst for OER with D.-J.L. and L.C. as the coinventors was granted to UCHICAGO ARGONNE, LLC. The authors declare no other competing interests. **Data and materials availability:** Data used for Pourbaix diagram calculations on La-doped cobalt spinel are available from Dryad (*35*). Other data are available in the main text and the supplementary materials. **License information:** Copyright © 2023 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.ade1499
Materials and Methods
Figs. S1 to S32
Tables S1 to S8
References (*36*–*70*)

Submitted 31 July 2022; resubmitted 13 February 2023
Accepted 12 April 2023
10.1126/science.ade1499

PALEONTOLOGY

A macroevolutionary pathway to megaherbivory

Oscar Sanisidro[1]*, Matthew C. Mihlbachler[2,3], Juan L. Cantalapiedra[1]

Several scenarios have been proposed to explain rapid net size increases in some early Cenozoic mammalian lineages: sustained and gradual directional change, successive occupation of adaptive zones associated with progressively larger body sizes, and nondirectional evolution associated with branching events in combination with higher diversification potential of the larger lineages. We test these hypotheses in brontotheres, which are among the first radiations of mammals that consistently evolved multitonne sizes. Body-mass evolution in brontotheres mainly occurred during speciation and had no preferential direction. Long-term directional change stemmed from the higher survival of larger lineages in less-saturated herbivore guilds. Our study emphasizes the role of differential species proliferation in explaining the long-term phenotypic trends observed in the fossil record, which are more than an accumulation of steady microevolutionary changes.

The diversification of body sizes during the early Cenozoic is one of the most extraordinary features of mammalian evolution (*1*). After the Cretaceous-Tertiary extinction event, mammalian body masses rarely exceeded 10 kg (*2*–*4*). Twenty million years later, megaherbivores (≥1000 kg) were keystone constituents of Cenozoic assemblages. The emergence of megaherbivores had a profound influence on terrestrial ecosystem functioning by promoting new interactions among herbivore guilds and plant communities (*5*), which triggered the evolution of larger carnivores (*6*), and restructuring energy and biomass flows among trophic levels (*7*). Yet the evolutionary processes underpinning rapid size increases in lineages are little understood. For example, the rate of body-mass change within populations (stemming from changes in allele frequencies through natural selection, i.e., microevolution) is substantially faster than the rates observed in the fossil record (*8*). This decoupling suggests that evolutionary change over deep time is not just an extrapolation of microevolutionary processes over millions of years (*9*) and points to a crucial contribution of processes operating at the species level and above [e.g., differential species proliferation, i.e., macroevolution (*10*)].

Cope's neo-Lamarckian views were inspired by his early study of the mammalian fossil record. His "law of the unspecialized" posed that the interplay between organisms' habits and environment propels lineages in parallel progressive lines toward specialization and size increases (*11*). Later on, the size aspect of the law was emphasized (*12*), rebranded ["Cope's rule" (*13*)], and recast in Darwinian terms: Selective advantages of a large size (*14*) operating over broad timescales should render gradual changes within lineages—phyletic evolution—and definitive adaptive trends toward larger sizes (*2*) (Fig. 1A). Alternative reformulations rejected rectilinear evolution (definite, straight trends) and proposed that increasing-size trends are the result of lineages gradually adjusting to successive regimes of natural selection or adaptive zones (*15*, *16*) (Fig. 1A).

In a later reinterpretation of Cope's rule, Stanley (*17*) deviated from views that understood macroevolution as a long-term accumulation of gradual directional changes. Inspired by the model of punctuated equilibrium (*18*), Stanley predicted that changes in body mass mainly occur during speciation events and that such "speciational evolution" has no preferential direction (larger and smaller species are evenly produced during branching events). The model starts with a small ancestor in close proximity to a functional or physiological size limit. Niches associated with smaller sizes are filled first, and larger lineages would show higher speciation potential and suppressed extinction because large-size guilds are initially less saturated. The pattern is expected to arise in major groups (e.g., mammals) but also explains the net size increases observed in some subclades (*17*, *19*) (Fig. 1, A and B). Which of the three presented models (Fig. 1A) better explains the rapid net size growth of some early Cenozoic mammalian clades? This question remains unanswered because we lack quantitative assessments of the modes of mammalian body-mass evolution during the early Cenozoic based on phylogenetic evolutionary models. Further, the implications of these alternative views on size evolution relate to an essential question in evolutionary biology: whether microevolution alone or macroevolution (including sorting or selection at the species level) shapes evolutionary patterns over deep time.

In this study, we investigate the well-known fossil record of the brontotheres (or titanotheres; superfamily Brontotheriioidea), an extinct Eocene perissodactyl lineage that underwent one of the most extreme net size increases seen among mammals. The first known stem brontotheres were around 18 kg in size [Fig. 1B; based on dental measurements (*20*)]. Yet during their 22-million-year history, more than half of the 57 known brontothere species had estimated masses of more than 1000 kg, posing a valuable example of efficient diversification into the megaherbivore guild. The definite size increase of brontotheres became one of the archetypical examples of rectilinear evolution during the early 20th century (*21*).

Our approach combines phylogeny-based trait evolution modeling (*22*–*24*) and diversification analyses (*25*) (fig. S1). The former is used to test three main scenarios of net body-mass evolution in brontotheres: steady size increase in lineages (or Cope's rule) (*13*), adjustment to one or successive adaptive zones related to progressively larger body sizes (*15*), and speciational evolution without preferential direction (*17*) [Fig. 1A and (*20*)]. Conversely, diversification analyses allow us to test for species sorting, that is, a correlation of extinction and speciation rates with body mass and other size-modulated ecological variables such as ecological niche saturation and predation pressure. The basis of our analyses comprises 1000 trees obtained through tip-dating procedures (fig. S2), mass estimates derived from dental dimensions, and an updated dataset of 276 fossil occurrences of brontotheres (*20*).

The evolution of body size in brontotheres was nondirectional and had a dominant speciational component (Fig. 1A). A model where change mainly occurs at speciation events without a preferential direction (*26*) is at least four times more probable than the other scenarios (Fig. 1C and table S2). Yet the non-negligible performance of a gradualistic version of Stanley's model (that is, a nondirectional model of gradual evolution; table S2) suggests that phyletic evolution likely operated in some regions of the tree. Our results are robust toward sampling biases. A series of simulations shows that the signal of speciational evolution is only recovered in densely sampled fossil phylogenies, whereas a gradualistic model of evolution is wrongly preferred because fewer tips are sampled in simulated clades [fig. S21 and (*20*)]. The unbiased nature of size evolution in the group is especially exemplified by the recurrent evolution of smaller-sized brontotheres (some even less than 100 kg) until the latest Eocene (Fig. 1B). Evolutionary leaps are the norm across the clade, as shown by the fact that the speciational model is preferred even after these smaller forms are removed from the analyses (table S3).

A speciational model of evolution is further supported by independent evidence revealing that net intraspecific mass changes—stemming

[1]Global Change Ecology and Evolution Research Group (GloCEE), Department of Life Sciences, Universidad de Alcalá, 28805 Alcalá de Henares, Madrid, Spain. [2]Department of Anatomy, New York Institute of Technology College of Osteopathic Medicine, Old Westbury, NY 11568, USA. [3]Division of Paleontology, American Museum of Natural History, New York, NY 10024, USA.
*Corresponding author. Email: oscarsanisidro@gmail.com

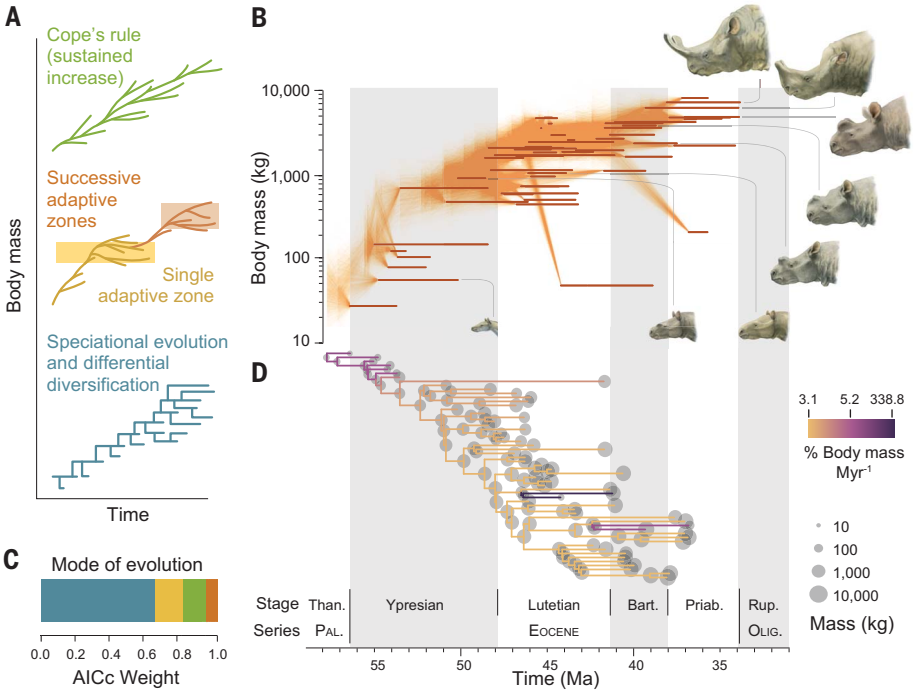


Fig. 1. The mode of body-mass evolution in brontotheres. (**A**) Hypothetical phenograms representing the main evolutionary models of net body-mass increment. Shown from top to bottom are gradual, sustained, and generalized increase in body mass (Cope's rule); size evolution across one or more successive adaptive zones with increasing optimal sizes (as in Simpson); and speciational evolution where change has no preferential direction (speciation events moving upward and downward are equal in number) and net size increases need differential survival and speciation potential. See (*20*) for details on the corresponding statistical models. (**B**) Phenogram showing brontothere body-mass evolution through time. The shaded areas reflect uncertainty in ancestral size reconstructions from fossilBM. (**C**) Support for different evolutionary models based on corrected Akaike information criterion (AICc) weights. (**D**) Rates of body mass evolution estimated by fossilBM. Circles at the tips show body mass. Bart., Bartonian; Olig., Oligocene; Pal., Paleocene; Priab., Priabonian; Rup., Rupelian; Than., Thanetian; Ma, million years ago; Myr, million years.

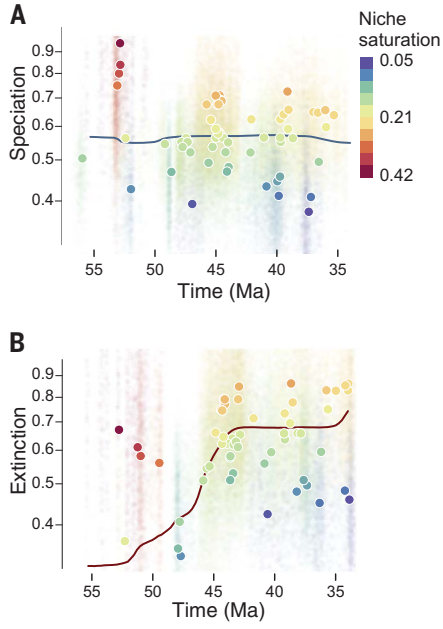


Fig. 2. Effect of niche saturation on the diversification of brontotheres. (**A** and **B**) Effect of niche saturation on speciation (A) and extinction (B) rates. Circles represent the average expected rate of each species and time of speciation and extinction in each case. Lighter background shades represent the speciation and extinction values for each brontothere species across the posterior of the Markov chain Monte Carlo run. Baselines (blue for speciation, red for extinction) are the expected rate for a species with the mean trait value.

from sustained increases or gradual phenotypic divergence—were rare in brontotheres. Instead, intraspecific mass evolution hovered around stable means in species recorded through several geological stages (*27*) (fig. S22). Overall, the observed pattern is congruent with phenotypic evolution being mostly condensed in events of peripatric speciation (*17*, *18*), where phenotypic sorting through drift or adaptation to immediate conditions is geologically rapid (*28*), and decoupled in direction from macroevolutionary trends (*19*). The signal of this discontinuity is also demonstrated by the weak reconstructed trend and attraction parameters in our phylogenetic models [trend parameter (μ) = 0 and attraction parameter (α) = 0.002, respectively; table S2]. In fossilBM, a Bayesian approach that captures tree-wide variation in rates and trends (*24*), all posterior distributions of the trend parameters largely overlap with zero (fig. S18).

The speciational model alone cannot explain the 150-fold net size increase observed in brontotheres (from around 20 kg to more than 3 tonnes). Processes where phenotypic traits drive sorting at the species level have been proposed to account for this apparent contradiction (*19*, *29*). We tested the impact of body mass on speciation and survival rates and also explored the role of two size-related ecological traits in species sorting: the niche saturation experienced by each species based on its size (as in Stanley's hypothesis) and species-specific predation pressure (*30*). These traits are based on body-mass information of 680 species of herbivorous and carnivorous taxa that co-occurred with each brontothere species at stratigraphic and continental levels (*20*). Although these ecological traits are expected to correlate with body size, they should better reflect specific biotic pressures undergone by each species during its evolutionary history.

We found that niche saturation was the best explanatory variable for diversification biases in brontotheres (table S4). According to this model, species in more-packed niches underwent higher volatility (higher speciation and disproportionately high extinction), mainly during the early stages of brontothere evolution (Fig. 2; importantly, our models consider potential trait-dependent sampling rates). Nevertheless, early brontotheres show large leaps in mass evolution (Fig. 1D), which were key to rapidly producing larger lineages that occupied less-saturated guilds. These new species show less capacity to produce species, but their extinction risk was far less severe (Fig. 2B; most of these species made it to the end of the Ypresian, some 48 million years ago). Early rapid saltations in body mass ensured the survival of brontothere species that popped up in less-packed niches, setting the initial course for subsequent size trends in the group.

From then on, brontothere species were able to avoid the most-filled herbivore guilds. Yet species in more-saturated niches still show speciation rates over the baseline rate (Fig. 2A), as reflected by the repeated emergence of small forms (Fig. 1B). Fast adaptation to local conditions could still render speciation into the smaller-sized-herbivore guilds of the Eocene. However, these highly saturated adaptive landscapes were macroevolutionary minefields for brontotheres, as revealed by the inflated extinction risk of the species associated with more-crowded guilds in the middle and late Eocene. Conversely, lineages in less-saturated ecological niches show a substantial increase

in survival probability until the end of the Eocene (Fig. 2B), further enhancing the macroevolutionary trend toward larger sizes.

Despite extinction suppression in lineages associated with less-crowded niches, the baseline extinction rate shows that overall lineage survival decreased toward the second half of the group's history (Fig. 2B and fig. S16). Classic notions advocated that the attainment of extreme large sizes was linked to overspecialization and eventually led to degeneracy and extinction (*11*, *12*), but others have proposed that brontotheres just could not keep up with the changing environment (*31*, *32*). We see an overall deceleration of body-mass evolution in brontotheres [$P < 0.001$; correlation coefficient (r) = −0.413; based on branch-specific rates from fossilBM; fig. S19A], which could be interpreted as a signal of intrinsic evolutionary stagnation. However, rates of anatomical evolution (estimated in our tip-dating approach based on characters in the cladistic matrix) did not track rates of body mass evolution ($P = 0.595$, $r = 0.051$; fig. S19D) and showed no signal of deceleration with time ($P = 0.107$; $r = -0.153$; fig. S1B9), showing that brontotheres had not exhausted their evolutionary potential. Rather, the decelerating rates of mass evolution could be the result of biomechanical or physiological constraints, which would also explain why large size leaps during the latest part of brontothere history only took place toward small sizes (Fig. 1B) but are also in line with longer generation times slowing down the pace of evolution (*9*).

More likely, the overall decline in brontothere diversity and extinction was driven by changes in climate and vegetation (*32*). The second half of the Eocene witnessed a stepwise transition toward drier and less-forested conditions in the Northern Hemisphere (*33*), bringing about unprecedented selective pressures for the last browsing, forest-dwelling brontotheres (*34*) (including further crowding of herbivore guilds at larger sizes). Further study, including climate and niche modeling as well as extensive dietary proxies, would clarify how environmental change led to the demise of the brontotheres.

Our results portray interconnected patterns of phenotypic and diversity trends in brontotheres acting at three different levels (*35*). First, body-mass evolution during peripatric speciation rapidly increased size disparity in brontotheres. Major size changes took place among the earliest forms, stemming from high macroevolutionary volatility in smaller lineages (Fig. 2). This finding contrasts with views that relate size increases in mammals with notions of progress and dominance. At a second level, ecology-mediated differential survival of species reshapes the available body mass disparity. It was this macroevolutionary process that rendered the observed net increase in brontothere body mass, which cannot be explained by upward extrapolation from gradual processes over millions of years. Yet the survival of the clade is not ensured by the enhanced survival of larger species in less-saturated niches. On a third tier, we find processes acting above the species level that dictate the decline and replacement of entire lineages, spurred by unpredictable changes in the physical environment. This last level of macroevolution marks the coming and going of clades and defines vast evolutionary trends over tens or hundreds of millions of years.

REFERENCES AND NOTES

1. F. A. Smith *et al*., *Science* **330**, 1216–1219 (2010).
2. J. Alroy, *Science* **280**, 731–734 (1998).
3. F. A. Smith, R. E. E. Smith, S. K. Lyons, J. L. Payne, *Science* **360**, 310–313 (2018).
4. T. R. Lyson *et al*., *Science* **366**, 977–983 (2019).
5. C. N. Johnson, *Proc. Biol. Sci.* **276**, 2509–2519 (2009).
6. J. O. Farlow, T. R. J. Holtz Jr., *Paleontol. Soc. Pap.* **8**, 251–266 (2002).
7. J. Sitters *et al*., *Glob. Change Biol.* **26**, 2060–2071 (2020).
8. P. D. Gingerich, *Contrib. Mus. Paleontol. Univ. Mich.* **26**, 197–225 (1983).
9. A. R. Evans *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **109**, 4187–4190 (2012).
10. S. J. Gould, *The Structure of Evolutionary Theory* (Belknap Press, 2002).
11. E. D. Cope, *The Primary Factors of Organic Evolution* (Open Court Publishing Co., 1896).
12. C. Depéret, *Les Transformations du Monde Animal* (FB&C Ltd., 1907).
13. B. Rensch, *Evolution* **2**, 218–230 (1948).
14. J. H. Brown, B. A. Maurer, *Nature* **324**, 248–250 (1986).
15. G. G. Simpson, *Tempo and Mode in Evolution* (Columbia Univ. Press, 1944).
16. R. B. J. Benson, R. A. Frigot, A. Goswami, B. Andres, R. J. Butler, *Nat. Commun.* **5**, 3567 (2014).
17. S. M. Stanley, *Evolution* **27**, 1–26 (1973).
18. N. Eldredge, S. J. Gould, in *Models in Paleobiology*, T. J. M. Schopf, Ed. (Freeman, Cooper, and Company, 1972), pp. 82–115.
19. S. M. Stanley, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 646–650 (1975).
20. Materials and methods are available as supplementary materials.
21. H. F. Osborn, *Science* **74**, 557–559 (1931).
22. M. Pennell *et al*., *Bioinform.* **30**, 2216–2218 (2014).
23. J. M. Beaulieu, D.-C. Jhwueng, C. Boettiger, B. C. O'Meara, *Evolution* **66**, 2369–2383 (2012).
24. D. Silvestro *et al*., *Syst. Biol.* **68**, 78–92 (2018).
25. D. Silvestro, N. Salamin, J. Schnitzler, *Methods Ecol. Evol.* **5**, 1126–1131 (2014).
26. M. Pagel, *Zool. Scr.* **26**, 331–348 (1997).
27. E. S. Vrba, *S. Afr. J. Sci.* **76**, 61–67 (1980).
28. S. Baleka *et al*., *Curr. Biol.* **31**, 3606–3612.e7 (2021).
29. E. S. Vrba, *Syst. Zool.* **33**, 318–328 (1984).
30. B. Van Valkenburgh, M. W. Hayward, W. J. Ripple, C. Meloro, V. L. Roth, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 862–867 (2016).
31. H. F. Osborn, *Mon. U.S. Geol. Surv.* **55**, 1–953 (1929).
32. L. S. Russell, *Can. J. Earth Sci.* **10**, 140–145 (1973).
33. E. A. Bestland, G. J. Retallack, C. C. I. Swisher III, *J. Geol.* **105**, 153–172 (1997).
34. A. Zanazzi, M. J. Kohn, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **257**, 22–37 (2008).
35. S. J. Gould, *Paleobiology* **11**, 2–12 (1985).
36. O. Sanisidro, M. C. Mihlbachler, J. L. Cantalapiedra, A macroevolutionary pathway to megaherbivory. Figshare (2023); https://doi.org/10.6084/m9.figshare.20222814.

ACKNOWLEDGMENTS

We acknowledge D. Silvestro, J. Barido-Sottani, and T. R. Simões for their technical support and G. Navalón, I. A. Lazagabaster, and three anonymous reviewers for valuable comments on the manuscript. **Funding:** This work was funded by the Talent Attraction Program of the Madrid government and Universidad de Alcalá grant 2017-T1/AMB5298. **Author contributions:** Conceptualization: O.S., M.C.M., J.L.C.; Methodology: O.S., M.C.M., J.L.C.; Visualization: O.S., J.L.C.; Funding acquisition: J.L.C.; Writing – original draft: O.S., M.C.M., J.L.C.; Writing – review and editing: O.S., M.C.M., J.L.C. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data, code, and materials used in the analysis can be found in data S1 and at figshare (*36*). **License information:** Copyright © 2023 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.ade1833
Materials and Methods
Supplementary Text
Figs. S1 to S22
Tables S1 to S5
References (*37*–*274*)
MDAR Reproducibility Checklist
Data S1

Submitted 29 July 2022; accepted 4 April 2023
10.1126/science.ade1833

ANCIENT GENOMES

Natural products from reconstructed bacterial genomes of the Middle and Upper Paleolithic

Martin Klapper[1]†, Alexander Hübner[2,3]†, Anan Ibrahim[1]†, Ina Wasmuth[1], Maxime Borry[2], Veit G. Haensch[4], Shuaibing Zhang[1], Walid K. Al-Jammal[5], Harikumar Suma[1], James A. Fellows Yates[1,2,3], Jasmin Frangenberg[1], Irina M. Velsko[2], Somak Chowdhury[1], Rosa Herbst[1], Evgeni V. Bratovanov[4], Hans-Martin Dahse[6], Therese Horch[4], Christian Hertweck[4,7], Manuel Ramon González Morales[8], Lawrence Guy Straus[9,10], Ivan Vilotijevic[5], Christina Warinner[2,3,7,11]*, Pierre Stallforth[1,5]*

Major advances over the past decade in the field of ancient DNA are providing access to past paleogenomic diversity, but the diverse functions and biosynthetic capabilities of this growing paleome remain largely elusive. We investigated the dental calculus of 12 Neanderthals and 52 anatomically modern humans ranging from 100,000 years ago to the present and reconstructed 459 bacterial metagenome-assembled genomes. We identified a biosynthetic gene cluster shared by seven Middle and Upper Paleolithic individuals that allows for the heterologous production of a class of previously unknown metabolites that we name "paleofurans." This paleobiotechnological approach demonstrates that viable biosynthetic machinery can be produced from the preserved genetic material of ancient organisms, allowing access to natural products from the Pleistocene and providing a promising area for natural product exploration.

Microbial natural products play key roles in cell signaling, communication, defense, and microbial evolution, enabling the interaction between microorganisms within communities and with their abiotic environments (*1*, *2*). Because of their high structural and functional diversity, they are an ideal source of therapeutic drugs, antimicrobials, and other functional compounds (*3*, *4*). Characterizing the natural products encoded in biosynthetic gene clusters (BGCs) and synthesized by ancient microbial communities would provide insights into past microbial lifestyles and enable access to previously hidden chemical, structural, and functional diversity. However, the direct detection and characterization of metabolites from ancient microbiota has met with only limited success (*5*). Small molecules are unlikely to persist in archaeological samples over long periods because they leach out and have a propensity to degrade and undergo modification over time. With the advent of paleogenomics, it is becoming feasible to access the vast structural and functional diversity of ancient low-molecular-weight natural products by instead resurrecting biosynthetic pathways from ancient metagenomes.

Since the publication of the first draft genome of an ancient bacterium in 2011 (*6*), major advances in paleogenomic technologies have resulted in the characterization of more than 1700 ancient microbial genomes, partial genomes, and metagenomes from diverse pathogens, commensals, culinary taxa, and microbiota (*7*). Yet, knowledge of the diverse functions and biosynthetic capabilities encoded within this growing microbial paleome lags behind. Most studies in the field of ancient microbial research have focused on sequence alignment to databases of reference genes or genomes, thus restricting findings to known taxa and their close relatives (*8*, *9*). Recent de novo assembly of authentic ancient DNA (aDNA) has shown that reconstruction of ancient metagenome-assembled genomes (MAGs) is possible (*10*–*13*). However, further optimization of assembly methods and exploration of deeply sequenced ancient metagenomes is needed to understand the unknown functional space and biosynthetic capabilities encoded in the microbial paleome. The development of computational tools for BGC prediction from (meta)genomes has accelerated natural product discovery through identification of gene targets for heterologous expression (*14*–*16*). However, this requires high-quality genomic data consisting of long contiguous sequences (contigs), which are difficult to assemble from fragmentary aDNA (*17*) with currently available methods (*11*–*13*).

De novo reconstruction of ancient metagenome-assembled genomes (MAGs)

To search for ancient BGCs, we selected previously published dental calculus metagenomic datasets (*9*) from 12 Neanderthals dating from ~102,000 to 40,000 years ago, 34 archaeological humans dating from ~30,000 to 150 years ago, and 18 present-day humans (Fig. 1A and dataset S1a) and generated additional sequencing data for a subset of 23 individuals predating 8000 years ago to improve metagenomic sequence coverage within this set (dataset S1b). For each individual, we obtained between 3.1 million and 338.3 million DNA molecules (dataset S1c) with a mode fragment length of 30 to 48 base pairs (bp) for archaeological sequences (mean, 30 bp) (Fig. 1B and dataset S1d). We removed five low-quality samples with fewer than 5 million sequences and then considered assembler performance on ultrashort sequences and performed de novo assembly of the metagenomic datasets using MEGAHIT (*18*) for Neanderthals and ancient humans and metaSPAdes (*19*) for present-day humans. For MEGAHIT assemblies, we observed a persistent artifact in which damage-related miscoding lesions were overrepresented in contig sequences, and we corrected this error by replacing the damaged allele extracted from the assembly graph with the allele inferred from FreeBayes (*20*) from short-read alignment to the contig sequence (fig. S1). We obtained 522,874 and 305,133 contigs >1 kb in length for the present-day and ancient metagenomes, respectively, of which 21,070 and 8165 contigs exceeded 10 kb (dataset S2a). Although on average the de novo assembly of present-day dental calculus metagenomes resulted in longer contigs (fig. S2 and dataset S2b), the assembly of a subset of deeply sequenced Pleistocene metagenomes (PES001, EMN001, and PLV001) performed similarly to present-day metagenomes (Fig. 1C), and the majority of the contigs belonging to these individuals were inferred to be of ancient origin by means of PyDamage (fig. S3 and dataset S2c) (*13*).

After assembly, we performed reference-free binning of all contigs longer than 1 kb (dataset S3a) and MAG refinement based on single-copy marker genes by using MetaWRAP (dataset S3b) (*21*). We next implemented an automated pipeline to identify and remove chimeric contigs (figs. S4 and S5 and dataset S3c). Requiring >50% genomic completeness and <10%

[1]Department of Paleobiotechnology, Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute, 07745 Jena, Germany. [2]Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany. [3]Associated Research Group of Archaeogenetics, Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute, 07745 Jena, Germany. [4]Department of Biomolecular Chemistry, Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute, 07745 Jena, Germany. [5]Institute of Organic Chemistry and Macromolecular Chemistry, Friedrich Schiller University Jena, 07743 Jena, Germany. [6]Department of Infection Biology, Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute, 07745 Jena, Germany. [7]Faculty of Biological Sciences, Institute of Microbiology, Friedrich Schiller University Jena, 07743 Jena, Germany. [8]Instituto Internacional de Investigaciones Prehistóricas de Cantabria, Universidad de Cantabria, 39071 Santander, Spain. [9]Department of Anthropology, University of New Mexico, Albuquerque, NM 87131, USA. [10]Grupo I+D+i EvoAdapta, Departmento de Ciencias Históricas, Universidad de Cantabria, 39005 Santander, Spain. [11]Department of Anthropology, Harvard University, Cambridge, MA 02138, USA.
*Corresponding author. Email: pierre.stallforth@leibniz-hki.de (P.S.); christina_warinner@eva.mpg.de (C.W.) †These authors contributed equally to this work.

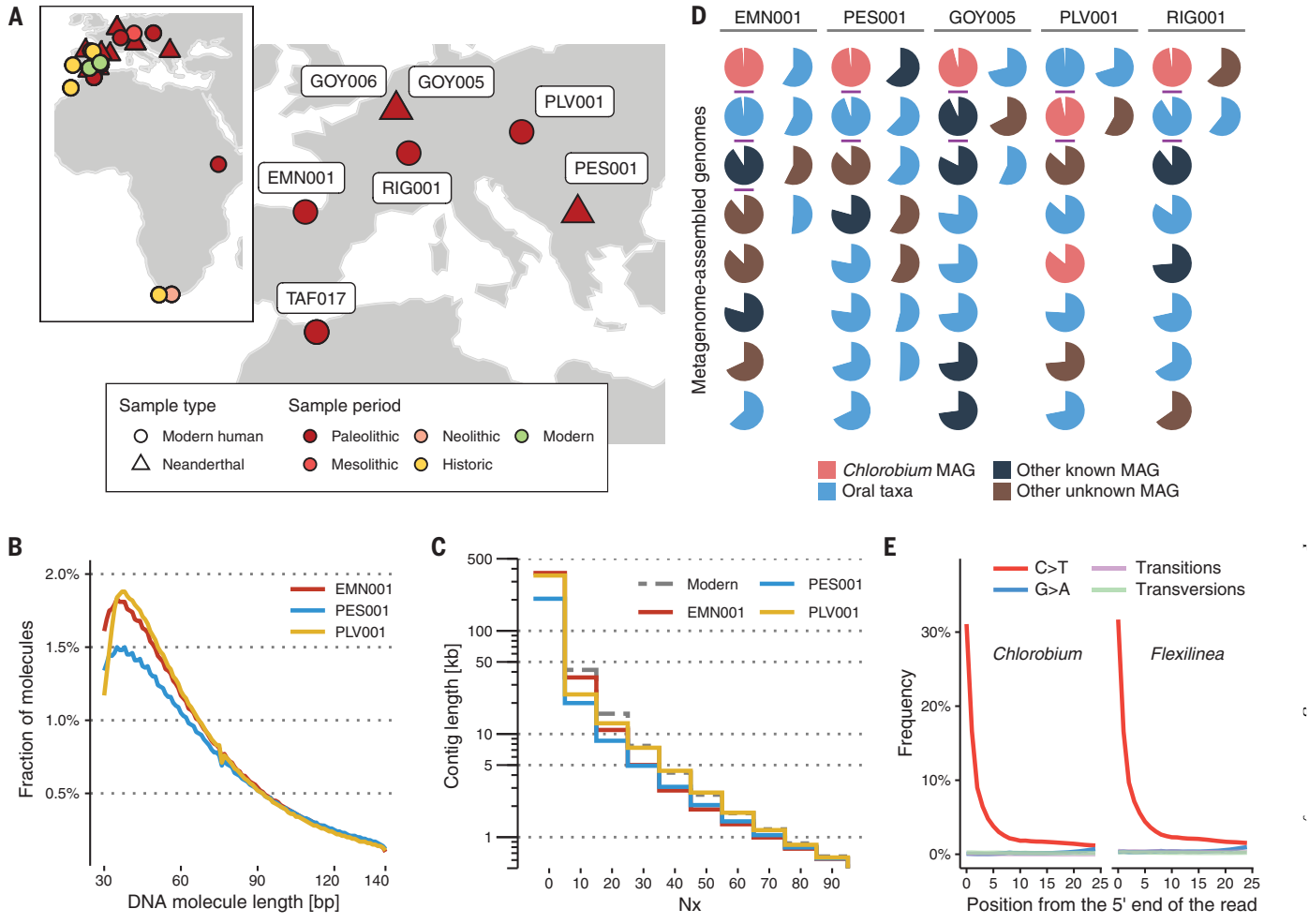


Fig. 1. De novo assembly and reference-free binning of ancient metagenome-assembled genomes (MAGs) from dental calculus samples. (**A**) Sample locations and temporal overview. (Inset) All analyzed samples. (Right) The larger map displays seven samples for which we were able to reconstruct >50% of the *Chlorobium* genome. (**B**) DNA length distribution of EMN001, PES001, and PLV001 metagenomes estimated from paired-end 75-bp Illumina-sequenced read pairs. The mode fragment length was <40 bp, and only 12.4, 21.4, and 12.6% of read pairs were longer than the maximum inferrable length of 140 bp (dataset S1b). (**C**) Contig length distribution of the longest contig (N0) and at the nine deciles (N10 to N90) for EMN001, PES001, and PLV001 compared with the average contig length of present-day dental calculus samples. Nx, deciles of the contig length distribution. (**D**) Summary of all MAGs recovered from dental calculus samples containing a HQ *Chlorobium* MAG. Pies indicate the completeness estimated by CheckM, horizontal bars indicate the HQ MAGs, and colors indicate the inferred MAG source. (**E**) Graph showing 5′ substitution frequency of EMN001 short-read data aligned to *Chlorobium* and *Flexilinea* contig consensus sequences. High frequencies of C-to-T transitions relative to other possible substitutions indicate age-related damage.

contamination as estimated with CheckM (*22*) and a clade separation score of <0.45 as estimated with the Genome UNClutterer (GUNC) (*23*), and following minimum information about a metagenome-assembled genome (MIMAG) criteria (*24*), we obtained a total of 459 MAGs, of which 150 high-quality (HQ) and 174 medium-quality (MQ) MAGs originated from present-day metagenomes and 25 HQ and 110 MQ MAGs originated from archaeological metagenomes (dataset S3c). However, both the modern and archaeological HQ MAGs lacked the required ribosomal RNA genes (dataset S3d), which is typical of MAGs assembled from short-read metagenomic sequencing data alone (*25*, *26*). Of the 135 archaeological MAGs, 22 HQ and 104 MQ MAGs were confirmed with PyDamage (*13*) to have characteristic age-related DNA damage (dataset S3c, q value < 0.05), indicating that the archaeological MAGs are predominantly ancient in origin and not recent contaminants.

From the 459 MAGs, we identified a set of 226 representative MAGs (dataset S3e) by clustering at an average nucleotide identity (ANI) of 95%; we then performed taxonomic classification using the Genome Taxonomy Database (GTDB). We next determined whether the representative MAGs were likely to have an oral origin by clustering them with all reference genomes present within the Human Oral Microbiome Database (HOMD) (fig. S6) (*27*). The majority of MAGs (75.4%) were found to be closely related to known oral taxa, although this is likely an underestimate because several taxa not identified as oral by this method were nevertheless assigned by means of PhyloPhlAn3 (*28*) to taxa with oral isolation sources (dataset S3e). We observed overlap in the representative MAGs found in both present-day and archaeological dental calculus, such as the oral taxa *Arachnia propionica* (n = 16), *Flexilinea sp001717545* (n = 15), and *Lautropia mirabilis* (n = 14) (dataset S3f), but some oral MAGs were recovered only from present-day (for example, *Capnocytophaga*) or archaeological (for example, *Methanobrevibacter*) metagenomes. Oral MAGs were well represented among many of the oldest Pleistocene metagenomes in our study (Fig. 1D), and at least nine MAGs of secure oral origin were reconstructed from a Neanderthal

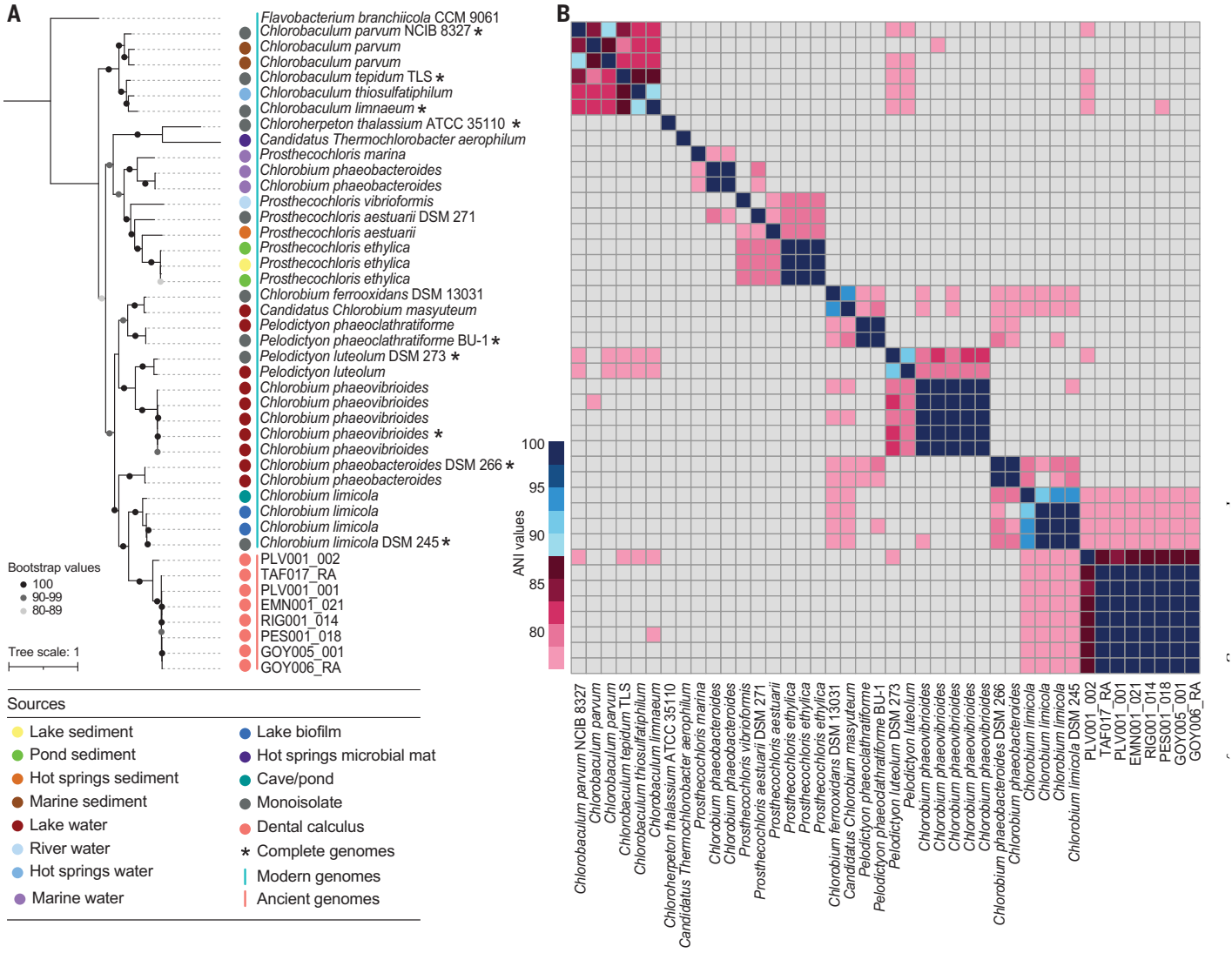


Fig. 2. Phylogeny and shared average nucleotide identities (ANI) of ancient *Chlorobium* MAGs within a subset of modern *Chlorobiales* genomes and MAGs. (**A**) A pruned phylogenetic tree based on the amino acid alignment of all core genes of *C. limicola*, illustrating the phylogeny of the ancient *Chlorobium* MAGs compared with 34 modern genomes of *Chlorobiales* species. The color coding indicates the sources from which the organisms and modern and ancient MAGs were obtained. Clades supported by at least 80% of the bootstrap replicates are indicated by dots on the respective branch. *Flavobacterium branchiicola* was used as an outgroup to root the tree. (**B**) Corresponding heatmap illustrating the ANIs ≥74% (colored) and <74% (gray).

found at a site in Pešturina, Serbia (*29*, *30*). These MAGs substantially exceed the age of the previously oldest reconstructed MAGs (~2000 to 8000 years old) (*11*, *31*) by more than 90,000 years.

Reconstruction of authentic ancient *Chlorobium* genomes from Paleolithic dental calculus

The genus with the highest number of ancient HQ MAGs was *Chlorobium*, a genus that is not typical of oral microbiota or burial sediments (fig. S6) but was detected in the dental calculus of nine individuals (dataset S4). Known members of *Chlorobium* are photolithoautotrophic, obligate anaerobic green sulfur bacteria that typically perform photosynthesis in anoxic water columns (*32*, *33*). Five of those nine individuals returned a HQ *Chlorobium* MAG (EMN001, PES001, GOY005, PLV001, and RIG001), and their HQ *Chlorobium* MAGs had the highest completeness of all MAGs in the metagenome (Fig. 1D). *Chlorobium* MAG EMN001_021, originating from a ~19,000-year-old individual found at El Mirón (*34*), Spain, had a very high median contig length (N50 = 115 kb) and an exceptionally low number of contigs (n = 21) (dataset S3e). We also assembled two MQ *Chlorobium* MAGs from Upper Paleolithic humans—one from TAF017 and another from PLV001 (dataset S3c)—and we genotyped >50% of the *Chlorobium* genome at a minimum coverage of threefold for the Neanderthal GOY006 (dataset S4c; the genotyping was performed by mapping rather than by assembly). The seven samples for which we could reconstruct at least one *Chlorobium* genome at >50% either by de novo assembly or reference-based alignment all originated from Pleistocene contexts in west Eurasia and north Africa and were excavated from either cave sites (EMN001, PES001, GOY005, GOY006, RIG001, and TAF017) or a site close to a water reservoir (PLV001) (Fig. 1 and dataset S1a). *Chlorobium* was largely absent from individuals aged younger than 10,000 years (n = 45), with the exception of ESA006 and ESA007 (dataset S4b), who were found at a 19th-century coastal site in Morocco (dataset S1a).

Because *Chlorobium* is not a known oral taxon, we investigated whether it could be a recent postmortem contaminant by inspecting the DNA damage patterns of the *Chlorobium*

HQ MAGs in comparison with those of the paired HQ MAGs of the securely oral taxon *Flexilinea* (National Center for Biotechnology Information taxonomy name *Anaerolinaceae* bacterium oral taxon 439) from the same individuals. The DNA sequences making up the *Chlorobium* and *Flexilinea* MAGs exhibit equivalently high frequencies of 5′ C-to-T miscoding lesions characteristic of age-related DNA damage (Fig. 1E and fig. S7) (*17*), confirming that the *Chlorobium* MAGs are of authentic ancient origin.

Paleoecology and evolution of ancient *Chlorobium* species

To determine the ecological origin of the ancient *Chlorobium* species, we constructed a phylogenetic tree using all available genomes of the *Chlorobiales* order from diverse environmental sources (Fig. 2A and fig. S8). The ancient *Chlorobium* MAGs are most closely related to *Chlorobium limicola* and form a previously undescribed monophyletic clade (Fig. 2A and figs. S8 and S9). Their set of functional genes, including those required for the photosynthetic machinery, is similar to those found in modern *C. limicola* strains occurring in freshwater biofilms (figs. S10 and S14, and datasets S6 and S8). Because *Chlorobium* spp. are photoautotrophs (*35*), we speculate that anaerobic colonization of the dental plaque biofilm occurred either during life or shortly after death but before burial through transient exposure to contaminated freshwater sources. Periodic in vivo host-mediated biofilm calcification or *Chlorobium*-mediated, light-dependent dolomite formation (*36*) may have further facilitated the long-term preservation of this microbe within dental calculus. Metagenomic analyses of burial-site sediments from El Mirón as well as non-human DNA obtained from a bone sample of the El Mirón individual (*37*) do not show the presence of *Chlorobium* DNA (fig. S11), supporting ante- or perimortem incorporation (dataset S7 and fig. S12).

Although related to *C. limicola*, the ancient *Chlorobium* MAGs are genomically distinct (ANI <80%) (Fig. 2B and fig. S13). Seven ancient MAGs share >95% ANI, indicating that they are members of a single species, whereas PLV001_002 had lower shared ANI values of 85 to 88%, suggesting that it is a different species (*38*). Thus, the ancient *Chlorobium* MAGs appear to represent two unknown, possibly extinct *Chlorobium* species. The ancient *Chlorobium* sp. represented by the seven closely related MAGs differs from *C. limicola* and the ancient *Chlorobium* sp. represented by PLV001_002 in appearing to have lost its ability to oxidize thiosulfate, and it may have also had altered iron-transport and photosynthesis capabilities (dataset S8, c and d). The additional presence of glycoside hydrolase orthologs in the ancient *Chlorobium* MAGs indicates that the ancient *Chlorobium* spp. were adapted to different growth conditions than were *C. limicola*. The low diversity observed among the seven closely related ancient *Chlorobium* MAGs is not unexpected, even considering their broad geographic distribution and time span. Wild populations of *Chlorobium* spp. are known to have extremely low sequence diversity (*39*) and to undergo genome-wide sweeps and strong, purifying selection (*40*). Low diversity was also observed among the HQ MAGs of the oral taxon *Flexilinea* obtained from the same individuals (fig. S15 and dataset S8e) and has been reported in well-studied ancient pathogens (*41*), such as *Yersinia pestis* (dataset S8f).

Discovery of ancient butyrolactone biosynthetic gene clusters (BGCs) with a highly similar genetic architecture

To gain insight into the biosynthetic potential of ancient microbes, we screened the ancient *Chlorobium* MAGs for BGCs by using antiSMASH (*14*) and identified four different biosynthetic gene cluster (BGC) types, of which putative butyrolactone and terpene BGCs are most prevalent (fig. S16B). Butyrolactones are known signaling molecules originally discovered in *Streptomyces* species (*42*). These compounds, however, have not been previously reported in the *Chlorobiales* order. To ensure that the butyrolactone BGCs are of ancient origin, we confirmed that all respective contigs encoding the butyrolactone BGCs have characteristic age-related DNA damage (table S1) by using PyDamage (*13*). We used an A-factor synthase (AfsA)–like gene, constituting the core gene of butyrolactone biosynthesis (*42*), as a query to understand the evolutionary origin of the *Chlorobium* butyrolactone BGC. Phylogenetic analysis of all ancient and 342 representative AfsA-like synthases from various organisms sharing a similar protein domain structure (a single HotDog domain) (*43*) showed that the AfsA-like synthases within the ancient *Chlorobium* MAGs are most closely related to those of the *Chlorobiales* sister lineage (fig. S18), suggesting vertical gene transfer.

To ascertain the fidelity of the ancient butyrolactone BGC architecture, we compared the gene synteny of the ancient BGCs to butyrolactone BGCs across representative genomes within the *Chlorobiales* order and performed a network analysis based on pairwise sequence similarities across the entire BGC region by using the biosynthetic gene similarity clustering and prospecting engine (BiG-SCAPE) (*44*). We found seven distinct gene cluster families (GCFs) that form two major gene cluster clans (GCCs) (Fig. 3A and dataset S7), of which one GCF comprises all ancient butyrolactone BGCs found in the highly related *Chlorobium* MAGs, as well as one modern BGC. A comparison of the genetic architecture and sequences of the modern and ancient butyrolactone BGC regions showed that all ancient and also several closely related modern butyrolactone BGCs share similar genetic features in the vicinity of the AfsA-like synthase gene (Fig. 3B; figs. S16C, S17, and S19; and table S2), suggesting correct reconstruction of the ancient BGCs.

Reconstruction of ancient BGC delivers previously unobserved metabolites

To access the butyrolactone-type metabolite from the ancient *Chlorobium* MAGs, we reconstructed the corresponding BGC. AfsA-like synthase-based biosynthetic pathways typically involve additional modifying enzymes leading to the production of butyrolactones (*42*), butenolides (*45*), or furanoids (*46*). To validate the BGC sequence derived from metagenomic assembly of highly fragmented aDNA, we performed pairwise sequence comparisons for the three core genes (*afsA*-like synthase, oxidoreductase, and *cqsS*-like sensor kinase/phosphatase gene) between all ancient *Chlorobium* MAGs. All BGC core genes are nearly or fully identical in their nucleotide and amino acid sequences (fig. S20), mirroring the high ANI shared between the respective MAGs (Fig. 2B and fig. S12).

We obtained the ancient coding sequence of the three core genes from HQ *Chlorobium* MAG EMN001_021 (which is identical to that in HQ MAG RIG001_021) (fig. S20) by performing gene synthesis and then cloned different sets of the encoded open reading frames into two different expression systems under the control of strong promoters (tables S3 to S5). We used a derivative of the extrachromosomal broad-host-range expression vector pRANGER-BTB3 to determine the required set of biosynthetic genes in an engineered *Pseudomonas protegens* Pf-5 expression chassis. Ultrahigh-performance liquid chromatography coupled with mass spectrometry (UHPLC-MS) indicated the presence of additional compounds when expressing the key biosynthetic gene encoding an AfsA-like synthase (*plfA*) with the adjacent oxidoreductase gene (*plfB*). These products were absent when expressing each gene individually (fig. S21A). Furthermore, we used chassis-independent recombinase-assisted genome engineering (CRAGE) to stably integrate the ancient biosynthetic genes chromosomally into engineered *P. protegens* Pf-5 (table S6) and additionally into *Photorhabdus khanii* subsp. *khanii* XP03 (*47*) to verify that compound production is host-independent. Because the compounds are only produced in both mutant strains, the small molecules must originate from the heterologous expression of the ancient *plfA* and *plfB* genes (Fig. 3C and fig. S21B). From a 3-liter fermentation of Pf-5 harboring both *plfA* and *plfB*, we isolated

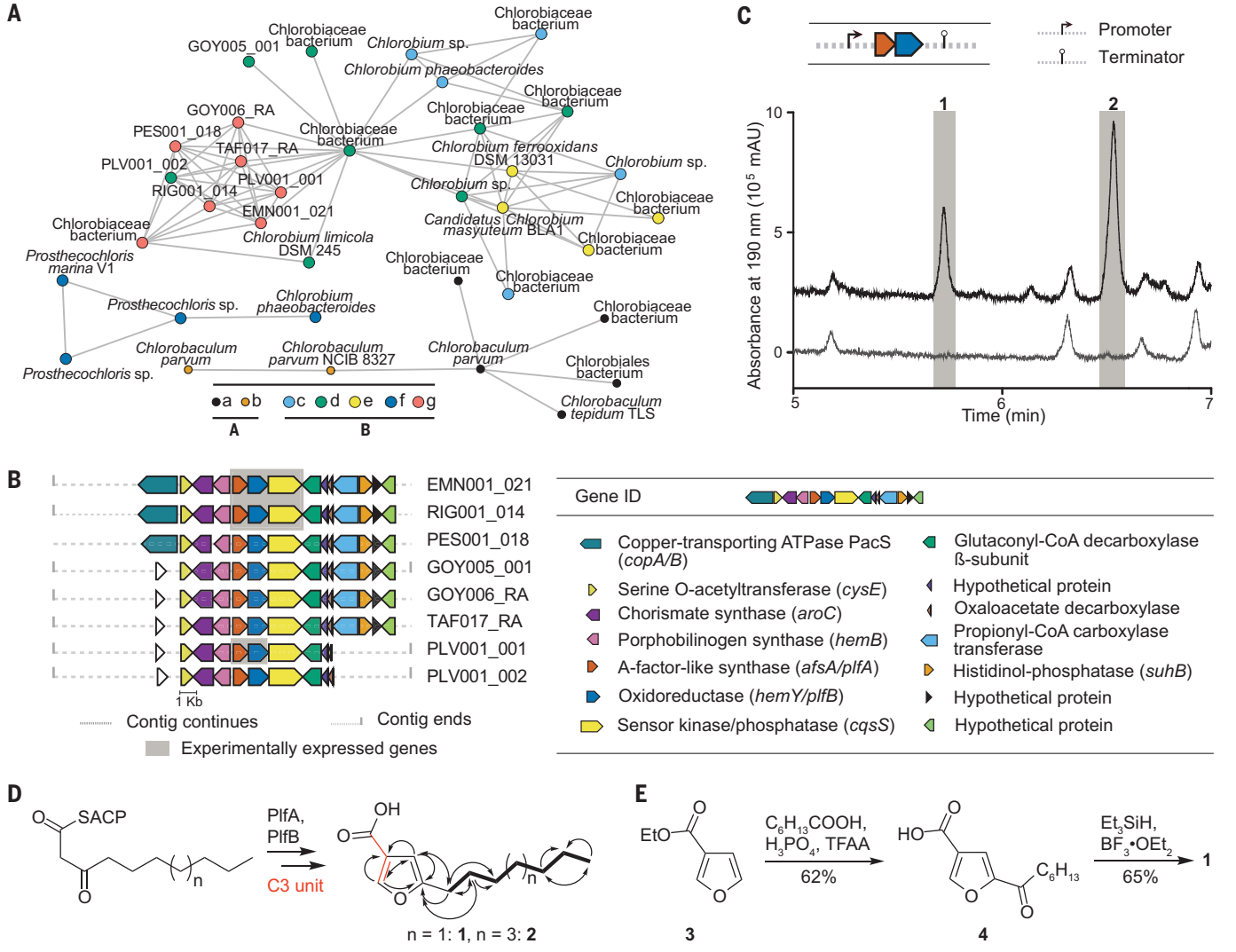


Fig. 3. Reconstruction of ancient butyrolactone BGCs to access metabolites. (**A**) A network analysis of ancient and modern butyrolactone BGCs performed with BiG-SCAPE identified two gene cluster clans (GCCs, A and B) including seven gene cluster families (GCF, a to g) in the order *Chlorobiales*. The butyrolactone BGCs from ancient MAGs fall in one GCF (g). The majority of the modern BGCs are found in different but related GCFs. For illustration purposes, the raw pairwise distances were filtered to include only those estimated as ≤0.3 (supplementary materials, section 7). GOY005_001 does not appear in GCF (g) because of initial assembly truncation (supplementary materials). (**B**) All ancient butyrolactone BGCs share a similar genetic architecture with high sequence similarities. Heterologously expressed genes are highlighted with a gray box. Gene annotations are given with orthologs in parentheses, if available. (**C**) HPLC profiles at 190 nm of culture extracts of Pf-5 harboring the *afsA*-like synthase *plfA* with the adjacent oxidoreductase *plfB* (black line) and of Pf-5 host control (gray line) show that paleofuran A (**1**) and B (**2**) are only produced upon heterologous expression of the ancient biosynthetic genes. mAU, milli-absorbance units. (**D**) Chemical structures of paleofuran A (**1**, $n = 1$) and B (**2**, $n = 3$) illustrated with correlation spectroscopy (COSY) (bold lines) and heteronuclear multiple-bond correlation (HMBC) (arrows) signals. Paleofuran biosynthesis requires both the AfsA-like synthase PlfA and the oxidoreductase PlfB, which most likely condense a glycerol-derived C3 unit (red) with an acyl carrier protein (ACP)–bound 3-oxo fatty acid, followed by a decarboxylation. (**E**) Scheme for the chemical synthesis of paleofuran A (**1**). TFAA, trifluoroacetic anhydride.

compounds with pseudomolecular masses $[M-H]^-$ of mass/charge ratio (m/z) = 209.1178 and 237.1493, corresponding to the molecular formulae $C_{12}H_{18}O_3$ and $C_{14}H_{22}O_3$, respectively. Using nuclear magnetic resonance (NMR) spectroscopy (figs. S26 to S32) and gas chromatography coupled with electron impact–mass spectrometry (GC-EI-MS) fragment analyses (fig. S23), we elucidated the structure of two previously unknown 5-alkylfuran-3-carboxylic acids with C7 or C9 alkyl side chains, which we named paleofuran A (**1**) and paleofuran B (**2**), respectively (Fig. 3D, fig. S22, and table S7). The chemical total synthesis of paleofuran A (**1**) in only two steps with 40% overall yield starting from 3-carboethoxyfuran (**3**) through ketone (**4**) confirmed the proposed structure (Fig. 3E, fig. S24A, and figs. S33 to S35).

As previously shown for structurally related methylenomycin furan (MMF) autoinducers (*48*), AfsA-dependent biosynthetic pathways can also lead to furanoid formation (*46*). Analogously, we found that feeding of glycerol-1,1,2,3,3-d5 led to incorporation of a single deuterium in the furan core next to the ring oxygen (figs. S24B, S25, S36, and S37). This result suggests that the building blocks are also derived from a glycerol C3 unit and a C10 (for **1**) or C12 (for **2**) 3-oxo fatty acid, which would decarboxylate after condensation (Fig. 3D). Because MMFs regulate the adjacent methylenomycin BGC in *Streptomyces* species (*46*), we analyzed the vicinity of the ancient

Chlorobium AfsA-like synthases. The adjacent genes, which are conserved in modern *Chlorobium* genomes, encode biosynthetic enzymes required for production of bacteriochlorophyll precursors, suggesting that the paleofurans could be involved in regulating the bacterial photosynthesis (fig. S19, table S2, and dataset S6).

Discussion

The paleobiotechnological strategy we have described unlocks the biosynthetic potential of authentic ancient metagenomes, effectively adding a time dimension to natural product discovery. Natural product recovery not only expands our understanding of the paleome but also raises questions regarding the management, stewardship, and intellectual property of humanity's collective biological and cultural heritage (supplementary materials, section 1.1). The multiple steps implemented throughout the workflow address the known challenges of aDNA recovery from diverse sources (*17*, *49*) and indicate the benefit of deeply sequenced ancient samples for the exploration of short read–based ancient metagenomes.

The recovery of unknown, possibly extinct ancient microbial species further opens an avenue to investigate lost microbial diversity due to environmental changes and increased selective pressures—for example, climate change or (mis)use of antibiotics, respectively (*50*, *51*). However, given the insufficient representation of *Chlorobium* genomes in current databases, we cannot be certain whether the two ancient ecotypes are extinct or only unrecovered. *Chlorobium* members are key players in the carbon, nitrogen, and sulfur cycles (*52*) and can harvest light at very low intensities (*32*), making it likely that they were incorporated into ancient dental biofilms through transient exposure to contaminated water. Pleistocene-era cave dwelling may have increased human exposure to *Chlorobium* spp., and the high prevalence of the genus only among Paleolithic individuals suggests that the environmental microbial exposures of humans may have substantially changed over time as habitation patterns shifted in response to Holocene-era climate change.

From the reconstructed ancient *Chlorobium* MAGs, we accessed previously unreported paleofurans. That the paleofuran BGC and its genetic vicinity remained largely unchanged over at least ~100,000 years may suggest that paleofurans play an important role in the ecology of *Chlorobium* species. This ability to reconstruct microbial genomes from well-dated archaeological contexts has important implications for refining bacterial evolution models and overcoming time-resolution challenges. Our approach can, in principle, be applied to any ancient metagenome, thus providing a roadmap for future BGC exploration and metabolite discovery. The analysis and comparison of the product spectrum of ancient and modern paleofuran BGC homologs offer key opportunities to explore BGC evolution.

By merging metagenomics, genome mining, gene synthesis, and metabolic analyses with the field of aDNA research, we chart a path for the discovery of ancient natural products to gain evolutionary insights on their formation and origin, as well as to inform their potential future applications.

REFERENCES AND NOTES

1. P. Stallforth, M. Mittag, A. A. Brakhage, C. Hertweck, U. A. Hellmich, *Trends Biochem. Sci.* **48**, 71–81 (2023).
2. S. A. West, A. S. Griffin, A. Gardner, S. P. Diggle, *Nat. Rev. Microbiol.* **4**, 597–607 (2006).
3. A. G. Atanasov, S. B. Zotchev, V. M. Dirsch, C. T. Supuran, International Natural Product Sciences Taskforce, *Nat. Rev. Drug Discov.* **20**, 200–216 (2021).
4. D. J. Newman, G. M. Cragg, *J. Nat. Prod.* **83**, 770–803 (2020).
5. I. M. Velsko *et al.*, *Metabolomics* **13**, 134 (2017).
6. K. I. Bos *et al.*, *Nature* **478**, 506–510 (2011).
7. J. A. Fellows Yates *et al.*, *Sci. Data* **8**, 31 (2021).
8. C. Warinner *et al.*, *Annu. Rev. Genomics Hum. Genet.* **18**, 321–356 (2017).
9. J. A. Fellows Yates *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2021655118 (2021).
10. A. Almeida *et al.*, *Nat. Biotechnol.* **39**, 105–114 (2021).
11. M. C. Wibowo *et al.*, *Nature* **594**, 234–239 (2021).
12. J. C. Brealey *et al.*, *Mol. Biol. Evol.* **37**, 3003–3022 (2020).
13. M. Borry, A. Hübner, A. B. Rohrlach, C. Warinner, *PeerJ* **9**, e11845 (2021).
14. K. Blin *et al.*, *Nucleic Acids Res.* **49**, W29–W35 (2021).
15. L. Paoli *et al.*, *Nature* **607**, 111–118 (2022).
16. Y. Sugimoto *et al.*, *Science* **366**, eaax9176 (2019).
17. L. Orlando *et al.*, *Nat. Rev. Methods Primers* **1**, 14 (2021).
18. D. Li, C.-M. Liu, R. Luo, K. Sadakane, T.-W. Lam, *Bioinformatics* **31**, 1674–1676 (2015).
19. S. Nurk, D. Meleshko, A. Korobeynikov, P. A. Pevzner, *Genome Res.* **27**, 824–834 (2017).
20. E. Garrison, G. Marth, Haplotype-based variant detection from short-read sequencingarXiv:1207.3907v2 [q-bio.GN] (2012)
21. G. V. Uritskiy, J. DiRuggiero, J. Taylor, *Microbiome* **6**, 158 (2018).
22. D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz, G. W. Tyson, *Genome Res.* **25**, 1043–1055 (2015).
23. A. Orakov *et al.*, *Genome Biol.* **22**, 178 (2021).
24. R. M. Bowers *et al.*, *Nat. Biotechnol.* **35**, 725–731 (2017).
25. F. Meyer *et al.*, *Nat. Methods* **19**, 429–440 (2022).
26. S. Saheb Kashaf *et al.*, *Nat. Microbiol.* **7**, 169–179 (2022).
27. T. Chen *et al.*, *Database* **2010**, baq013 (2010).
28. F. Asnicar *et al.*, *Nat. Commun.* **11**, 2500 (2020).
29. P. Radović, J. Lindal, D. Mihailović, M. Roksandic, *J. Hum. Evol.* **131**, 139–151 (2019).
30. B. Blackwell *et al.*, in *Palaeolithic and Mesolithic Research in the Central Balkans*, D. Mihailović, Ed. (Serbian Archaeological Society, 2014), 21–38.
31. L. Granehäll *et al.*, *Microbiome* **9**, 197 (2021).
32. J. M. Tsuji *et al.*, *ISME J.* **14**, 2732–2747 (2020).
33. N. Lambrecht *et al.*, *Front. Microbiol.* **12**, 695260 (2021).
34. L. G. Straus, M. R. González Morales, D. Cuenca-Solana, *J. Archaeol. Sci.* **60**, 1–9 (2015).
35. J. T. Beatty *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 9306–9310 (2005).
36. M. Daye, J. Higgins, T. Bosak, *Geology* **47**, 509–512 (2019).
37. Q. Fu *et al.*, *Nature* **534**, 200–205 (2016).
38. C. Jain, L. M. Rodriguez-R, A. M. Phillippy, K. T. Konstantinidis, S. Aluru, *Nat. Commun.* **9**, 5114 (2018).
39. P. Panwar *et al.*, *Microbiome* **9**, 231 (2021).
40. M. L. Bendall *et al.*, *ISME J.* **10**, 1589–1601 (2016).
41. A. Andrades Valtueña *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2116722119 (2022).
42. J. Y. Kato, N. Funa, H. Watanabe, Y. Ohnishi, S. Horinouchi, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 2378–2383 (2007).
43. S. C. Dillon, A. Bateman, *BMC Bioinformatics* **5**, 109 (2004).
44. J. C. Navarro-Muñoz *et al.*, *Nat. Chem. Biol.* **16**, 60–68 (2020).
45. M. Klapper *et al.*, *Angew. Chem. Int. Ed.* **59**, 5607–5610 (2020).
46. C. Corre, L. Song, S. O'Rourke, K. F. Chater, G. L. Challis, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17510–17515 (2008).
47. G. Wang *et al.*, *Nat. Microbiol.* **4**, 2498–2510 (2019).
48. S. Zhou, N. R. Malet, L. Song, C. Corre, G. L. Challis, *Chem. Commun.* **56**, 14443–14446 (2020).
49. C. Warinner *et al.*, *Nat. Genet.* **46**, 336–344 (2014).
50. D. A. Hutchins *et al.*, *Nat. Rev. Microbiol.* **17**, 391–396 (2019).
51. J. L. Sonnenburg, E. D. Sonnenburg, *Science* **366**, eaaw9255 (2019).
52. A. M. Linz *et al.*, *PeerJ* **6**, e6075 (2018).
53. A. Hübner, Paleome reconstruction reveals Middle and Late Upper Paleolithic bacterial species and adds time axis in natural product discovery, version 1, Edmond (2023); https://doi.org/10.17617/3.COCL7R.
54. M. Klapper, A. Hübner, A. Ibrahim, Natural products from reconstructed bacterial genomes of the Middle and Upper Paleolithic, version 1.1, Zenodo (2023): https://doi.org/10.5281/zenodo.7777932.

ACKNOWLEDGMENTS

We thank M. Hajdinjak for the access to the shotgun sequencing data of the toe bone of EMN001 from El Mirón; S. Sázelová, J. Svoboda, I. Crevecoeur, H. Rougier, P. Semal, D. Mihailović, M. Roksandic, C. Bauer, C. Cupillard, M. Díaz-Zorita Bonilla, D. Drucker, and L. Humphrey for providing Pleistocene-era dental calculus for analysis; F. Aron for paleogenomic data generation; Z. Zhao and Y. Yoshikuni for sharing the CRAGE system with us; A. Perner, N. Ueberschaar, and H. Heinecke for MS and NMR measurements; M. Cyrulies, M. Steinacker, and U. Knüpfer for large-scale fermentations; and F. Dewhirst for discussions about the ecology and metabolism of *Chlorobium*. **Funding:** This research was supported by the Werner Siemens Foundation; the Max Planck Society; the Leibniz Association; the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2051 (project 390713860, "Balance of the Microverse"); CRC1127 ChemBioSys, B09, project 239748522; and by the Leibniz Award. The DAAD graduate fellowship program supported W.K.A., and the Joachim Herz Foundation supported M.K. **Author contributions:** C.W. and P.S. conceived and designed the project. C.W. and P.S. acquired funding and supervised the research. M.K. and A.H. administered and coordinated the research. M.K. established cloning and heterologous expression protocols. A.H. performed assembly, authentication, binning, MAG classification, genome reconstruction, and pangenome and phylogenetic analyses. A.I. performed phylogenetic analyses, BGC bioinformatics analyses, and genomic DNA and plasmid sequencing analyses. M.K., S.C., J.F., J.A.F.Y., I.M.V., and M.B. contributed to bioinformatics analyses. M.R.G.M. and L.G.S. directed the excavations at El Mirón Cave and provided archaeological sediment samples. M.K., I.W., H.S., E.V.B., T.H., and P.S. performed cloning and heterologous expression and prepared host chassis strains. M.K. and S.Z. performed isolation, purification, and structure elucidation. M.K., I.W., and V.H. performed biosynthetic studies. V.H. performed GC–MS analyses. W.K.A. performed syntheses. M.K., A.H., A.I., V.H., W.K.A., M.B., S.Z., I.W., H.S., J.F., I.M.V., J.A.F.Y., R.H., H.M.D., I.V., C.H., C.W., and P.S. analyzed data. M.K., A.H., A.I., C.W., and P.S. wrote the manuscript, with the input of all coauthors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The accession number for all newly reported sequencing data is available from the European Nucleotide Archive: PRJEB55583. The sequences and annotations for the MAGs are available in the Edmond Max Planck Data Repository (*53*). The code and files used to reproduce the analyses are available in the GitHub repository https://github.com/paleobiotechnology/EMN001_Paleofuran, archived in Zenodo (*54*), which also includes the raw NMR data, GC–MS data, HPLC chromatogram data, and a material transfer agreement for CRAGE components granted on 20 October 2020. Datasets S1 to S9 can be found under 06-figures_tables in the code repository. **License information:** Copyright © 2023 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.adf5300
Materials and Methods
Supplementary Text
Figs. S1 to S37
Tables S1 to S7
References (*55*–*164*)
MDAR Reproducibility Checklist

Submitted 28 October 2022; accepted 13 April 2023
Published online 4 May 2023
10.1126/science.adf5300

AGING

Effects of hunger on neuronal histone modifications slow aging in *Drosophila*

K. J. Weaver[1]*, R. A. Holt[2], E. Henry[3], Y. Lyu[4], S. D. Pletcher[1]*

Hunger is an ancient drive, yet the molecular nature of pressures of this sort and how they modulate physiology are unknown. We find that hunger modulates aging in *Drosophila*. Limitation of branched-chain amino acids (BCAAs) or activation of hunger-promoting neurons induced a hunger state that extended life span despite increased feeding. Alteration of the neuronal histone acetylome was associated with BCAA limitation, and preventing these alterations abrogated the effect of BCAA limitation to increase feeding and extend life span. Hunger acutely increased feeding through usage of the histone variant H3.3, whereas prolonged hunger seemed to decrease a hunger set point, resulting in beneficial consequences for aging. Demonstration of the sufficiency of hunger to extend life span reveals that motivational states alone can be deterministic drivers of aging.

The relations between an animal and its diet influence behavior and aging in remarkable ways. The physiological need for nutrients motivates animals to forage and feed, and forced limitation in food availability slows aging across taxa (*1*). Both effects derive not only from the energetic content of the diet but also its composition (*2*–*4*). Many animals eat until they have consumed a specific amount of protein, for example, and the protein:carbohydrate ratio that results in part from seeking this target is a major factor in modulating life span (*5*). An animal need not even consume its diet to be affected by it. In mice, appetite is promoted by environmental cues that are associated with future food consumption, presumably by influencing broader neural states that specify nutrient-specific drive, or hunger (*6*, *7*). Similarly, the taste and smell of specific nutrients modulate life span in the fruit fly, *Drosophila melanogaster*, and the nematode, *Caenorhabditis elegans,* among other species (*8*–*11*).

The effects of diet on aging and behavior could share mechanistic foundations in the motivational states they promote, yet little is known about the molecular nature of these states. Many animals, including humans, develop a motivational drive for protein, as do *Drosophila* in response to starvation and mating (*12*–*14*). The existence of this drive and its ability to influence physiology is demonstrated by the observation that perception of protein-containing food without its consumption reverses the beneficial effects of protein-restricted diets in *Drosophila* and *C. elegans* (*11*, *15*, *16*). The neurotransmitter serotonin, together with neurons that express a specific serotonin receptor (5-HT2A), modulate this hunger state, and loss of 5-HT2A increases life span up to 50% in nutrient-rich conditions (*17*). Observations such as these led us to test whether the neural states that encode the motivation to seek food are sufficient to slow aging, independent of any changes in nutrient intake that may result.

We established that hunger modulates aging in *Drosophila* and demonstrated that this motivational state promotes feeding and modulates life span through partially separable downstream mechanisms. Flies that were made hungry by reducing dietary branched-chain amino acids (BCAAs) or by activation of neurons that evoke hunger were long-lived, despite consuming more calories and more total protein. Specific subsets of *Drosophila* neurons used BCAA metabolism to modify the neuronal histone acetylome—a plasticity that was required to encode hunger in these models. Prolonged hunger appeared to alter a set-point around which appetite is regulated and this adaptation may be an important component of a slowed aging process.

Results

To investigate how neural states motivate feeding and modulate aging, we focused our investigations on the BCAAs primarily because decreasing their abundance in the diets of mammals and flies increases protein appetite and extends life span (*18*, *19*). We used a chemically defined diet to adjust the concentration of BCAAs without altering other dietary components (*20*). We designed a reference holidic diet (RD), within which we manipulated BCAAs over a range of concentrations consistent with those in diets used for *Drosophila* aging studies that avoid malnutrition and overnutrition. We kept equal concentrations of all other non-essential amino acids and non-BCAA essential amino acids across diets, which allo[illegible] us to avoid confounding effects of gen[illegible] amino acid deficiency (Fig. 1A) (*21*). This led to diets of slightly different caloricity. However, caloric content is less impactful than dietary composition in modulating life span, and subsequent experiments ruled out differences in calories as a cause of the dietary effects we observed (*2*).

Dietary BCAA concentration modulated fly life span (*19*, *22*) (Fig. 1B). Life-span extension was greater in female flies and was maximized on a diet containing roughly one-third of the BCAA content of the reference diet, which is consistent with similar effects observed with dilution of the standard laboratory diet (Fig. 1, B and C, and fig. S1A) (*2*, *3*). We focused our investigation on female flies because of their heightened life span and neuronal responses to protein availability (*23*, *24*). We chose two diets, hereafter termed "low BCAA" (8 mM BCAAs/5.6% w/v total amino acid) and "high BCAA" (44 mM BCAAs/7.2% w/v total amino acid), to investigate in detail.

Variation in the dietary BCAAs also modulated food intake. We used a method that determines how much volume a group of flies consume in 24 hours (termed "Con-Ex" for Consumption-Excretion) (*25*, *26*). Flies on low-BCAA diets consumed more food volume than those on high-BCAA diets (Fig. 1D, left panel). This appeared to be due to an increase in food intake by flies on low-BCAA diets rather than a decrease by flies on high-BCAA diets because food intake was also increased compared with that of flies fed the RD (fig. S1B). Differences were unlikely to be a result of food taste; flies carrying a mutation in the transcription factor Pox-Neuro (*PoxN*), who fail to develop chemosensory bristles and thus have extreme deficits in chemosensation (*27*, *28*), also consumed more food on low-BCAA diets (Fig. 1D, right panel). Increased volume intake by flies on low-BCAA diets resulted in significantly higher caloric intake (Fig. 1E) but in similar amounts of total amino acids (micrograms per fly) eaten between flies fed low- or high-BCAA diets (Fig. 1F, left panel), whereas BCAA intake was significantly reduced (micrograms per fly) for flies fed the low-BCAA diet (Fig. 1F, right panel). Thus, flies fed our low-BCAA diet consumed more calories and similar amounts of total amino acids yet lived significantly longer.

Dietary BCAAs influence hunger states

The relative increase in food intake we observed for flies fed low-BCAA diets led us to speculate that reducing dietary BCAAs promoted a heightened and persistent state of hunger. Measuring hunger can be challenging in a simple model system; animals eat for many reasons, and total food intake on homogenous mixtures of nutrients may not be as indicative of hunger levels as feeding paradigms

[1]Department of Molecular and Integrative Physiology and Geriatrics Center, Biomedical Sciences and Research Building, University of Michigan, Ann Arbor, MI 48109, USA. [2]College of Literature, Science, and the Arts, Biomedical Sciences and Research Building, University of Michigan, Ann Arbor, MI 48109, USA. [3]Program in Cellular and Molecular Biology, University of Michigan, Ann Arbor, MI 48109, USA. [4]Department of Molecular Biology & Biochemistry, Rutgers University, Piscataway, NJ 08855, USA.
*Corresponding author. Email: spletch@umich.edu (S.D.P.); kjweaver@umich.edu (K.J.W.)

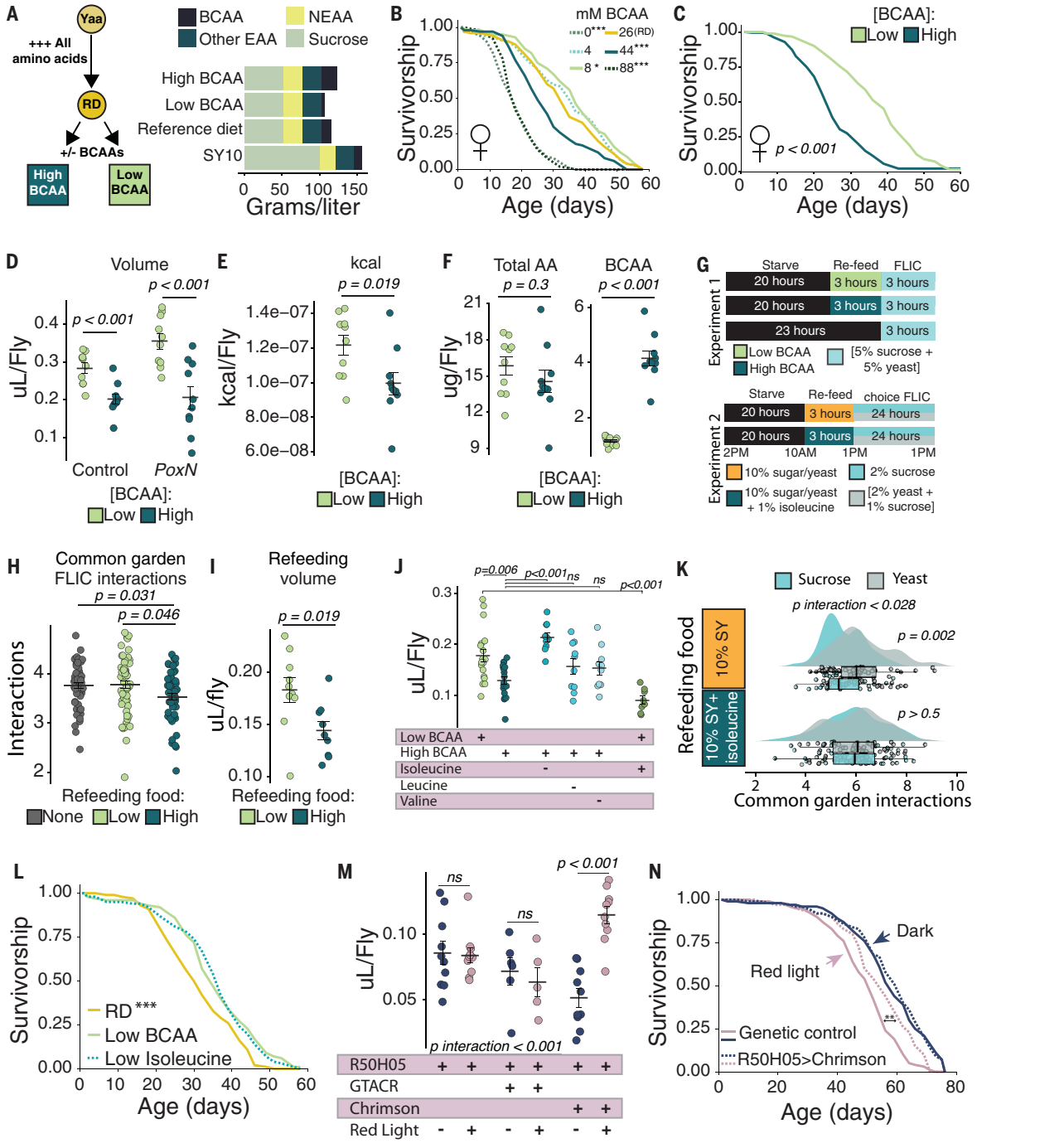


Fig. 1. Hunger extends life span independent of appetite in dietary and genetic models of hungry flies. (**A**) Schematic (left) and relative composition (right) of experimental diets. Details can be found in tables S1 and S2 and methods. Yaa, baseline diet as in (*20*); RD, reference diet, containing 51.36g/L carbohydrate, 34.9g/L nonessential amino acids, 22.56g/L non-BCAA essential amino acids, and 12.39g/L BCAAs. (**B**) Life span of female *Canton-S* flies on diets of indicated BCAA concentration (log-rank test, *P*-values derived by comparison to RD, n = 159 to 172). (**C**) Life span of *Canton-S* flies on low- or high-BCAA diets (log-rank test, n = 173 to 175). (**D** to **F**) 24-hour Con-Ex measurement of w^{1118} or *PoxN* flies on indicated diets. Volume (D), kcal (E), and total or BCAA-only amino acid intake (F) (two-tailed Student's *t*-tests, n = 10 flies per treatment). (**G**) Schematic of FLIC experimental designs used in (H and I) (top panel) and (K) (bottom panel). (**H**) FLIC interactions in common garden [one-way analysis of variance (ANOVA) with Tukey's post hoc, two experimental replicates are pooled, n = 52 to 55 flies per treatment]. (**I**) Con-Ex measurement of volumetric intake during the refeeding period (two-tailed Student's *t*-test). (**J**) Con-Ex measurement on diets with individual BCAA reductions (one-way ANOVA with Tukey's post hoc, n = 8 to 18 flies per treatment). (**K**) FLIC interactions in food choice environment [2% sucrose or (2% yeast + 1% sucrose)] after refeeding SY food ± 1% isoleucine (two-way ANOVA with Tukey's post hoc, 3 experimental replicates are pooled, n = 59 to 60 flies per treatment). (**L**) Life span of *Canton-S* flies on RD, low-BCAA, or low-isoleucine diets (log-rank test, *P*-values derived by comparison to low-BCAA diet, n = 171 to 177). (**M** to **N**) Con-Ex and life-span measurements of flies carrying *R50H05-GAL>UAS-CsChrimson* or *R50H05-GAL4/w-; CS* controls exposed to red light for 12 hours per day or kept in constant darkness. (M) 24-hour Con-Ex measurement (two-way ANOVA with Tukey's post hoc, n = 5 to 10 flies per treatment). (N) Life-span measurement (log-rank test, n = 101 to 123). All FLIC data are expressed as box-cox transformation to the 0.25 power to achieve normality. **P* < 0.05, ***P* < 0.01, ****P* < 0.001. Error bars represent the mean ± SEM.

that quantify precisely when and how often individual animals interact with food sources (*6*, *13*, *17*, *29*, *30*). We therefore devised a refeeding assay to determine how BCAAs influence hunger. In this refeeding paradigm, flies deprived of food for 20 hours were refed a 3-hour bolus of test food, such as low- or high-BCAA, and then placed into a common sucrose and yeast food environment (termed "common garden") where we measured food interactions using the Fly Liquid Food Interaction Counter (FLIC) (*31*) (Fig. 1G, Experiment 1, top panel). We reasoned that if BCAAs did influence a state of hunger, then manipulating their concentration in the test food during refeeding would alter food interactions in the common garden.

Flies refed a low-BCAA food subsequently interacted more often with the sucrose and yeast food in the common garden than did flies refed a high-BCAA food and, in fact, interacted as frequently as did flies that had not been refed at all (Fig. 1H). This was not due to a reduction in total calories or amino acids consumed during the refeeding period: flies ingested significantly more food volume and also more calories, as measured by Con-Ex, when refeeding on a diet of low-BCAAs than they did when refeeding on a high-BCAA diet (Fig. 1I), although total amino acid intake (micrograms per fly) was indistinguishable from that of flies refed high BCAAs, and BCAA consumption (micrograms per fly) was significantly lower (fig. S1C).

Although the three BCAAs are commonly investigated together because they share similarities in their biochemical structures and functions, we tested them individually for their role in promoting feeding. We measured food intake after reduction of each BCAA from the high-BCAA diet and found that decreasing isoleucine, but not valine or leucine or other pairwise combinations of amino acids, was required to increase feeding (Fig. 1J and fig. S1, D and E). Reducing dietary isoleucine increased the total calories consumed (fig. S1E) and was the only dietary manipulation that we observed as sufficient to promote increased total amino acid intake (fig. S1D). Furthermore, increasing isoleucine alone from our low-BCAA diet to match the concentration in our high-BCAA diet was sufficient to decrease food intake (Fig. 1J), demonstrating that isoleucine acts as a dietary signal to modulate feeding.

Having established that isoleucine was sufficient to modulate volumetric food intake, we used our refeeding paradigm to test whether hunger for specific nutrients was affected by this amino acid. Choices between carbohydrate- or protein-rich food are thought to be tuned by increased protein demand in starved animals, who exhibit increased interactions with protein-rich food compared to carbohydrate-rich food, and to serve as a useful proxy for measuring protein-specific hunger (*31*–*33*). We reasoned that if isoleucine has a specific influence on protein hunger, then manipulating its concentration in the test food during refeeding might alter future choices between protein or carbohydrate food during the subsequent assessment period. We allowed flies to refeed for three hours on either conventional 10% sugar and yeast food or the same food to which we added 1% isoleucine. We then measured over the next 24 hours the frequency of individual fly interactions with either carbohydrate- or protein-rich food in a choice environment (Fig. 1G, Experiment 2, bottom panel). Control flies that were refed conventional sugar and yeast food exhibited more interactions with protein-rich food compared with the carbohydrate-rich food, establishing that the three hours of refeeding before the test period were insufficient to fully satiate the animals and that they maintained a heightened protein drive (Fig. 1K, top). On the other hand, flies that were refed food with added isoleucine had statistically indistinguishable interactions with carbohydrate-rich compared to protein-rich food (Fig. 1K, bottom). Thus, amounts of dietary isoleucine modulate protein-specific appetite, suggesting that our low-BCAA diet promoted a heightened state of hunger, perhaps by influencing protein drive.

Hunger states influence life span

We then tested whether restriction of isoleucine alone could increase life span. We reduced all BCAAs or just isoleucine from our reference diet and found that limitation of isoleucine (but not valine or threonine) was sufficient to extend life span to a similar degree as reducing all BCAAs, despite increased intake of calories, total amino acids, and carbohydrates (Fig. 1L and fig. S1F). Thus, the motivational state of hunger itself, rather than the availability or energetic characteristics of the diet, might slow aging. Further support for this idea was provided by the observation that our diets had no significant effects on egg laying, activity levels, triglyceride and protein levels, activation of mTOR/AKT pathways, or extension of life span upon dietary restriction (fig. S2, A to F), which are commonly associated with manipulations that modulate life span through changes in nutrient availability or toxicity.

To determine whether hunger itself might slow aging, we sought a way to induce it independent of dietary manipulations and examine its effect on life span. Several populations of neurons have been found to stimulate hunger-driven behaviors in the fly (*34*). Of these, the activity of one population seems to encode the motivation to feed, rather than behaviors associated with feeding (*35*). These neurons can be manipulated using the Gal4 transgenic driver line, R50H05-GAL4, which labels 80 neurons in the central brain. Using the Gal4/UAS system, we expressed the red light–sensitive cation channel UAS-CsChrimson to create flies in which R50H05 hunger neurons were optogenetically activated at will to artificially generate hunger. These flies consumed twice as much food as did flies of the same genotype that were kept in the dark or that expressed an activity-inhibiting green light–sensitive opsin channel (GtACR), consistent with prior reports (Fig. 1M). Inducing a heightened state of hunger by activating these neurons for 12 hours each day extended life span relative to that of a genetic control strain. This was not observed in flies maintained in constant darkness, where flies are slightly longer-lived in general, a known phenomenon that has been previously reported (*36*) (Fig. 1N, fig. S3). Thus, hunger itself may slow aging.

Molecular encoding of hunger by neuronal histone modifications

We explored the molecular interactions between BCAAs, hunger neurons, and the brain to determine how hunger might be molecularly encoded. Alterations in histone proteins are involved in the generation of motivated behaviors, are a hallmark of aging, and are regulated by nutrients (*37*–*40*). Furthermore, histone acetylation is especially sensitive to the availability of dietary nutrients that generate acetyl-CoA, and metabolism of BCAAs produces this substrate (*41*, *42*).

We examined whether dietary BCAAs influenced histone posttranslational modifications (PTMs). Histone H3 abundance and histone H3K9 acetylation was significantly decreased in fly heads within one week of feeding a low-BCAA diet (Fig. 2A). H3K27 acetylation, but not H3K9 methylation, was also slightly reduced when flies were fed low-BCAA diets. Effects of dietary BCAAs on histone PTMs and total histone H3 were abrogated by feeding flies the histone deacetylase inhibitor, sodium butyrate (Fig. 2C). Histone H3 abundance decreased in heads and brains from flies on low-BCAA diets, although H3 mRNA was increased (Fig. 2, A to C, and fig. S4). Histone H4 did not change between diets (Fig. 2D). Inducing hunger independent of diet by activation of R50H05 hunger neurons also reduced histone H3 abundance in fly heads (Fig. 2E) and adding additional isoleucine to a low-BCAA diet partially recapitulated the effects of BCAAs on histone H3 abundance, suggesting that histone modifications may be correlated with one or both hunger phenotypes (Fig. 2D).

Because abundance of H3 mRNA increased but protein abundance decreased in animals on low-BCAA diets, we tested whether BCAAs might influence turnover of histone H3. In the brain, histone H3 is marked for eviction from chromatin by acetylation marks and can be replaced by the histone variant H3.3 (*43*, *44*). This variant is thought to accumulate in regions of actively transcribed chromatin (*45*).

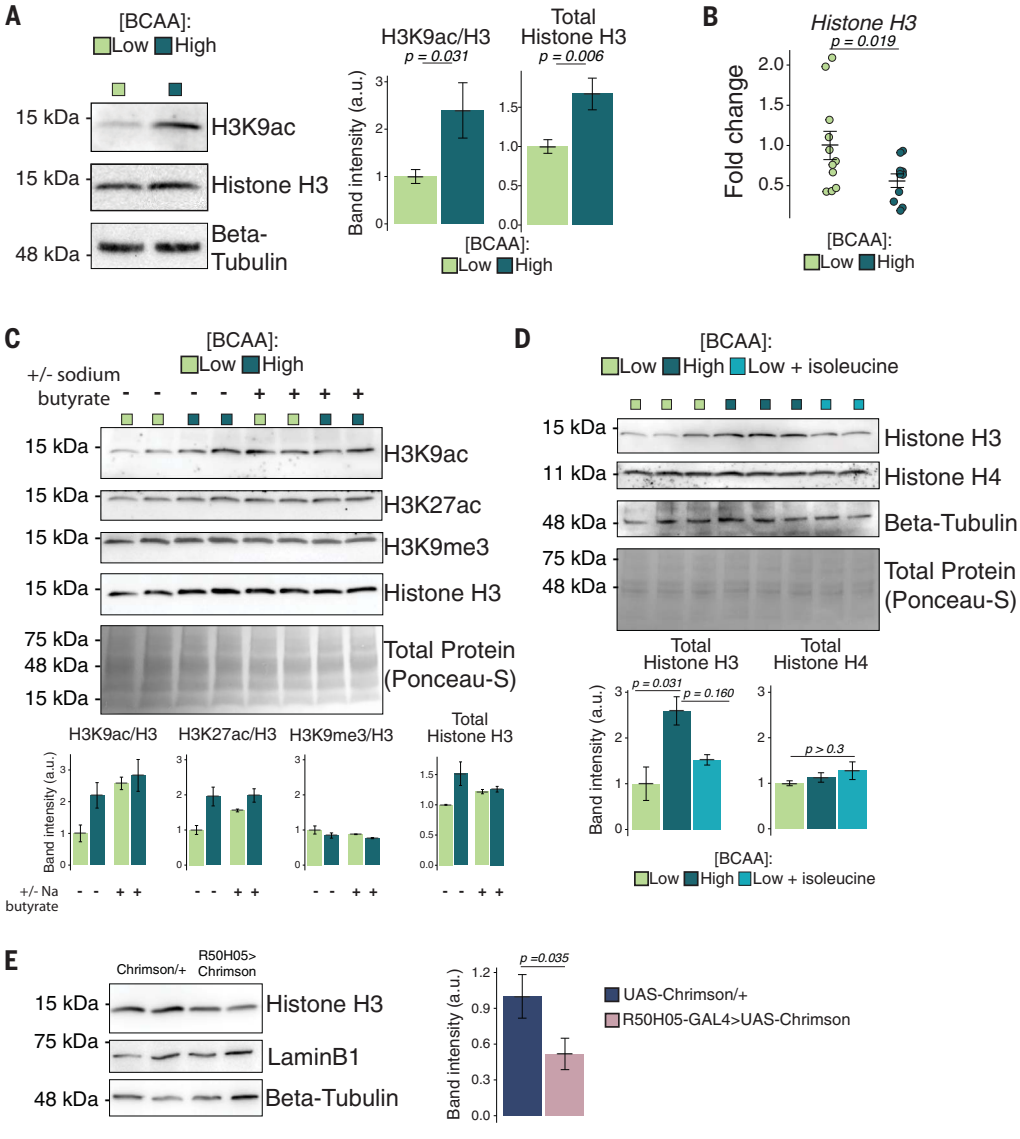


Fig. 2. Dietary BCAAs alter histone acetylation and histone H3 abundance. (**A**) Representative western blots for H3K9ac or total histone H3 in *Canton-S* fly heads after exposure to indicated diets for 5 to 7 days, quantified in right panel (one-way ANOVA; 10 heads were used per biological replicate, and 5 experimental replicates were performed, n = 11 biological replicates per treatment). Values are normalized to low-BCAA treatment. (**B**) RT-qPCR measurement of relative mRNA abundance in *Canton-S* heads after 5 to 7 days on low- or high-BCAA, values are normalized to low-BCAA treatment (one-way ANOVA; 10 heads were used per biological replicate and 2 experimental replicates were performed, n = 10 to 11 biological replicates per treatment). (**C**) Western blots for histone PTMs in *Canton-S* heads after exposure to indicated diets ± 100 mM sodium butyrate for 5 to 7 days, bands are quantified in bottom panel and normalized to low-BCAA treatment (10 heads were used per biological replicate, n = 2 biological replicates). (**D**) Western blots for histone H3 and H4 in *Canton-S* heads on low-BCAA, high-BCAA, or low-BCAA + high isoleucine diets, quantified in bottom panel and normalized to low-BCAA treatment (one-way ANOVA with Tukey's post hoc. 10 heads were used per biological replicate, n = 2 to 3 biological replicates). (**E**) Western blots for histone H3 in heads of flies carrying *R50H05-GAL>UAS-CsChrimson* or *UAS-CsChrimson/w-;CS* controls exposed to red light for 12 hours, quantified in right panel and normalized to *UAS-CsChrimson/w-;CS* control (one-way ANOVA; 10 heads were used per biological replicate, n = 4 to 5 biological replicates). Error bars represent the mean ± SEM.

H3.3 was transcriptionally increased in the heads of flies fed low-BCAA diets (Fig. 3A). To visualize the incorporation of histone H3.3 into neuronal chromatin, we used the inducible pan-neuronal driver Nsyb-GeneSwitch-GAL4, which allowed for the expression of an *H3.3-GFP* transgene to be pulsed and visualized by addition of the drug RU486 to food and chased after removal of RU486 (*43*). We observed higher H3.3-GFP signal in fly brains after one week of feeding on low- versus high-BCAA diets, indicating an increase in its stability or persistence in chromatin (Fig. 3B, lower panel; quantified in Fig. 3D), which seemed to be not attributable to changes in initial H3.3-GFP incorporation (Fig. 3B, top panel; quantified in Fig. 3C). We observed no differences in total ubiquitin levels, suggesting that protein degradation more generally was unaffected (Fig. 3D). However, more sensitive techniques to assess protein degradation rates may be required, and thus we cannot completely rule out the possibility that bulk differences in degradation are present.

To determine whether the observed differences in histone acetylation or variant usage were required for feeding or life span differences between the diets, we tested the effects of inhibition of histone deacetylases (HDACs), which prevented diet-dependent differences in histone acetylation and total histone H3 (Fig. 2C). Feeding flies the HDAC inhibitor sodium butyrate for one week abrogated increased feeding on low-BCAA diets, indicating that the histone acetylome could either encode hunger itself or be a permissive response to hunger that modulates feeding (Fig. 3E). To distinguish between these possibilities, we fed flies HDAC inhibitors for life (Fig. 3H) or removed the histone deacetylase *Rpd3* using the loss-of-function mutant $Rpd3^{def24}$ (Fig. 3I) (*46*, *47*) and found that both manipulations slightly extended the median life span of flies on high-BCAA food. Modulation of histone acetylation therefore influences both feeding and life span, supporting the idea that the histone acetylome might influence hunger rather than feeding per se (Fig. 3H). We also tested whether incorporation of histone variant H3.3 was required for the effects of hunger on feeding and life span by using pan-neuronally expressed RNAi to deplete the protein chaperone *Hira*, which is required for the exchange of H3 for H3.3 (*48*). This also abolished increased feeding on a low-BCAA diet (Fig. 3F) but had no effect on life span (Fig. 3G). Differential usage of histone H3/H3.3 therefore appears to be required for modulating feeding in response to hunger but is dispensable for hunger's effects on life span.

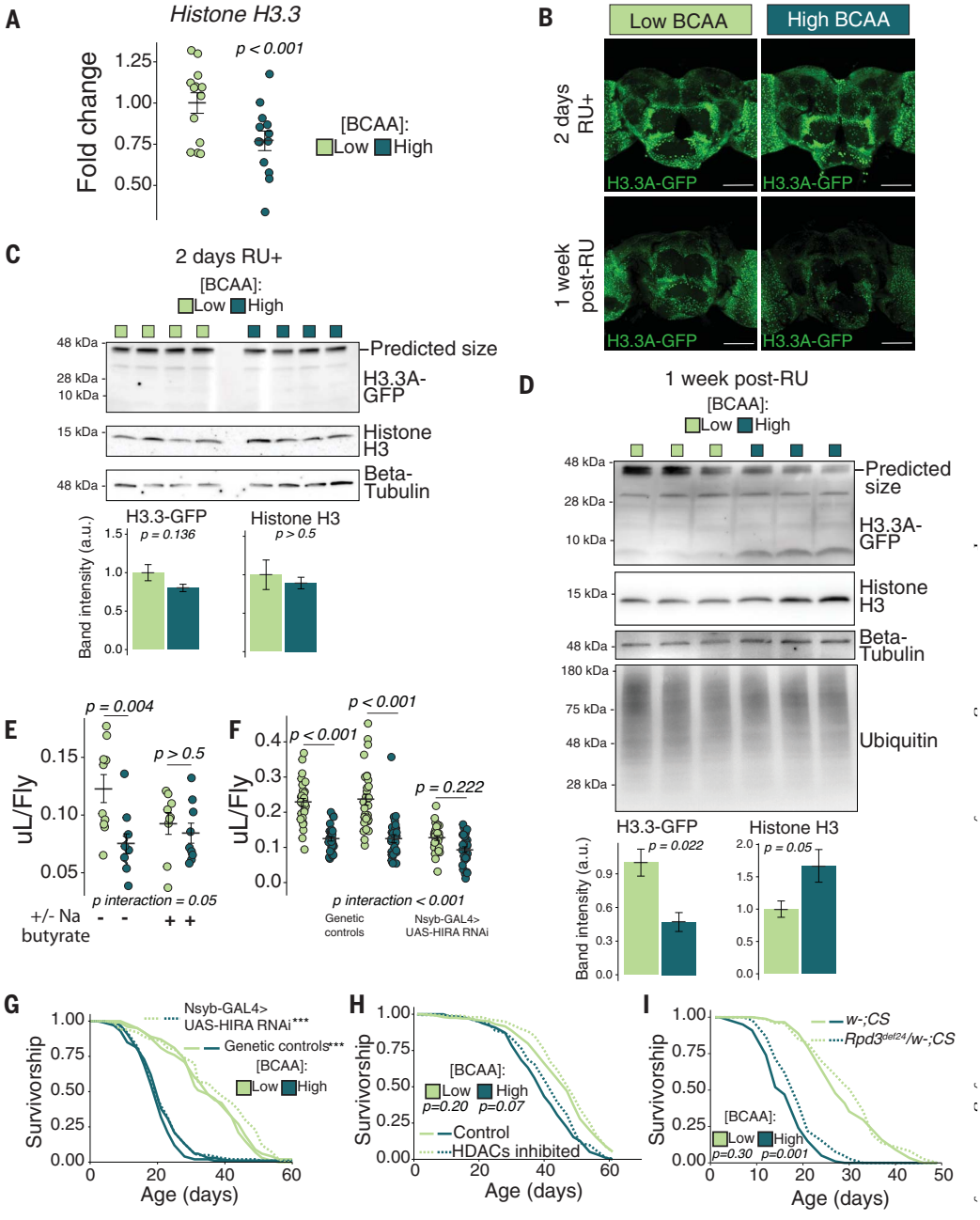


Fig. 3. Utilization of the histone variant H3.3A in the brain is modulated by BCAAs to stimulate feeding but is not required to extend life span. (**A**) RT-qPCR of relative mRNA abundance of *H3.3A* in *Canton-S* heads after 5 to 7 days on BCAA diets, values normalized to low-BCAA treatment (one-way ANOVA; 10 heads were used per biological replicate, 3 experimental replicates are pooled, n = 12 to 13 biological replicates per treatment). (**B** to **D**) Pulse-chase of fluorescently labelled H3.3A in brains (B) or heads (C and D) of flies carrying *Nsyb-GeneSwitch-GAL4>H3.3A-GFP* on BCAA diets. (B) Representative confocal images of brains immediately following the 2-day pulse of H3.3A-GFP induced by BCAA food + RU486 (top panel) and 1 week after the pulse (bottom panel). Scale bar 100 μm. (C) Western blots of H3.3A-GFP (predicted molecular weight = 42 kDa) and histone H3 following the 2-day pulse of H3.3A-GFP induced by BCAA food + RU486, quantified in bottom panel and normalized to low-BCAA treatment (one-way ANOVA; 10 heads were used per biological replicate, n = 4 biological replicates per treatment). (D) Western blots of H3.3A-GFP (one-way ANOVA), histone H3 (one-tailed Student's t-test), and ubiquitin (one-way ANOVA, P = 0.571) 1 week after the H3.3A-GFP pulse, quantified in bottom panel and normalized to low-BCAA treatment (10 heads were used per biological replicate, n = 3 biological replicates per treatment). (**E**) Con-Ex measurement of 24-hour food intake after 5 days on BCAA diets ± 100 mM sodium butyrate (two-way ANOVA with Tukey's post hoc). (**F**) Con-Ex measurement of 24-hour food intake after 5 to 7 days on BCAA diets from flies carrying *Nsyb-GAL4>UAS-HIRA RNAi* or controls (*Nsyb-GAL4/w-;CS* and *UAS-HIRA RNAi/w-;*CS) (two-way ANOVA with Tukey's post hoc, 3 experimental replicates are pooled, n = 29 to 30 flies per treatment). (**G**) Life spans of flies carrying *Nsyb-GAL4>UAS-HIRA RNAi* or controls (*Nsyb-GAL4/w-;CS* and *UAS-HIRA RNAi/w-;CS*) (log-rank test, n = 184 to 202). (**H**) Life span of flies on BCAA diets ± 100 mM sodium butyrate (log-rank test, n = 186 to 196). (**I**) Life span of flies heterozygous for a null mutation in *Rpd3* or *w-;CS* controls on BCAA diets (log-rank test, n = 85 to 105). ***P < 0.001. Error bars represent the mean ± SEM.

Metabolic intermediates link BCAAs to feeding and life span

BCAAs are unusual amino acids because in mammalian systems they bypass metabolism in the liver and are instead metabolized in target tissues by the enzymes branched-chain aminotransferase (BCAT) and branched-chain alpha-ketoacid dehydrogenase (BCKDH) (*42*, *49*, *50*). *BCAT* mRNA abundance was decreased in the heads of flies fed low-BCAA diets (Fig. 4A). Depletion of *BCAT*, a homolog of mammalian *BCAT2*, with the pan-neuronal Nsyb-GAL4 driver reduced histone H3 abundance to similar levels on both diets (Fig. 4B). The Fly Cell Atlas indicates that *BCAT* is expressed in the ~80 neurons that also coexpress the promoter used to label R50H05 hunger neurons (*51*) (fig. S5). RNAi-mediated *BCAT* depletion only in the R50H05 hunger neurons eliminated the effect of low-BCAA diets to reduce histone H3 and reduced the magnitude of the effect on reduction of H3K9ac (Fig. 4C). It also decreased diet-dependent differences in life span by significantly extending the life span of flies on high-BCAA diet (Fig. 4D). This is unlikely to be a result of deficiencies in neuronal activity upon depletion of *BCAT* because optogenetic inhibition of the R50H05 hunger neurons had no effect on the life span of flies fed high-BCAA diets (fig. S6). These findings indicate that R50H05 hunger neurons require BCAT, possibly to use BCAA metabolites to modulate aging through a mechanism that does not require neuronal depolarization. However, *BCAT* depletion in hunger neurons, rather than reduce feeding differences between diets, seemed to do the

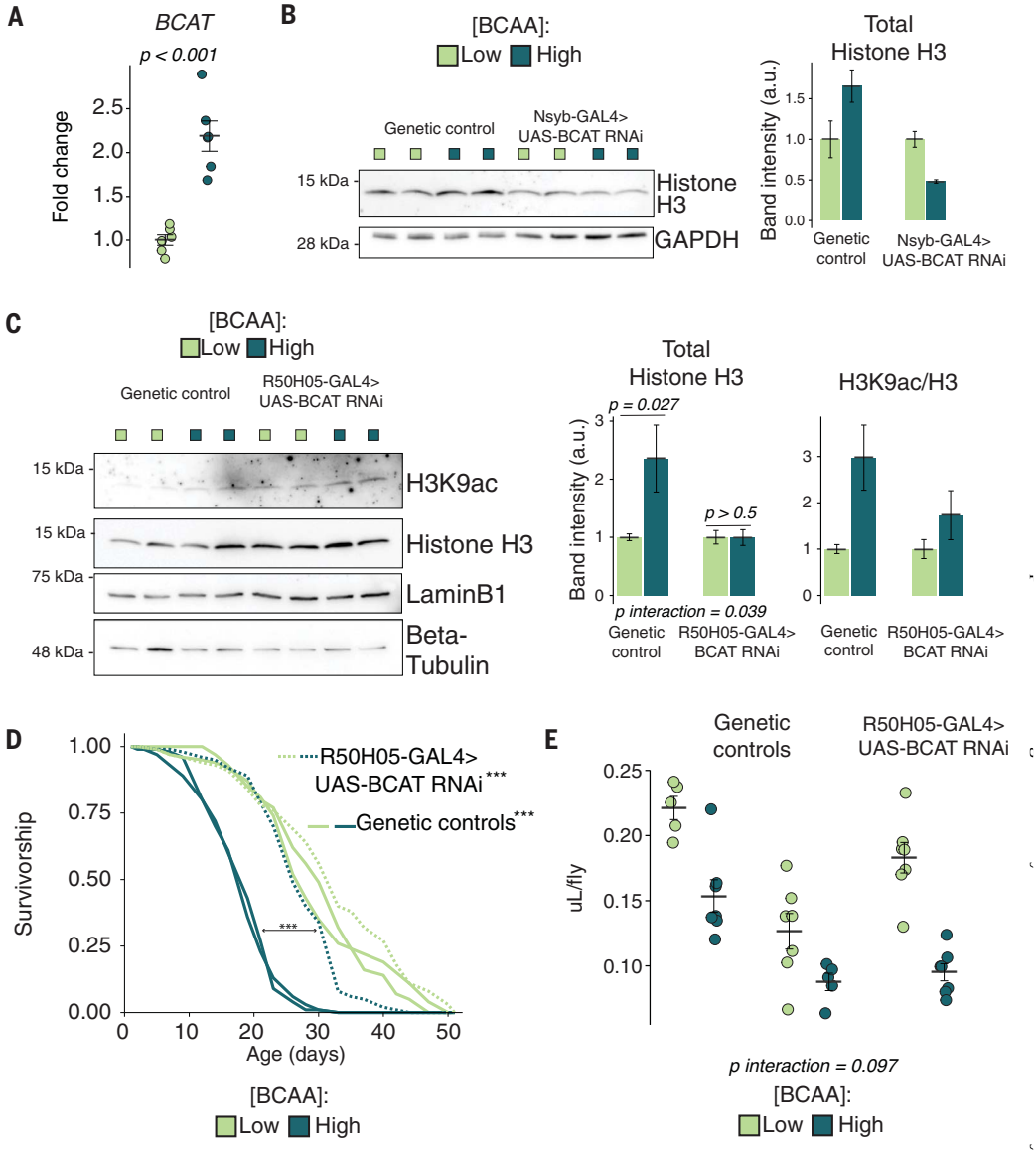


Fig. 4. Hunger neurons use BCAA metabolism to regulate histone abundance and modulate life span independent of feeding. (**A**) RT-qPCR quantification of relative mRNA abundance of *BCAT* (*CG1673*) in *Canton-S* heads after 5 to 7 days on BCAA diets, values normalized to low-BCAA treatment (one-way ANOVA, 10 heads were used per biological replicate, n = 5 to 6 biological replicates). (**B**) Western blot of histone H3 in heads of flies carrying *Nsyb-GAL>UAS-BCAT RNAi* or control (*Nsyb-GAL4/w-;CS*), quantified in right panel and normalized to low-BCAA treatment within genotypes (two-way ANOVA with Tukey's post hoc, n = 2 biological replicates per treatment). (**C**) Western blots of histone H3 (3 experimental replicates were pooled, n = 5 biological replicates per treatment) and H3K9ac (n = 2 biological replicates per treatment) in heads of flies carrying *R50H05-GAL4>UAS-BCAT RNAi* or control (*R50H05-GAL4/w-;CS*), quantified in right panel and normalized to low-BCAA treatment within genotypes (two-way ANOVA with Tukey's post hoc; 10 heads were used per biological replicate). (**D**) Life spans of flies carrying *R50H05-GAL4>UAS-BCAT RNAi* or controls (*R50H05-GAL4/w-;CS* or *UAS-BCAT RNAi/w-;CS*) on low- or high-BCAA diets (log-rank test, n = 91 to 100, ***P < 0.001). (**E**) Con-Ex measurement of 24-hour food intake after 5 days on BCAA diets in flies carrying *R50H05-GAL4>UAS-BCAT RNAi* or controls (*R50H05-GAL4/w-;CS* or *UAS-BCAT RNAi/w-;CS*) (two-way ANOVA, n = 5 to 7 flies per treatment). Error bars represent the mean ± SEM.

opposite by increasing the magnitude of the difference (Fig. 4E). These results indicated complexity in the circuits that increase feeding and modulate aging in response to hunger and reinforce the notion that they are, at least in part, anatomically distinct.

Considerations on the divergence of mechanisms linking hunger with feeding and aging

A subset of the R50H05 hunger neurons produce the neuromodulator serotonin and express the genes *Trh* and *Henna*, which encode tryptophan hydroxylase and tryptophan phenylalanine hydroxylase, respectively, that function as key modulatory steps in serotonin synthesis. Expression of both genes was significantly increased in the heads of flies that were BCAA-restricted (Fig. 5A). The low-BCAA diet also increased antibody staining of serotonin in the cell bodies of the serotonergic posterior lateral protocerebrum (PLP) cluster, a subset of the R50H05 hunger neurons (Fig. 5, B and C, quantified in D). Given that inhibition of BCAT in the R50H05 neurons seemed to slightly exacerbate the feeding differences in our BCAA diets and that serotonin can either promote or suppress feeding depending on which serotonergic neurons are manipulated (*35*), we tested whether *BCAT* depletion in the serotonergic network more broadly may produce opposite effects on feeding compared with what we observed after *BCAT* depletion in the R50H05 subset. We used the Trh-GAL4 driver, which is putatively expressed in all serotonergic neurons of the central nervous system, to target depletion of *BCAT* in these cells. This manipulation decreased feeding overall and prevented increased feeding on low-BCAA diets, indicating that BCAA metabolites may have variable effects on distinct populations of neurons to regulate feeding and aging (Fig. 5E). We found further support that BCAAs may have spatially restricted effects in the observation that transcriptional profiling of whole brains from flies fed RD, low-BCAA, or high-BCAA (fig. S7) showed no obvious separation of brain-wide transcriptional profiles, although we identified many differentially expressed genes (fig. S7B). This reinforces the notion that neuronal effects of BCAAs may differ between distinct populations of neurons.

Intrigued by the possibility that hunger effects on feeding and aging may be regulated by partially distinct pathways, we questioned whether some of this complexity might result from the time scale on which they occur. Although we consider life span to be an outcome of chronic dietary effects, measurements of feeding quantify more acute responses that may or may not be stable over longer time periods. The increased feeding on low-BCAA food that was present after five days on the diet

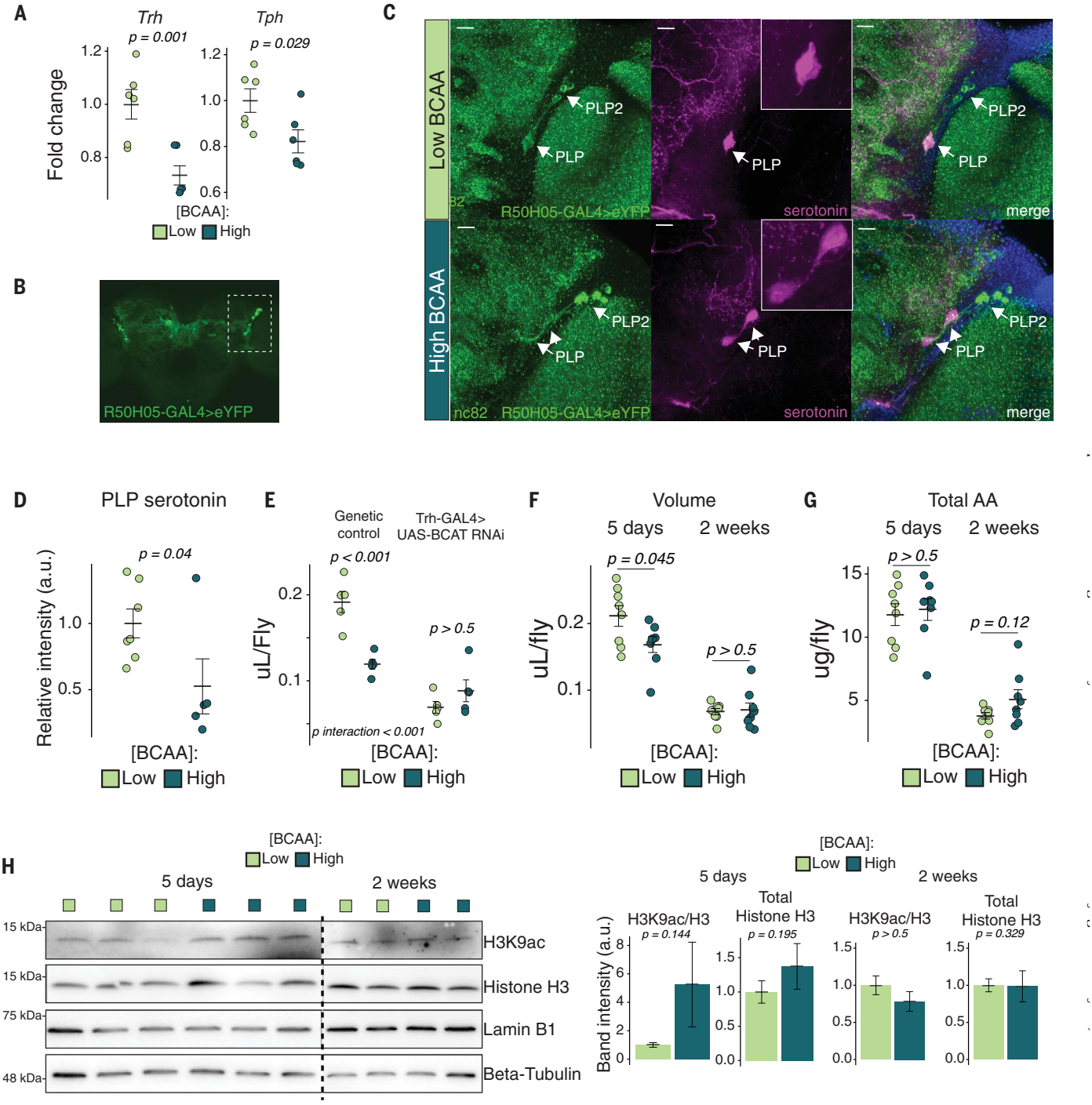


Fig. 5. Spatially and temporally distinct processes link hunger to feeding and aging. (**A**) RT-qPCR quantification of relative mRNA abundance of *Trh* and *Tph* in *Canton-S* heads after 5 to 7 days on BCAA diets, values are normalized to low-BCAA treatment (one-way ANOVA; 10 heads were used per biological replicate, $n = 6$ biological replicates per treatment). (**B** to **D**) Immunostaining for serotonin in PLP neuron cell bodies in flies carrying *R50H05-GAL4>UAS-GTACR.eYFP* after 5 to 7 days on BCAA diets. (B) Maximum intensity projection of ten 1.5-μm stacks highlighting the PLP neurons (white box) in the posterior fly brain. Scale bar 100 μm. (C) Representative confocal images of immunostaining for serotonin in PLP neurons of flies carrying *R50H05-GAL4>UAS-GTACR.eYFP* (green, nc82; magenta, serotonin; blue, DAPI; scale bar = 10 μm). Images are maximum intensity projections of ten 1.5-μm stacks through the PLP neurons (D) Quantification of (C) as described in Methods (one-way ANOVA, 3 experimental replicates are pooled, $n = 5$ to 7 brains per treatment). (**E**) Con-Ex measurement of 24-hour food intake after 5 days on BCAA diets in flies carrying *Trh-GAL4>UAS-BCAT RNAi* or control (*Trh-GAL4/w-;CS)* (two-way ANOVA, $n = 5$ flies per treatment). (**F** to **G**) Con-Ex measurement of 24-hour food intake after 5 days or 2 weeks on BCAA diets in *Canton-S* flies; volume (F) and total amino acid intake (G) (two-tailed Student's t-tests, $n = 8$ flies per treatment). (**H**) Western blots of histone H3 and H3K9ac in heads of *Canton-S* flies after 5 days versus 2 weeks on BCAA diets, quantified in right panel (one-tailed Student's t-tests, 10 heads were used per biological replicate, 2 experimental replicates are pooled, $n = 3$ to 5 biological replicates per treatment). Error bars represent the mean ± SEM.

receded after two weeks (Fig. 5F), resulting in reduced amino acid intake overall (Fig. 5G). Changes in brain histone acetylation observed after 5 days were also reversed or eliminated after two weeks on the diet (Fig. 5H). Thus, flies on low-BCAA diets presumably adapt to their food environment in a way that decreases the amount of protein they require, which may reduce hunger and have beneficial consequences for aging (see model, fig. S8).

Discussion

We present behavioral and molecular evidence consistent with the ideas that the neural state of hunger modulates aging and that this motivational state is molecularly encoded by patterns of neuronal histone modifications. Increased feeding and extended life span are consequences that seem to be orchestrated by downstream responses to hunger. Our findings—that dietary BCAAs modulated use of the histone variant H3.3, that this was required for increasing feeding but dispensable for increasing life span, and that distinct populations of serotonergic neurons require BCAA metabolism to modulate either feeding or aging—all indicate that various neurons use different strategies to generate responses to hunger, but all seem to involve modifications to histone proteins. Such complexity is consistent with emerging reports of interactions between metabolism and histone modifications that are driven in part by specialized enzymes that reside in distinct cellular compartments and in some cases distinct nuclear subcompartments (*52*–*54*). In *C. elegans,* H3.3 is important for engaging transcriptional networks that underlie neuronal plasticity in response to environmental manipulations (*55*, *56*). Perhaps neuronal plasticity is one strategy used by circuits that promote feeding to generate behavioral responses to hunger. In *Drosophila,* protein appetite requires plastic changes in a dopaminergic circuit, although a role for histone variant usage in this process has not been determined (*33*). Thus, the cell-type specificity of the usage of histone variant H3.3 and the genomic loci to which it is targeted in response to dietary BCAAs remain unknown.

The observation that limitation of BCAAs acutely increased appetite but that this eventually subsided indicates that hunger may act as an allostatic stressor that acutely increases feeding whereas chronic hunger may promote physiological changes that lower a set point around which appetite is regulated (*57*, *58*). It might be this adaptive response, mediated by modifications to histone proteins in discrete neural circuits, that slows aging. This idea is consistent with several dietary manipulations that are known to extend life span: Low protein diets (and methionine restriction in particular) generate protein-specific appetites and intermittent fasting paradigms putatively increase the frequency with which an animal experiences hunger without affecting total caloric intake (*5*, *59*–*63*).

Our data support a role for the modification of histone proteins as a link between hunger, feeding, and aging. In the brain, transcriptional changes occur in response to dietary manipulations; however, analysis of single "longevity genes"—the list of which is continually growing—is typically insufficient to generalize about how diet modulates aging. The effects could require a cell-type–specific overhaul of transcriptional programming that occurs in specialized hunger neurons, including the R50H05 neurons described here (*64*–*66*). In this view, patterns of histone modifications, gene expression profiles that determine cell identity, and intracellular signaling pathways that detect peripheral hunger information, converge to determine transcriptional states which, although they are most often studied in the context of neuronal-fate determination, may also tune neuronal states in response to dietary or other environmental manipulations (*67*, *68*).

REFERENCES AND NOTES

1. L. Fontana, L. Partridge, *Cell* **161**, 106–118 (2015).
2. W. Mair, M. D. Piper, L. Partridge, *PLOS Biol.* **3**, e223 (2005).
3. D. A. Skorupa, A. Dervisefendic, J. Zwiener, S. D. Pletcher, *Aging Cell* **7**, 478–490 (2008).
4. S. M. Solon-Biet *et al*., *Cell Metab.* **19**, 418–430 (2014).
5. S. J. Simpson *et al*., *Ageing Res. Rev.* **39**, 78–86 (2017).
6. M. L. Andermann, B. B. Lowell, *Neuron* **95**, 757–778 (2017).
7. J. Berrios *et al*., *Nature* **595**, 695–700 (2021).
8. J. Apfeld, C. Kenyon, *Nature* **402**, 804–809 (1999).
9. M. J. Waterson *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8137–8142 (2014).
10. I. Ostojic *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8143–8148 (2014).
11. S. Libert *et al*., *Science* **315**, 1133–1137 (2007).
12. M. A. Vargas, N. Luo, A. Yamaguchi, P. Kapahi, *Curr. Biol.* **20**, 1006–1011 (2010).
13. C. Ribeiro, B. J. Dickson, *Curr. Biol.* **20**, 1000–1005 (2010).
14. A. M. Felton *et al*., *Behav. Ecol.* **20**, 685–690 (2009).
15. E. D. Smith *et al*., *BMC Dev. Biol.* **8**, 49 (2008).
16. B. Zhang, H. Jun, J. Wu, J. Liu, X. Z. S. Xu, *Nat. Aging* **1**, 255–268 (2021).
17. J. Ro *et al*., *eLife* **5**, e16843 (2016).
18. S. M. Solon-Biet *et al*., *Nat. Metab.* **1**, 532–545 (2019).
19. P. Juricic, S. Grönke, L. Partridge, *J. Gerontol. A Biol. Sci.* **75**, 24–31 (2020).
20. M. D. Piper *et al*., *Nat. Methods* **11**, 100–105 (2014).
21. S. C. Johnson, P. S. Rabinovitch, M. Kaeberlein, *Nature* **493**, 338–345 (2013).
22. N. E. Richardson *et al*., *Nat. Aging* **1**, 73–86 (2021).
23. J. M. C. McDonald, P. Nabili, L. Thorsen, S. Jeon, A. W. Shingleton, *Evodevo* **12**, 6 (2021).
24. M. F. Camus, M. D. Piper, M. Reuter, *eLife* **8**, e47262 (2019).
25. B. C. Shell *et al*., *Sci. Rep.* **8**, 11536 (2018).
26. B. C. Shell, Y. Luo, S. Pletcher, M. Grotewiel, *Sci. Rep.* **11**, 20044 (2021).
27. T. Awasaki, K. Kimura, *J. Neurobiol.* **32**, 707–721 (1997).
28. Y. D. Chen, S. J. Park, W. W. Ja, A. Dahanukar, *Front. Cell. Neurosci.* **12**, 382 (2018).
29. K. P. Lee *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2498–2503 (2008).
30. K. Steck *et al*., *eLife* **7**, e31625 (2018).
31. J. Ro, Z. M. Harvanek, S. D. Pletcher, *PLOS ONE* **9**, e101107 (2014).
32. Y. Lyu *et al*., *eLife* **10**, e59399 (2021).
33. Q. Liu *et al*., *Science* **356**, 534–539 (2017).
34. S. Lin, B. Senapati, C. H. Tsao, *Open Biol.* **9**, 180259 (2019).
35. S. D. Albin *et al*., *Curr. Biol.* **25**, 2435–2440 (2015).
36. J. C. Johnson, A. S. Munneke, H. M. Richardson, C. M. Gendron, S. D. Pletcher, *Aging* **15**, 396–420 (2023).
37. P. Mews *et al*., *Nature* **546**, 381–386 (2017).
38. K. E. Wellen *et al*., *Science* **324**, 1076–1080 (2009).
39. C. López-Otín, M. A. Blasco, L. Partridge, M. Serrano, G. Kroemer, *Cell* **153**, 1194–1217 (2013).
40. M. Herre, E. Korb, *Curr. Opin. Neurobiol.* **59**, 79–86 (2019).
41. M. A. Reid, Z. Dai, J. W. Locasale, *Nat. Cell Biol.* **19**, 1298–1306 (2017).
42. J. T. Brosnan, M. E. Brosnan, *J. Nutr.* **136**, 207S–211S (2006).
43. K. Ahmad, S. Henikoff, *Mol. Cell* **9**, 1191–1200 (2002).
44. A. Sakai, B. E. Schwartz, S. Goldstein, K. Ahmad, *Curr. Biol.* **19**, 1816–1820 (2009).
45. A. Tvardovskiy, V. Schwämmle, S. J. Kempf, A. Rogowska-Wrzesinska, O. N. Jensen, *Nucleic Acids Res.* **45**, 9272–9289 (2017).
46. B. Rogina, S. L. Helfand, S. Frankel, *Science* **298**, 1745 (2002).
47. K. Suda *et al*., *Exp. Cell Res.* **385**, 111673 (2019).
48. B. Loppin *et al*., *Nature* **437**, 1386–1390 (2005).
49. D. Biswas, L. Duffley, T. Pulinilkunnil, *FASEB J.* **33**, 8711–8731 (2019).
50. M. D. Neinast *et al*., *Cell Metab.* **29**, 417–429.e4 (2019).
51. H. Li *et al*., *Science* **375**, eabk2432 (2022).
52. S. Sivanand, I. Viney, K. E. Wellen, *Trends Biochem. Sci.* **43**, 61–74 (2018).
53. R. Boon, G. G. Silveira, R. Mostoslavsky, *Nat. Metab.* **2**, 1190–1203 (2020).
54. K. W. Ryu *et al*., *Science* **360**, eaan5780 (2018).
55. I. Maze *et al*., *Neuron* **87**, 77–94 (2015).
56. A. Piazzesi *et al*., *Cell Rep.* **17**, 987–996 (2016).
57. B. S. McEwen, *Ann. N. Y. Acad. Sci.* **840**, 33–44 (2006).
58. K. C. Berridge, *Physiol. Behav.* **81**, 179–209 (2004).
59. R. Leitão-Gonçalves *et al*., *PLOS Biol.* **15**, e2000862 (2017).
60. B. C. Lee *et al*., *Nat. Commun.* **5**, 3592 (2014).
61. N. Orentreich, J. R. Matias, A. DeFelice, J. A. Zimmerman, *J. Nutr.* **123**, 269–274 (1993).
62. A. Di Francesco, C. Di Germanio, M. Bernier, R. de Cabo, *Science* **362**, 770–775 (2018).
63. J. H. Catterson *et al*., *Curr Biol* **28**, 1714–1724.e4 (2018).
64. M. D. Piper, A. Bartke, *Cell Metab.* **8**, 99–104 (2008).
65. D. Wilinski *et al*., *Nat. Commun.* **10**, 4052 (2019).
66. A. Vaziri *et al*., *Sci. Adv.* **6**, eabc8492 (2020).
67. J. Janssens *et al*., *Nature* **601**, 630–636 (2022).
68. T. Tsukahara *et al*., *Cell* **184**, 6326–6343.e32 (2021).
78. H. Li *et al*., *Science* **375**, eabk2432 (2022).

ACKNOWLEDGMENTS

We wish to thank past and present members of the Pletcher laboratory for their support and comments about the experimental design of these studies, and we express gratitude to D. Paris for his help engineering and creating some tools used in this study. We also thank members of the Lombard laboratory for sharing antibodies and protocols. We acknowledge B. Jacobovitz in the Michigan Medicine Microscopy Core for training and advice on confocal imaging and the Advanced Genomics Core at the University of Michigan. **Funding:** This research was supported by The National Science Foundation Graduate Research Fellowship Program (DGE 1256260) and the Howard Hughes Medical Institute through the James H. Gilliam Fellowships for Advanced Study program to K.J.W. (GT11426), and the US National Institutes of Health, National Institute on Aging (R01 AG051649, R01 AG030593, and R01 AG063371) and the Glenn Medical Foundation (to S.D.P.) **Author contributions:** Conceptualization: K.J.W. and S.D.P. Methodology: K.J.W. and S.D.P. Investigation: K.J.W., R.A.H., E.H., and S.D.P. Visualization: K.J.W., E.H., and Y.L. Funding acquisition: K.J.W. and S.D.P. Project administration: K.J.W. and S.D.P. Supervision: K.J.W. and S.D.P. Writing – original draft: K.J.W. and S.D.P. Writing – review and editing: K.J.W. and S.D.P. **Competing interests:** The senior author (S.D.P.) is a share holder in the company Flidea, which has developed technology related to the FLIC feeding system. **Data and materials availability:** All data are available in the manuscript or the supplementary material. FLIC data were analyzed using custom R code, which is available at https://github.com/PletcherLab/FLIC_R_Code. High-throughput sequencing raw data files can be found on Gene Expression Omnibus GSE229336. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade1662
Materials and Methods
Figs. S1 to S8
Tables S1 to S3
References (*69*–*77*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 28 July 2022; accepted 13 April 2023
Published online 4 May 2023
10.1126/science.ade1662

ZEOLITE STRUCTURE

Three-dimensional inhomogeneity of zeolite structure and composition revealed by electron ptychography

Hui Zhang[1,2,3]*†, Guanxing Li[3]†, Jiaxing Zhang[4], Daliang Zhang[5,6], Zhen Chen[7], Xiaona Liu[8], Peng Guo[8], Yihan Zhu[9,10], Cailing Chen[3], Lingmei Liu[5,6], Xinwen Guo[4], Yu Han[3]*

Structural and compositional inhomogeneity is common in zeolites and considerably affects their properties. Thickness-limited lateral resolution, lack of depth resolution, and electron dose-constrained focusing limit local structural studies of zeolites in conventional transmission electron microscopy (TEM). We demonstrate that a multislice ptychography method based on four-dimensional scanning TEM (4D-STEM) data can overcome these limitations. Images obtained from a ~40-nanometer-thick MFI zeolite exhibited a lateral resolution of ~0.85 angstrom that enabled the identification of individual framework oxygen (O) atoms and the precise determination of the orientations of adsorbed molecules. Furthermore, a depth resolution of ~6.6 nanometers allowed probing of the three-dimensional distribution of O vacancies, as well as the phase boundaries in intergrown MFI and MEL zeolites. The 4D-STEM ptychography can be generally applied to other materials with similar high electron-beam sensitivity.

Zeolites are aluminosilicate microporous crystals with regular intracrystalline cavities and channels of molecular dimensions. The molecular sieving ability of zeolites renders highly desirable size and shape selectivity for adsorption, separation, and catalysis (*1*–*4*). However, the complexity of zeolite structures makes them prone to structural and compositional inhomogeneity that can profoundly affect their properties. For example, the precursors and synthetic conditions affect the location and distribution of functional sites (such as Al pairs and heteroatoms) in the zeolite framework (*5*). Postsynthesis calcination generates oxygen (O) vacancies to render Lewis acid sites in an uncontrolled manner (*6*, *7*). The intergrowth of multiple polymorphs and various phases is common in zeolite crystals (*8*–*10*). The prevalence of inhomogeneity explains why zeolites with the same topology and composition prepared from different batches often exhibit variability in applications (*11*, *12*).

Although inhomogeneity in zeolites has been widely recognized, the precise determination of the related local nonperiodic structures remains challenging given the lack of suitable characterization techniques. For instance, the electron-beam sensitivity of zeolites limits the use of transmission electron microscopy (TEM) to study their structures (*13*–*15*) because the low-electron doses required to prevent structural damage to zeolites often result in images with limited resolution and poor signal-to-noise ratios. New imaging methods that use electron doses more efficiently are needed to mitigate this issue (*16*, *17*).

Among emerging low-dose TEM techniques, integrated differential phase-contrast scanning TEM (iDPC-STEM) has proven effective for imaging zeolite structures, especially for studying guest species located in the micropores (*18*–*20*). Despite the demonstrated successes, iDPC-STEM has limitations. First, the image resolution of iDPC-STEM is severely limited by the specimen thickness. For zeolites, iDPC-STEM can only resolve framework T atoms (where T is Si or Al) (strictly speaking, atomic columns) but not the O atoms between them unless the specimen thickness is within a few unit cells (<10 nm), which is extremely rare in actual samples. Second, like other conventional TEM imaging modes, iDPC-STEM lacks resolving power along the projection direction and cannot identify longitudinal structural inhomogeneity inside the specimen. Third, iDPC-STEM requires precise focusing of the electron beam to achieve high resolution, which could lead to a lower success rate for beam-sensitive materials because structural damage may occur during the fine-tuning of the focus.

Here, we demonstrate that electron ptychography (*21*–*26*) can address these problems and achieve subangstrom resolution to resolve individual O atomic columns in various zeolites with specimen thicknesses up to ~40 nm, which has not been reported with other direct-imaging techniques. Electron ptychography is a coherent diffractive imaging method that uses a series of convergent-beam electron diffraction patterns, commonly referred to as four-dimensional STEM (4D-STEM) data, to generate high-resolution images. The identification of O atoms enabled a semiquantitative analysis of the three-dimensional (3D) distribution of O vacancies throughout the zeolite specimen. Furthermore, 4D-STEM ptychography provided ~6.6-nm depth resolution to reveal the growth mode of MEL zeolite domains along the *b* axis of the MFI zeolite. In addition, acquiring a 4D-STEM ptychography dataset did not require precise focusing of the electron probe, which considerably improved the efficiency of obtaining atomic-resolution phase-contrast images of zeolites and other beam-sensitive materials.

Comparison of iDPC-STEM and 4D-STEM ptychography

Image simulations were first performed to demonstrate the superiority of 4D-STEM ptychography over iDPC-STEM, the state-of-the-art technique for zeolite imaging. A series of 4D-STEM datasets were simulated for [010]-projected zeolite ZSM-5 by varying the specimen thicknesses and convergence angles (see materials and methods for details). Ptychography and iDPC-STEM images were calculated from each dataset. Regardless of the convergence angle, when the sample thickness exceeded 12 nm, iDPC-STEM images started to deviate from the structural model and became difficult to interpret. By contrast, ptychography images were robust, displaying substantial tolerance for variation of up to 40 nm in specimen thickness (Fig. 1A and fig. S1).

We verified the advantage of ptychography experimentally using specially synthesized ZSM-5 zeolite crystals. These ZSM-5 crystals had a highly uniform thickness of 30 to 40 nm along the *b* axis, as evidenced by multiple characterizations (fig. S2), which allowed a reasonable comparison between iDPC-STEM and ptychography. The iDPC-STEM images were acquired using conditions optimized for zeolite imaging (convergence

[1]Electron Microscopy Center, South China University of Technology, Guangzhou 510640, China. [2]School of Emergent Soft Matter, South China University of Technology, Guangzhou 510640, China. [3]Advanced Membranes and Porous Materials Center, Physical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia. [4]State Key Laboratory of Fine Chemicals, Frontiers Science Center for Smart Materials, School of Chemical Engineering, Dalian University of Technology, Dalian 116024, China. [5]Multi-scale Porous Materials Center, Institute of Advanced Interdisciplinary Studies, Chongqing University, Chongqing 400044, China. [6]School of Chemistry and Chemical Engineering, Chongqing University, Chongqing 400044, China. [7]School of Materials Science and Engineering, Tsinghua University, Beijing 100084, China. [8]National Engineering Research Center of Lower-Carbon Catalysis Technology, Dalian National Laboratory for Clean Energy, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China. [9]Center for Electron Microscopy, State Key Laboratory Breeding Base of Green Chemistry Synthesis Technology and College of Chemical Engineering, Zhejiang University of Technology, Hangzhou 310014, China. [10]Institute for Frontier and Interdisciplinary Sciences, Zhejiang University of Technology, Hangzhou 310014, China.
*Corresponding authors. Email: hui.materials.zhang@gmail.com (H.Z.); yu.han@kaust.edu.sa (Y.H.) †These authors contributed equally to this work.

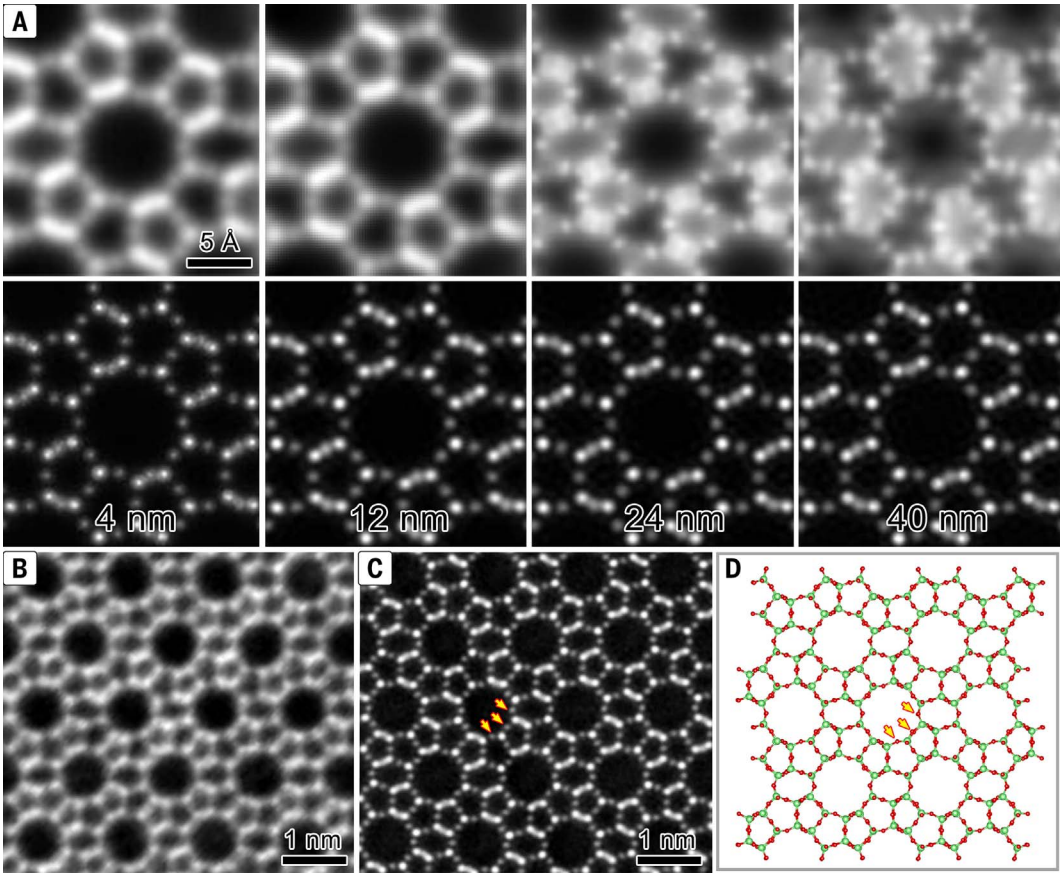


Fig. 1. Comparison of iDPC-STEM and 4D-STEM ptychography. (**A**) Simulated iDPC-STEM (top) and 4D-STEM ptychography (bottom) images of ZSM-5 along the [010] axis with varying thicknesses, as indicated, under a convergence semiangle of 15 mrad. (**B**) iDPC-STEM image of ZSM-5 along the [010] axis, with a specimen thickness of ~10 nm. (**C**) 4D-STEM ptychography image of ZSM-5 along the [010] axis, with a specimen thickness of ~40 nm. (**D**) The [010]-projected structural model of ZSM-5. The yellow arrows in (C) and (D) indicate several O atomic columns in the framework, which are not resolved in (B).

semiangle: 15 mrad; probe current: 2 pA; dwell time: 10 μs; pixel size: 0.380 Å by 0.380 Å), with a total electron dose of ~900 $e^-/Å^2$. The 4D-STEM datasets were collected by using a high-performance hybrid electron microscope pixel array detector (EMPAD) with a probe current of ~0.09 pA, scan step of 0.399 Å, and 256-by-256 probe positions. Limited by the EMPAD frame rate (1000 frames/s), the specimen was subjected to an electron dose of ~3500 $e^-/Å^2$ during data acquisition. However, this dose was still low enough to preserve the structure of ZSM-5. The iterative ptychographic reconstruction used the maximum likelihood (ML) method to update the exit wave function and the linear least squares (LSQ) method to update the object and probe functions (*27*, *28*). Multislice (*29*) and mixed-state (*30*) algorithms were implemented to address the multiple scattering associated with thick specimens and the partial incoherence of electron probes, respectively. Successful reconstruction at low-dose conditions relies on the best possible initial guess for the probe function, which can be facilitated by minimizing aberrations using the data acquisition strategy described in fig. S3.

Unambiguous identification of O in the zeolite framework requires a resolution near 1 Å and good image contrast, which is rarely achievable with iDPC-STEM, given its sensitivity to sample thickness and time-constrained focusing unless the specimen is extremely thin (*20*). The iDPC-STEM image taken from ~40-nm-thick ZSM-5 did not match the [010]-projected structural model and is difficult to interpret (fig. S4), which is consistent with the simulation. When the imaging was performed on the crystal periphery, where the thickness was ~10 nm, the acquired iDPC-STEM image revealed 5-, 6-, and 10-membered rings (MRs) surrounded by T atoms; however, the O atoms located between the T atoms remained unresolvable (Fig. 1B).

In the ptychographic phase image of ~40-nm-thick ZSM-5 (Fig. 1C), all of the framework atoms, including O atoms, are unequivocally resolved. This image exhibited high precision in matching the projected structural model (Fig. 1D) with structural information transferred up to 0.85 Å, as determined from its Fourier transform pattern (fig. S5). By comparison, the highest resolution achieved in the STEM imaging of zeolites in the past was ~1 Å (*31*). Notably, the 4D-STEM data were collected at a high magnification after only coarse focusing at low magnification. Ptychography reconstruction revealed that the electron probe was ~35-nm defocused (fig. S6), which could not provide atomic resolution using conventional STEM modes. For beam-sensitive materials, an advantage of 4D-STEM ptychography is not needed to fine-tune the beam focus.

Similar results were obtained when ZSM-5 was imaged along the [100] axis. Thin specimens in this orientation were obtained by crushing commercial micrometer-sized ZSM-5 crystals into small fragments. Although the exact specimen thicknesses were unknown, we selected the highest-quality iDPC-STEM and ptychographic images obtained for comparison (Fig. 2, A and B). In this projection, four closely spaced T atomic columns (T-T distances: 0.8 to 1.8 Å; Fig. 2C) were indiscernible in the iDPC-STEM image (Fig. 2A) but were clearly separated in the ptychographic image (Fig. 2B), owing to the higher resolution achieved by ptychography compared with iDPC-STEM (fig. S5). The separation of these four atomic columns had not previously been achieved by imaging. When used for other types of zeolites—such as EMM-17, which has unusual 11-MR channels—4D-STEM ptychography also demonstrated higher resolution than iDPC-STEM and the ability to identify framework O atoms (Fig. 2, D to F).

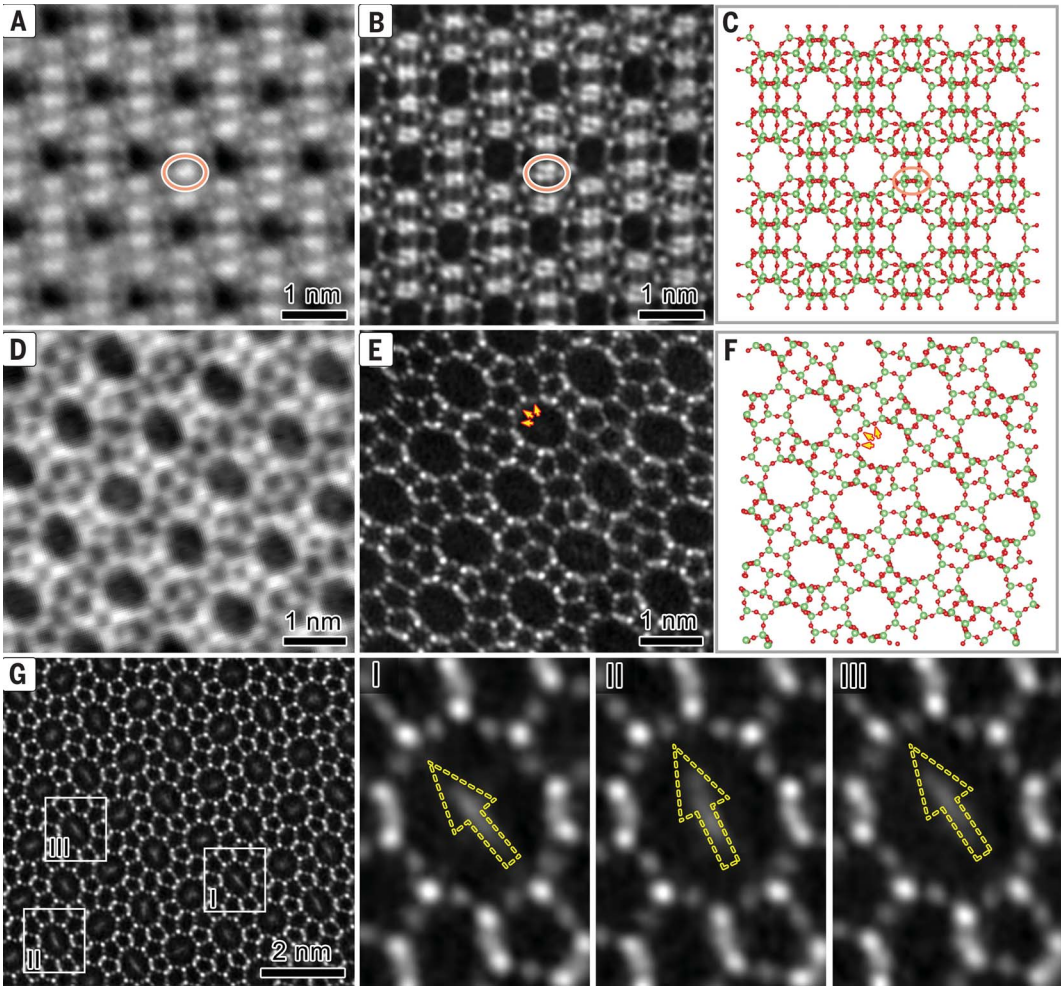


Fig. 2. Subangstrom imaging of zeolites and adsorbates using 4D-STEM ptychography. (**A** and **B**) iDPC-STEM (A) and 4D-STEM (B) ptychography images of ZSM-5 along the [100] axis. (**C**) The [100]-projected structural model of ZSM-5. The orange ellipses in (A) to (C) highlight the same positions, where four closely spaced columns of T atoms are resolved in (B) but not in (A). (**D** and **E**) iDPC-STEM (D) and 4D-STEM (E) ptychography images of EMM-17 along the characteristic 11-MR channel direction (i.e., the [001] axis). (**F**) The [001]-projected structural model of EMM-17. The yellow arrows in (E) and (F) indicate several O atomic columns in the framework, which are not resolved in (D). (**G**) 4D-STEM ptychographic image of ZSM-5 with adsorbed PX molecules along the straight 10-MR channel (i.e., the [010] axis). The square regions labeled I, II, and III are enlarged on the right to reveal the subtly different orientations of the PX molecules, as reflected by the rod-like contrast in the channels. The arrows are used as visual guides.

Resolving adsorbed molecules and oxygen vacancies

We used 4D-STEM ptychography to image ZSM-5 with adsorbed *p*-xylene (PX) molecules. The reconstructed ptychographic image displays rod-like and dot-like contrasts within the straight 10-MR channels (Fig. 2G), which are typically interpreted as adsorbed aromatic molecules in vertical and nearly horizontal configurations, respectively (*19*, *32*). However, the dot-like contrast is likely caused by the projection effect that results from the presence of multiple vertically adsorbed PX molecules with varying in-plane orientations. Whereas previous studies have reported similar observations using iDPC-STEM (*19*, *32*), the ultrahigh spatial resolution of 4D-STEM ptychography allows for more precise determination of the molecular orientations. As shown in Fig. 2G, the PX molecules pointing to the T atom, the O atom, and the middle of the T–O bond were all identified. The diverse orientations of adsorbed molecules imply the presence of complex host-guest interactions in zeolites that are caused by their inhomogeneous chemical environments.

Interestingly, we observed that the 10-MR channels with rod-like contrast displayed a more elongated shape (higher ellipticity) compared with those with dot-like contrast (fig. S7). This finding suggests that ZSM-5 has a flexible framework that can undergo local deformation in response to guest-molecule adsorption. However, we also noted the coexistence of 10-MR channels with markedly different ellipticities in pristine ZSM-5 without any adsorbates (fig. S8). Thus, although the the observation of inhomogeneous channel shape in ZSM-5 is intriguing, a definitive causal relationship between channel shape and molecular orientation has yet to be established.

Zeolites have Brönsted and Lewis acidity, both of which play vital roles in catalysis. The O vacancies generated by framework dehydration during high-temperature calcination is generally considered to be a major source of Lewis acidity, whereas direct observation of O vacancies in zeolites has not been realized to date. The ability of 4D-STEM ptychography to image O atomic columns in the zeolite framework also enabled the identification of O vacancies. To demonstrate this application, we collected 4D-STEM data from a ZSM-5 sample with a Si/Al ratio of ~100 along the [010] axis and reconstructed its phase image using the multislice ptychography method described above. The reconstruction result consists of seven slices, each with a thickness of 4 nm (two unit cells); thus, the imaged area is ~28 nm thick. We precisely identified the T and O atomic columns in each slice using an image-recognition software (see fig. S9), which were subsequently used for intensity analysis.

The intensity distribution of T columns plotted from the experimental ptychographic image was consistent with the simulation based on the defect-free ZSM-5 structure, both of which displayed a 3% coefficient of variation (Fig. 3A). By contrast, the O columns in the experimental ptychographic image exhibited a more pronounced intensity fluctuation than the simulation result (coefficient of variation: 7 versus 5%; Fig. 3B), suggesting the presence of O vacancies in the actual sample.

To explore the effect of O vacancies on the intensity of O columns, we performed image simulations using a series of [010]-projected ZSM-5 structural models, which were 4 nm

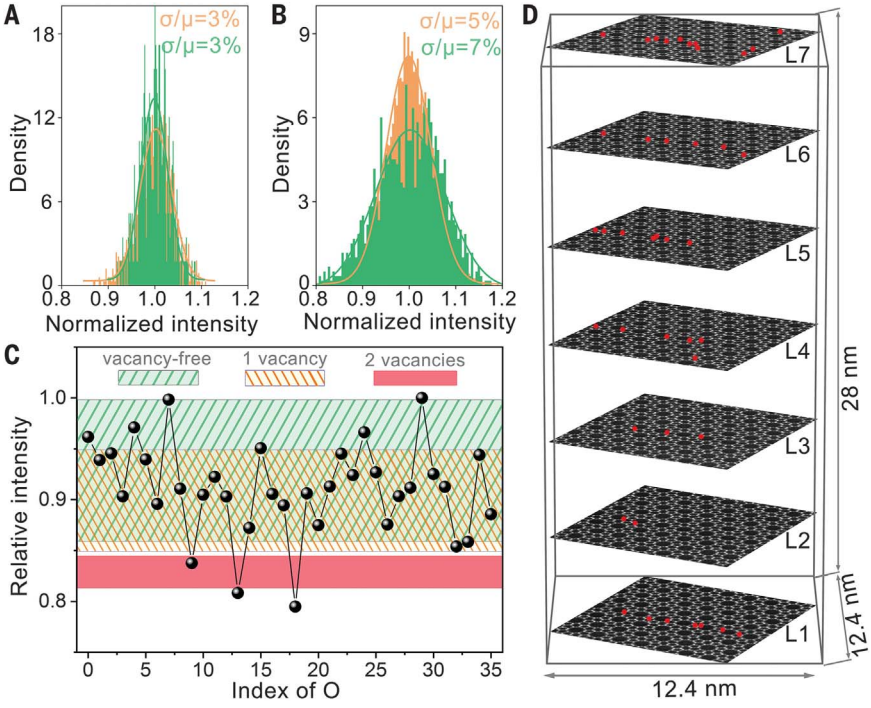


Fig. 3. Identification of O vacancies in ZSM-5. (**A** and **B**) Intensity histograms of T columns (A) and O columns (B) in the experimental (green) and simulated (orange) ptychographic phase images; σ is the standard deviation, and μ is the mean. The intensities are normalized by μ. The statistics are based on more than 1000 atoms. (**C**) Intensity variation of 36 O columns in the experimental ptychographic images. The image with the atomic index is presented in fig. S10. The intensity of each column is normalized by the maximum intensity in the 1-nm range. The light green region represents the intensity range of the O columns in the simulated ptychographic image using a vacancy-free model. The light yellow and red regions mark the intensity ranges of O columns that contain one and two O vacancies within a thickness of 4 nm, respectively. The model used for simulation is presented in fig. S11, and the intensity data are listed in table S1. In the experimental images, the columns with an intensity less than than the lower bound of the red region are identified as vacancy-containing columns. (**D**) Distribution of O vacancies identified by this method throughout the investigated region. The red dots represent identified columns containing O vacancies. L1 to L7 refer to layer 1 to layer 7.

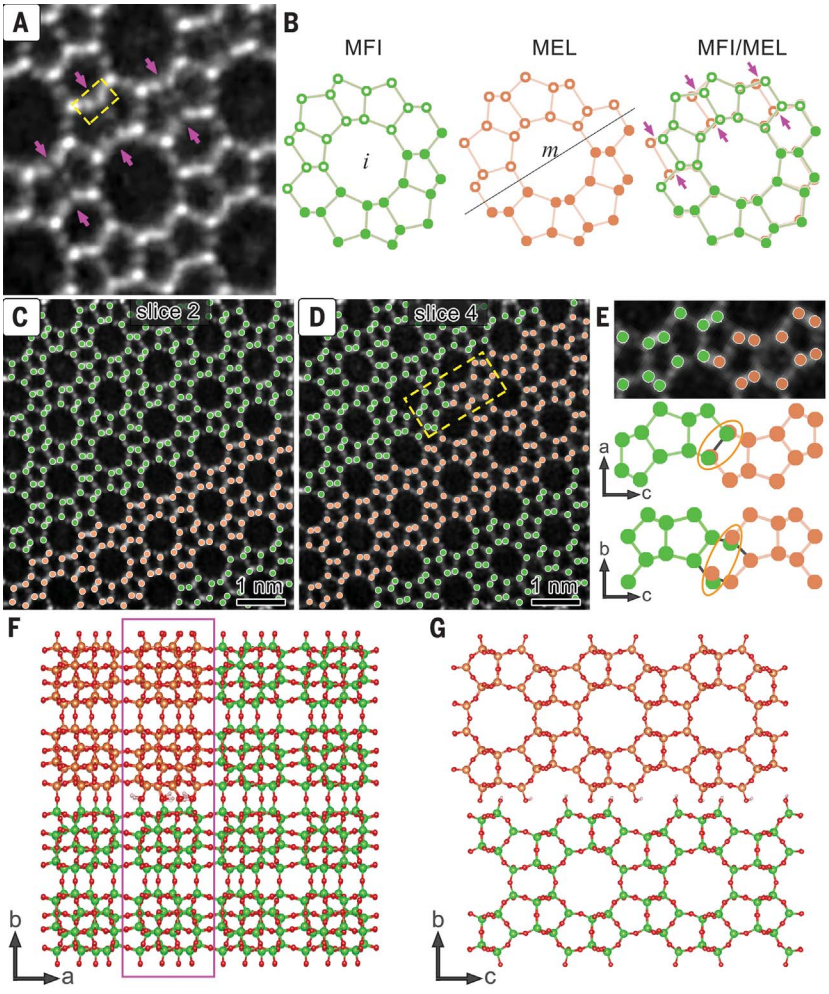


Fig. 4. Intergrowth of MFI and MEL phases in three dimensions. (**A**) 4D-STEM ptychography image displaying extra T atoms (indicated by the magenta arrows) relative to the typical MFI framework. The yellow dashed rectangle indicates the region analyzed in Fig. 5A. (**B**) Structural projections of straight 10-MR channels of MFI and MEL and their superposition. The letters *i* and *m* denote inversion and mirror symmetry, respectively. (**C** and **D**) Second (C) and fourth (D) slices of the multislice ptychography reconstruction result. Raw images are provided in fig. S16. The entire reconstructed volume is presented in movie S1 to display slice-by-slice structural changes. (**E**) At the top is an enlarged image from the region in (D) outlined by the yellow dashed rectangle, which shows the interface of MFI and MEL within the pentasil chain. The middle and bottom panels show a structural model of the interface that is viewed along different orientations, in which T atoms of MFI and MEL are all kept at the interface to demonstrate minimal mismatch. The ellipses highlight the columns at the interface. (**F**) Structural model of MFI-MEL intergrowth projected along the *c* axis. (**G**) The region enclosed by the magenta rectangle in (F) projected along the *a* axis, displaying the disconnected silanol-terminated interface perpendicular to the *b* axis. In (F) and (G), red and pink spheres represent O and H atoms, respectively. Green and orange spheres represent T atoms in MFI and MEL, respectively.

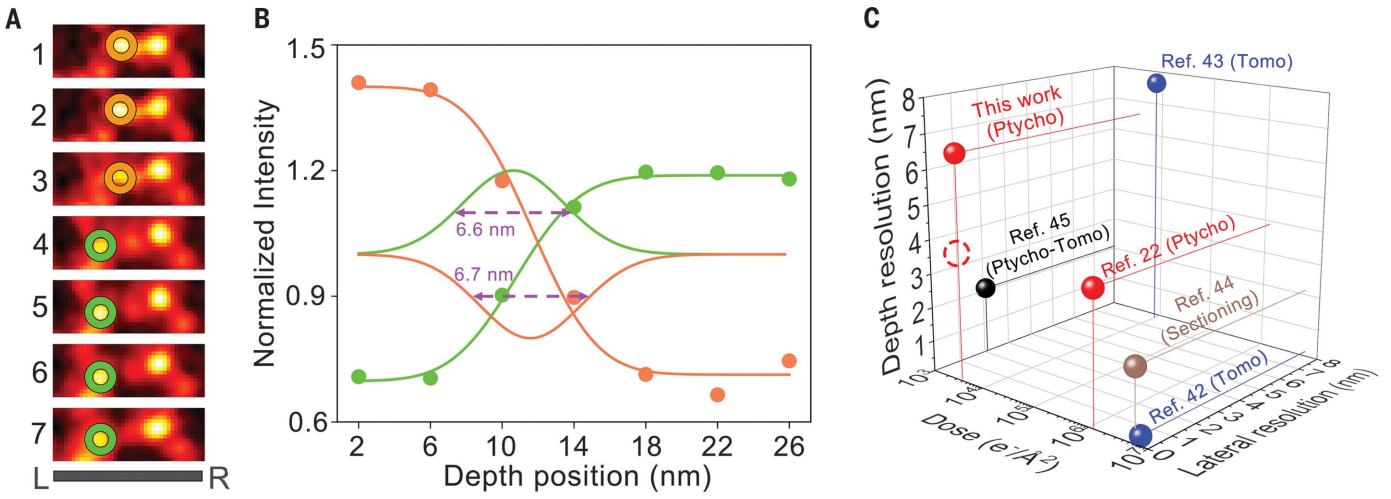


Fig. 5. Depth resolution achieved by 4D-STEM ptychography imaging of zeolites. (**A**) Slice-by-slice analysis of the region highlighted by the yellow dashed rectangle in Fig. 4A. Each slice is 4 nm thick. The green and orange circles mark the T columns in MFI and MEL, respectively. (**B**) Depth resolution estimation based on intensity profiles. Normalized intensities at the two locations indicated by the green and orange circles in (A) were extracted from seven successive slices and plotted as a function of their depth positions. The intensities were normalized by their mean values. The intensity variations at the two locations, represented by green and orange dots, are fitted with the error function. The full width at half maximum of the derivative (dashed lines) of the fitted error function is defined as the depth resolution. (**C**) Lateral and depth resolution reported for various 3D electron microscopy imaging methods. Tomo, Sectioning, Ptycho, and Ptycho-Tomo denote tomography, high-angle annular dark-field STEM depth sectioning, ptychography, and ptychographic tomography, respectively. The red dashed circle indicates the depth resolution of the ptychographic phase image that was reconstructed from a simulated 4D-STEM dataset.

thick and contained eight O sites in each O column, with varying degrees of O vacancy. The results indicated that one O vacancy within a thickness of 4 nm could not be reliably determined because the intensity reduction it caused was within the intrinsic intensity fluctuation of ZSM-5. When two or more aligned O vacancies were within a thickness of 4 nm, the resulting intensity reduction was sufficient for a reliable assignment (Fig. 3C).

We analyzed the ptychographic image of the ~28-nm-thick ZSM-5 slice by slice to render a 3D distribution of O vacancies within the crystal (Fig. 3D and fig. S12). Based on the simulation results (Fig. 3C), we established a relative intensity threshold of 81% to identify O vacancies. The result revealed that the imaged region of the zeolite contained a O vacancy concentration of ~51 μmol/g, with 58% of these O vacancies exposed to the straight 10-MR channels (table S2). It should be noted that the concentration of O vacancies obtained using this method needs to be treated with caution and preferably confirmed by other characterization techniques because the result is sensitive to the choice of intensity threshold. For instance, if a relative intensity threshold of 80% is chosen, the concentration of O vacancies is ~24 μmol/g (fig. S13). Probe-assisted nuclear magnetic resonance (*33*–*35*) and infrared spectroscopy of adsorbed pyridine (Py-IR) (*36*, *37*) are commonly used techniques for identifying and quantifying Lewis acid sites in zeolites that are closely related to O vacancies. The Py-IR method was used to determine the concentration of Lewis acid sites in the investigated ZSM-5 sample. The resulting value was ~35 μmol/g (fig. S14), which is in reasonable agreement with the concentration of O vacancies determined by 4D-STEM ptychography.

Intergrowth of MFI and MEL zeolites

Diffraction-based characterizations have revealed that ZSM-5 and ZSM-11 often coexist in synthetic products because they are similar in topology (MFI and MEL, respectively) (*38*, *39*). However, it is challenging to determine the exact coexistence state, such as whether they are mixed or intergrown, and how the minor phase is distributed relative to the major phase. The only direct evidence for the intergrowth of MFI and MEL by electron microscopy was obtained from a 2D single–unit cell–thick zeolite membrane, in which the two phases were interconnected along the *a* axis of MFI, forming an interface parallel to its *c* axis (*40*). However, conventional electron microscopy imaging has insufficient resolving power in the depth direction to explore the 3D intergrowth of zeolite crystals.

We demonstrate that electron ptychography provides a solution to this problem through the use of a zeolite material containing a primary phase, MFI, and concomitant phase, MEL, with an overall Si/Al ratio of ~90 (see materials and methods for the synthesis method). A comparison between experimental powder x-ray diffraction (PXRD) and PXRD simulation based on random-stacking disorder (*41*) indicated that MEL constitutes 10 to 20% in the material (fig. S15). The 4D-STEM ptychographic image of this material revealed subtle deviations from the typical structural projection along the MFI *b* axis, where there appeared to be some extra T atoms in the pentasil chain (Fig. 4A). The identification of extra T atoms indicated the presence of another structure besides MFI in the projection direction. Based on the PXRD results, we speculated that the imaged zeolite crystal contained MFI and MEL structures along the MFI *b* axis. The two structures of the pentasil family differ only in how the pentasil layers are connected (inversion symmetry for MFI versus mirror symmetry for MEL), and their stacking along the *b* axis resulted in a perfect match of the structure projection to the ptychographic image (Fig. 4B).

Because the reconstructed ptychographic image consisted of seven 4-nm-thick slices, a careful slice-by-slice analysis allowed us to confirm the intergrowth of MFI and MEL along the *b* axis while probing their interfaces in the *a*-*c* plane at atomic resolution (movie S1). In the first two slices, a narrow MEL domain consisting of only two pentasil chains was observed, sandwiched by MFI domains (Fig. 4C). In the fourth slice, the MEL domain expanded, resulting in the generation of a new MFI-MEL interface within the pentasil chain (Fig. 4D). These observations confirmed that MEL and MFI structures were intertwined in three directions within the crystal. Figure 4, C and D, displays how MFI and MEL

structures were interconnected in the *a*-*c* plane. Along the *a* axis, the two structures were perfectly interconnected by sharing a pentasil chain without generating strain or defects, which was consistent with the previous observation (*40*).

The connection of the two structures along the *c* axis (that is, within the pentasil chain) could adopt three possible configurations (fig. S17), of which the one with the least interfacial mismatch was observed (Fig. 4E). Although the MFI-MEL interface perpendicular to the *b* axis could not be directly observed from ptychography given the limited depth resolution, it could be deduced based on the determined orientation relationship. When viewed along the *c* axis, the interface displayed alternating "connected" and "disconnected" regions along the *a* axis, each with a width of one-half the unit cell parameter *a* (Fig. 4F and fig. S18). The disconnected interface could not accommodate additional TO_4 tetrahedra and should be terminated by dangling silanol groups (Fig. 4G).

Depth-resolution analysis

The depth resolution of 4D-STEM ptychography, which is affected by imaging conditions and reconstruction parameters, must be determined explicitly for a given system. Using the 4D-STEM data acquired in this study and the LSQ-ML multislice algorithm, the minimum slice thickness for stable reconstruction was 4 nm, which defined the upper limit of depth resolution. The distinct structures of the two slices separated by 8 nm demonstrate a depth resolution better than 8 nm. To more precisely determine the depth resolution, we analyzed the intensity variation of two T atomic columns in seven consecutive slices spanning from one side of the MEL-MFI interface to the other, where the two studied columns were from MEL and MFI (Fig. 5A). In the interface region (i.e., slices 3 and 4), the two columns interfered in intensity because of the limited depth resolution (Fig. 5A). The two intensity profiles were fitted with error functions and their derivative curves were then derived (Fig. 5B). The full width at half maximum of the derivative curve, which is generally defined as the depth resolution, was measured as ~6.6 nm (Fig. 5B). By using a 4D-STEM dataset simulated at a total electron dose of ~3500 $e^-/Å^2$, a better depth resolution of ~3.5 nm could be achieved based on the same determination method because of the absence of experimental errors (fig. S19).

We compared the results of 4D-STEM ptychography in this study with previous results obtained using various 3D imaging methods in terms of lateral resolution, depth resolution, and consumed electron dose (Fig. 5C). Electron tomography can provide atomic resolution in three dimensions (*42*, *43*). Optical sectioning with high-angle annular dark-field STEM can generate subnanometer depth resolution with dense focal series (*44*). However, these two methods require very high electron doses, typically millions of $e^-/Å^2$, and are unsuitable for beam-sensitive materials (Fig. 5C and table S3). Ptychographic tomography exhibited ~2-nm resolution in three dimensions at low-dose conditions (Fig. 5C) (*45*).

Discussion

This study demonstrated that 4D-STEM multislice ptychography represents an efficient low-dose 3D imaging technique that exhibits excellent tolerance to the specimen thickness while not requiring precise focusing. These advantages make 4D-STEM multislice ptychography broadly applicable and particularly useful for imaging beam-sensitive materials. The lateral and depth resolutions achieved in this study are ~0.85 Å and ~6.6 nm, respectively, using a total electron dose of ~3500 $e^-/Å^2$. This performance has enabled an investigation of the intrinsic structural and compositional inhomogeneity in zeolites, including guest-molecule orientations, O vacancies, and multiphase intergrowth. The simulations indicate that under the imaging conditions that were used, the depth resolution can, in principle, be improved to ~3.5 nm. We believe that combining multislice ptychography with tomography is worth exploring because of its potential to further enhance 3D resolution.

REFERENCES AND NOTES

1. Y. Zhou *et al*., *Science* **373**, 315–320 (2021).
2. Y. Li, J. Yu, *Nat. Rev. Mater.* **6**, 1156–1174 (2021).
3. M. Y. Jeon *et al*., *Nature* **543**, 690–694 (2017).
4. K. Varoon *et al*., *Science* **334**, 72–75 (2011).
5. D. E. Perea *et al*., *Nat. Commun.* **6**, 7589 (2015).
6. S. M. Maier, A. Jentys, J. A. Lercher, *J. Phys. Chem. C* **115**, 8005–8013 (2011).
7. M. Ravi, V. L. Sushkevich, J. A. van Bokhoven, *Nat. Mater.* **19**, 1047–1056 (2020).
8. X. Liu *et al*., *Angew. Chem. Int. Ed.* **60**, 24227–24233 (2021).
9. Q. Yu, X. Tang, H. Yi, *Chem. Eng. J.* **314**, 212–222 (2017).
10. T. Willhammar *et al*., *Nat. Chem.* **4**, 188–194 (2012).
11. V. L. Zholobenko, M. A. Makarova, J. Dwyer, *J. Phys. Chem.* **97**, 5962–5964 (1993).
12. H.-M. Kao, C.-Y. Yu, M.-C. Yeh, *Microporous Mesoporous Mater.* **53**, 1–12 (2002).
13. W. Yuan *et al*., *Science* **371**, 517–521 (2021).
14. L. Liu *et al*., *Nat. Catal.* **3**, 628–638 (2020).
15. X. Li *et al*., *J. Am. Chem. Soc.* **141**, 12021–12028 (2019).
16. D. Zhang *et al*., *Science* **359**, 675–679 (2018).
17. K. C. Bustillo *et al*., *Acc. Chem. Res.* **54**, 2543–2551 (2021).
18. L. Liu *et al*., *Angew. Chem. Int. Ed.* **59**, 819–825 (2020).
19. H. Xiong *et al*., *Science* **376**, 491–496 (2022).
20. B. Shen *et al*., *Nature* **592**, 541–544 (2021).
21. Y. Jiang *et al*., *Nature* **559**, 343–349 (2018).
22. Z. Chen *et al*., *Science* **372**, 826–831 (2021).
23. G. Li, H. Zhang, Y. Han, *ACS Cent. Sci.* **8**, 1579–1588 (2022).
24. C. Ophus, *Microsc. Microanal.* **25**, 563–582 (2019).
25. T. J. Pennycook *et al*., *Ultramicroscopy* **151**, 160–167 (2015).
26. C. M. O'Leary *et al*., *Appl. Phys. Lett.* **116**, 124101 (2020).
27. M. Odstrčil, A. Menzel, M. Guizar-Sicairos, *Opt. Express* **26**, 3108–3123 (2018).
28. K. Wakonig *et al*., *J. Appl. Crystallogr.* **53**, 574–586 (2020).
29. A. M. Maiden, M. J. Humphry, J. M. Rodenburg, *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **29**, 1606–1614 (2012).
30. S. Cao, P. Kok, P. Li, A. M. Maiden, J. M. Rodenburg, *Phys. Rev. A* **94**, 063621 (2016).
31. B. Shen *et al*., *Adv. Mater.* **32**, e1906103 (2020).
32. B. Shen *et al*., *Nature* **607**, 703–707 (2022).
33. J. H. Lunsford, W. P. Rothwell, W. Shen, *J. Am. Chem. Soc.* **107**, 1540–1547 (1985).
34. A. Zheng, S.-B. Liu, F. Deng, *Chem. Rev.* **117**, 12475–12531 (2017).
35. X. Yi *et al*., *J. Am. Chem. Soc.* **140**, 10764–10774 (2018).
36. M. Guisnet, P. Ayrault, J. Datka, *Pol. J. Chem.* **71**, 1455–1461 (1997).
37. J. Zhang *et al*., *ACS Catal.* **13**, 3794–3805 (2023).
38. G. R. Millward, S. Ramdas, J. M. Thomas, M. T. Barlow, *J. Chem. Soc. Faraday Trans. II* **79**, 1075–1082 (1983).
39. P. M. Piccione, M. E. Davis, *Microporous Mesoporous Mater.* **49**, 163–169 (2001).
40. P. Kumar *et al*., *Nat. Mater.* **19**, 443–449 (2020).
41. M. M. J. Treacy, J. M. Newsam, M. W. Deem, *Proc. R. Soc. London Ser. A* **433**, 499–520 (1991).
42. M. C. Scott *et al*., *Nature* **483**, 444–447 (2012).
43. D. Chen, H. Friedrich, G. de With, *J. Phys. Chem. C* **118**, 1248–1257 (2014).
44. R. Ishikawa, N. Shibata, T. Taniguchi, Y. Ikuhara, *Phys. Rev. Appl.* **13**, 034064 (2020).
45. Z. Ding *et al*., *Nat. Commun.* **13**, 4787 (2022).
46. H. Zhang *et al*., Data Set: Three-dimensional structural and compositional inhomogeneity in zeolites unraveled by low-dose electron ptychography. KAUST Repository (2022); https://repository.kaust.edu.sa/handle/10754/690886.

ACKNOWLEDGMENTS

We acknowledge the KAUST Supercomputing Laboratory for providing the computing resources, J. H. Xin for assistance in setting up the computing environment, and F. Lin and P. Pelz for discussions on ptychography. **Funding:** This work was funded by the National Key Research and Development Project of China (2022YFE0113800), the KAUST Center Competitive Fund (FCC/1/1972-43-01), Liaoning Revitalization Talents Program (XLYC2008032), and Fundamental Research Funds for the Central Universities (DUT22LAB602). **Author contributions:** Conceptualization: Y.H., H.Z.; Methodology: H.Z., G.L., Z.C., J.Z., X.L., C.C., Y.Z.; Investigation: Y.H., H.Z., G.L., J.Z., L.L.; Visualization: H.Z., G.L., D.Z., Y.Z.; Supervision: Y.H., H.Z., X.G., D.Z., P.G.; Writing – original draft: H.Z.; Writing – review and editing: Y.H., G.L., C.C., P.G., X.G., Z.C. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and the supplementary materials. The original image data are available at the KAUST Repository (*46*). **License information:** Copyright © 2023 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.adg3183
Materials and Methods
Figs. S1 to S19
Tables S1 to S3
References (*47*–*52*)
Movie S1

Submitted 15 December 2022; accepted 5 April 2023
10.1126/science.adg3183

ORGANIC CHEMISTRY

Regio-controllable [2+2] benzannulation with two adjacent C(sp^3)–H bonds

Ji-Min Yang†, Yu-Kun Lin†, Tao Sheng, Liang Hu, Xin-Pei Cai, Jin-Quan Yu*

Regiocontrol in traditional cycloaddition reactions between unsaturated carbon compounds is often challenging. The increasing focus in modern medicinal chemistry on benzocyclobutene (BCB) scaffolds indicates the need for alternative, more selective routes to diverse rigid carbocycles rich in C(sp^3) character. Here, we report a palladium-catalyzed double C–H activation of two adjacent methylene units in carboxylic acids, enabled by bidentate amide-pyridone ligands, to achieve a regio-controllable synthesis of BCBs through a formal [2+2] cycloaddition involving σ bonds only (two C–H bonds and two aryl–halogen bonds). A wide range of cyclic and acyclic aliphatic acids, as well as dihaloheteroarenes, are compatible, generating diversely functionalized BCBs and hetero-BCBs present in drug molecules and bioactive natural products.

Benzocyclobutenes (BCBs) represent an important class of rigid four-membered carbocycles found in natural products (*1*, *2*) that have demonstrated great potential as therapeutic molecular scaffolds (*3*, *4*), versatile synthons (*5*), and functional motifs in materials science and mechanochemistry (*6*, *7*) (Fig. 1A). BCB-based rigid and three-dimensional pharmacophores have led to the discovery of Ivabradine, an FDA (Food and Drug Administration)–approved drug for the treatment of heart failure and heart-related chest pain (*3*). In addition, the BCB analog of the psychoactive 2C-B (4-bromo-2,5-dimethoxyphenethylamine) was found to show superior affinity for the human 5-hydroxytryptamine 2A (5-HT_{2A}) receptor compared with the conformationally flexible parent compound and the benzocyclopentane analog (Fig. 1A) (*8*). Currently, the [2+2] cycloaddition of alkenes and benzynes is one of the most common synthetic routes to BCBs (*9*). However, controlling the regioselectivity of this cycloaddition reaction is an unsolved problem. As illustrated in Fig. 1B, the [2+2] cycloadditions between a substituted alkene and benzyne typically lead to mixt… of head-to-head and head-to-tail regioisom… as well as multiple diastereoisomers. A number of alternative intramolecular processes, such as cyclization of *o*-quinodimethane (*10*), cyclative C–H arylations (*11*–*14*), visible-light–induced radical processes (*15*), and others (*16*), have been developed. A modular two-component intermolecular synthesis of BCBs has recently been reported using an elegant Pd-catalyzed [2+2] annulation of alkenes bearing an amide directing group and arylboronic acids (*17*); however, this method does not allow regiocontrol. In addition, existing methods for BCB synthesis suffer from scope and efficiency limitations owing to the requirement of preinstallation of reactive functional groups such as double or triple bonds. Notably, the synthesis of medicinally important heterocyclic BCBs using these reactions is challenging (*18*, *19*). Our recently demonstrated Pd-catalyzed activation of β- or γ-C–H bonds of aliphatic acids (*20*–*22*) inspired us to explore a formal [2+2] annulation of aliphatic acids with dihaloarenes through a dual sequential methylene C–H activation at the β and γ positions. The state of the art of dual functionalization of two adjacent

Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA.
*Corresponding author. Email: yu200@scripps.edu
†These authors contributed equally to this work.

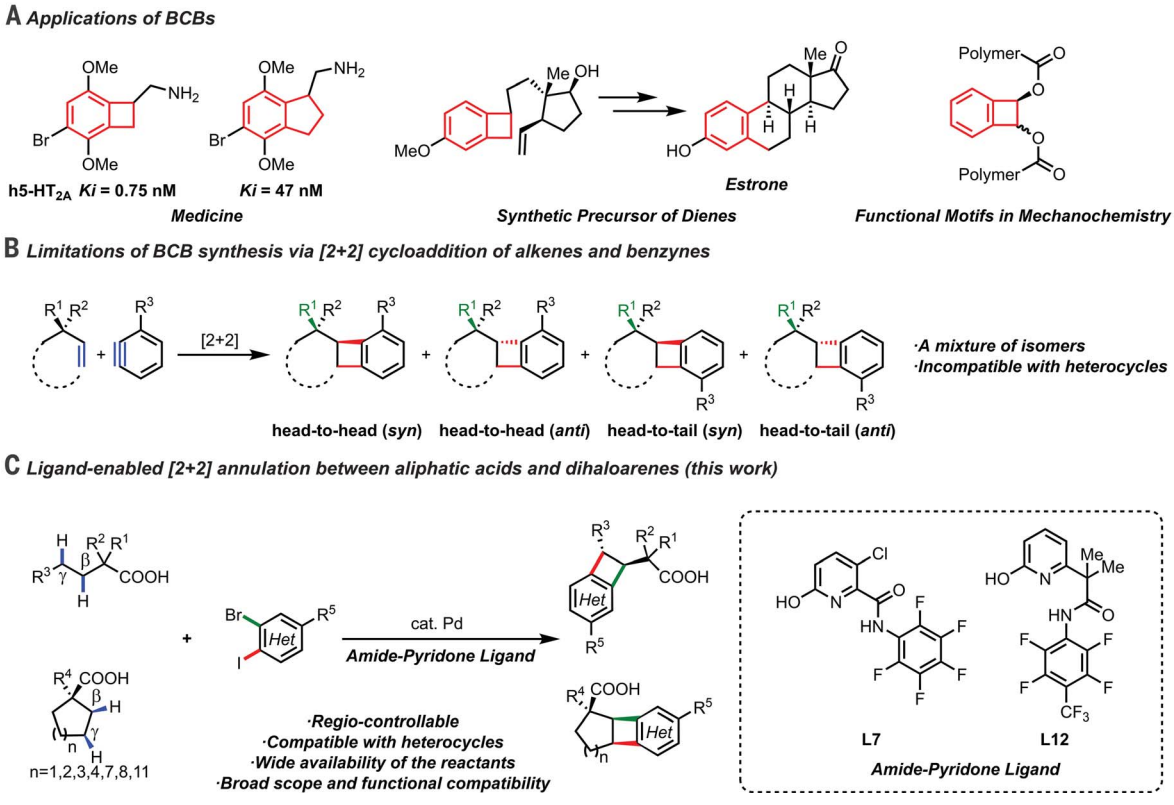


Fig. 1. [2+2] annulation reaction via dual methylene C–H bond functionalization. (**A**) Applications of BCBs. K_i is the inhibitor constant. (**B**) Limitations of BCB synthesis via [2+2] cycloaddition of alkenes and benzynes. *Syn*, *anti* refers to the conformation of R^1 group and the adjacent aryl group. (**C**) Ligand-enabled [2+2] annulation between aliphatic acids and dihaloarenes (this work).

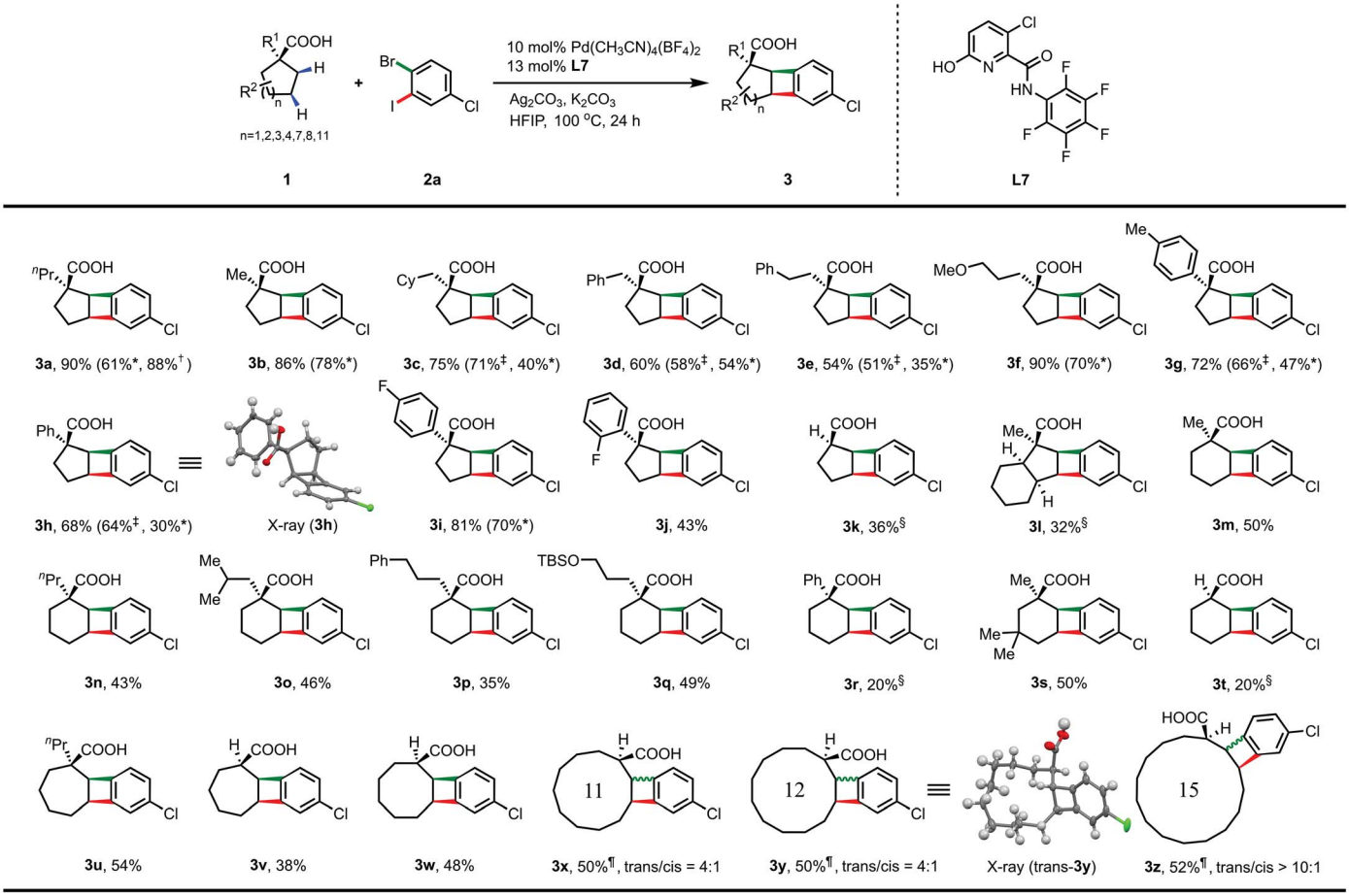


Fig. 2. Cyclic aliphatic acid scope for the [2+2] annulation reaction. Reaction conditions: aliphatic acid (0.1 mmol), **2a** (0.2 mmol), $Pd(CH_3CN)_4(BF_4)_2$ (10 mol %), **L7** (13 mol %), K_2CO_3 (2.5 equiv), Ag_2CO_3 (2.0 equiv), HFIP (1.0 ml), at 100°C for 20 hours. Isolated yields are reported. Relative configurations are shown, as the synthesis is racemic. See supplementary materials for details. *$Pd(CH_3CN)_4(BF_4)_2$ (1 mol %) was used, reaction time 56 hours. †Reaction conducted with aliphatic acid **1a** (4.5 mmol), 1.05 g of **3a** was obtained. ‡ $Pd(CH_3CN)_4(BF_4)_2$ (5 mol %) was used, reaction time 40 hours. §Aliphatic acid (0.2 mmol), **2a** (0.1 mmol), K_2CO_3 (3.5 equiv). ¶**L28** instead of **L7**, $KHCO_3$ instead of K_2CO_3.

methylene C–H bonds is limited to allylic and benzylic positions, or sites α- to a heteroatom (*23*–*26*). Here we report successful Pd-catalyzed BCB synthesis through the annulation of aliphatic acids with dihaloarenes enabled by amide-pyridone ligands, in which the exclusive regiocontrol is achieved through the differentiation between the aryl iodide and bromide sites. The direct use of abundant and structurally varied acyclic and cyclic acids as substrates without prefunctionalization substantially expands access to diverse BCBs, including heterocyclic BCB scaffolds (Fig. 1C).

Reaction development

During our mechanistic investigations of the methylene C–H lactonization of dicarboxylic acids, we observed the β,γ-C–H deuteration of 7-ethoxy-7-oxoheptanoic acid (*20*). This observation prompted us to wonder whether a sequential twofold C–H coupling process with ortho-dihaloarenes could be developed to construct BCB scaffolds. Using 1-propyl-1-cyclopentanecarboxylic acid **1a** and dihaloarene coupling partner 1-bromo-4-chloro-2-iodobenzene **2a** with Ag_2CO_3 as a scavenger for halides, we began to test the feasibility of this transformation with a series of pyridine-pyridone ligands recently developed in our laboratory (table S1). Encouragingly, the desired BCB **3a** was indeed formed as a single isomer (fig. S1). However, attempts to improve the reaction yield with pyridine-pyridone ligands and other known ligand scaffolds (**L27** to **L36**) under various conditions proved unsuccessful (≤22% yield).

These initial results with pyridine-pyridone ligands, though promising, pointed to the need for more effective ligands. Because the 2-pyridone serves a critical role as the internal base for C–H cleavage (*27*), we focused on developing alternatives for the pyridine arm of the scaffold. Inspired by the superior reactivity of the electron-deficient amide motif as a directing group for Pd(II) in $C(sp^3)$–H activation (*28*), we synthesized a class of bidentate ligands bearing both a pyridone and an electron-deficient amide derived from perfluorinated anilines. Amide-pyridone ligand **L1**, posited to coordinate to Pd(II) as a five-membered chelate, afforded substantially improved yield in the annulation reaction (38%). Through extensive structural tuning, we identified ligand **L7** as optimal, providing BCB **3a** in 90% isolated yield.

Cyclic acid substrate scope

Under these optimized conditions, we subsequently evaluated the substrate scope of the [2+2] annulation reaction (Fig. 2). A wide range of cyclic aliphatic acids including five- (**1a-1l**), six- (**1m-1t**), seven- (**1u-1v**), and eight-membered (**1w**) rings were compatible to afford *cis*-BCB products as single regioisomers (**3a-3w**). Larger cyclic aliphatic acids including 11- (**1x**), 12- (**1y**), and 15-membered (**1z**) rings produced *trans*-BCB products as major isomers. Remaining unreacted aliphatic acid reactants were recovered when the yields were relatively

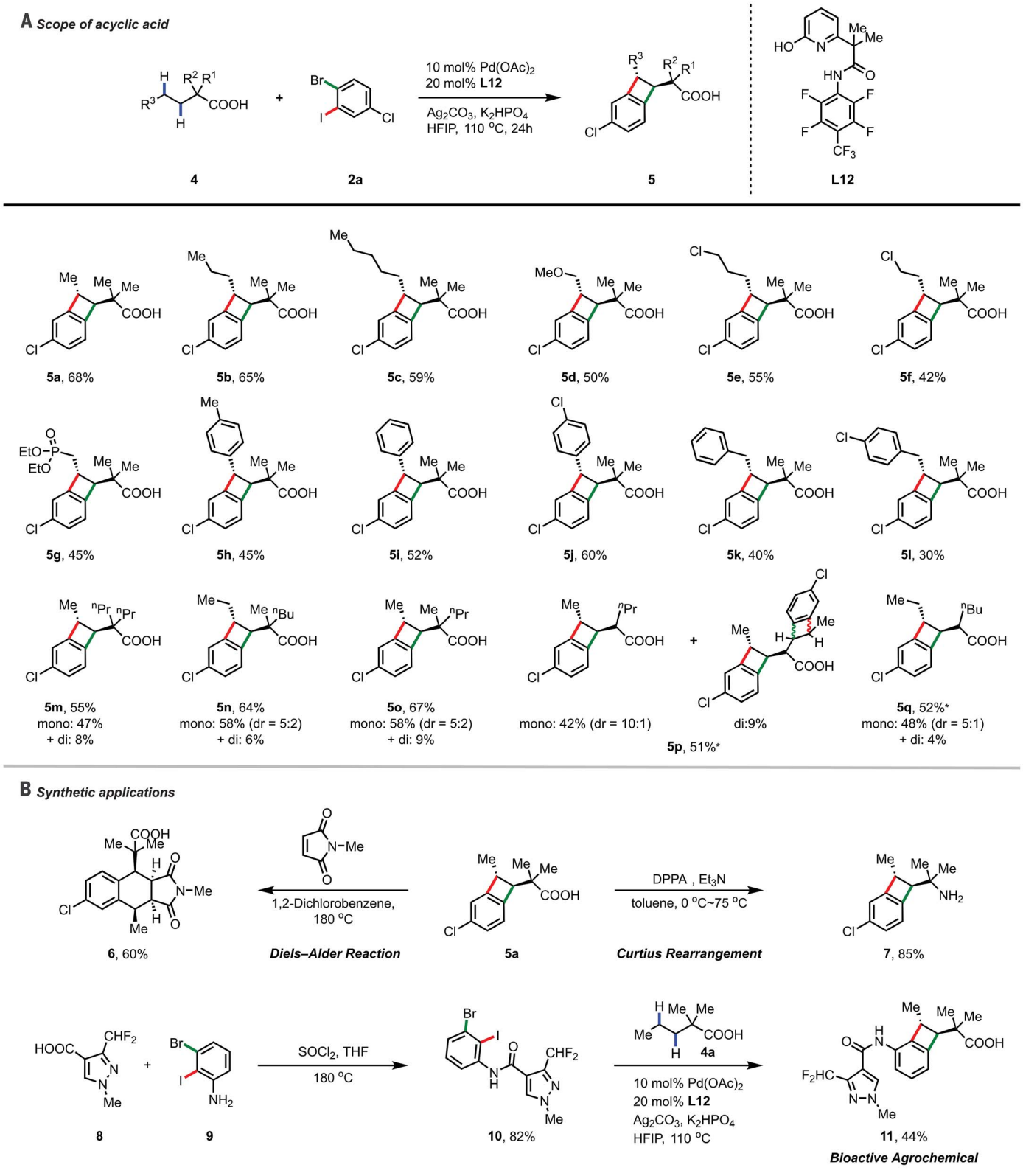


Fig. 3. Acyclic aliphatic acid scope for the [2+2] annulation reaction and transformations of the products. (**A**) Acyclic aliphatic acid scope. Reaction conditions: aliphatic acid (0.1 mmol), **2a** (0.3 mmol), $Pd(OAc)_2$ (10 mol %), **L12** (20 mol %), K_2HPO_4 (3.5 equiv), Ag_2CO_3 (2.0 equiv), HFIP (1.2 ml), at 110°C for 24 hours. Isolated yields are reported. Relative configurations are shown, as the synthesis is racemic. Di-**5p** was isolated as a representative of di-BCB products. *__2a__ (0.35 mmol), K_2HPO_4 (3.0 equiv), HFIP (1.8 ml) as a replacement. (**B**) Synthetic applications. DPPA, diphenylphosphoryl azide.

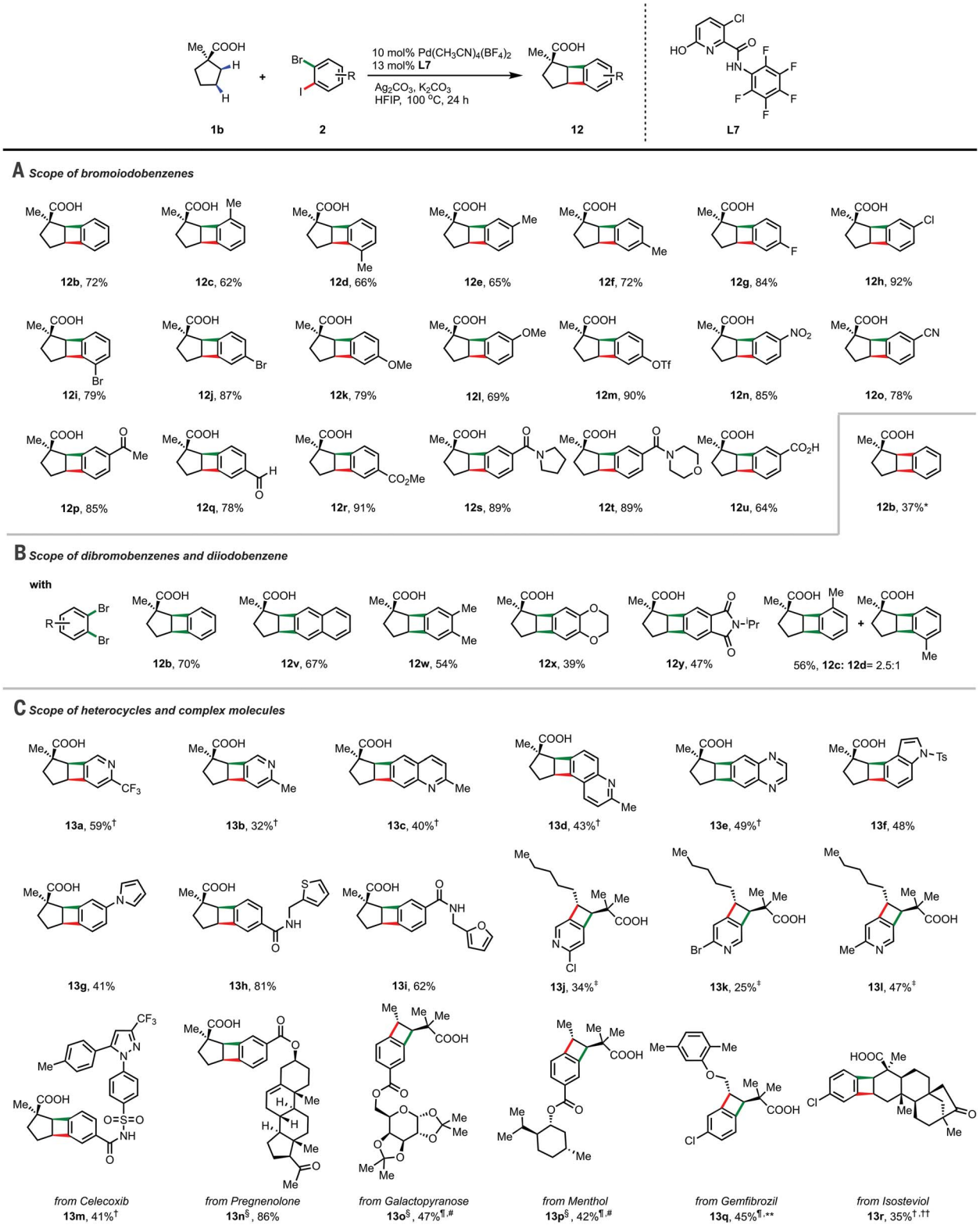


Fig. 4. Dihaloarene scope for the [2+2] annulation reaction. (**A**) Scope of bromoiodobenzenes. (**B**) Scope of dibromobenzenes and diiodobenzene. (**C**) Scope of heterocycles and complex molecules. Reaction conditions: aliphatic acid (0.1 mmol), dihaloarene (0.2 mmol), $Pd(CH_3CN)_4(BF_4)_2$ (10 mol %), **L7** (13 mol %), K_2CO_3 (2.5 equiv), Ag_2CO_3 (2.0 equiv), HFIP (1.0 ml), at 100°C for 20 hours. Isolated yields are reported. Relative configurations are shown, as the synthesis is racemic. *1,2-diiodobenzene (**2b^{I}**) as substrate. †Aliphatic acid (0.2 mmol), dihaloarene (0.1 mmol), K_2CO_3 (3.5 equiv). ‡**L28** instead of **L7**, **4c** instead of **1b**, $KHCO_3$ instead of K_2CO_3. §Diastereoisomers. ¶Dihaloarene (0.3 mmol), $Pd(OAc)_2$ (10 mol %), **L12** (20 mol %), K_2HPO_4 (3.5 equiv), Ag_2CO_3 (2.0 equiv), HFIP (1.2 ml), at 110°C for 24 hours. #**4a** instead of **1b**. **Gemfibrozil instead of **1b**. ††Isosteviol instead of **1b**.

low. Notably, ligand **L7** allowed us to reduce Pd loading from 10 to 5 or 1% while maintaining synthetically useful yields for the [2+2] annulation between dihaloarene **2a** and acids (**1a-1i**) as selected examples. Various substitutions at the α position of carboxylic acid were tested in the reaction. Alkyl groups (**3a-3f**, **3l-3q**, **3s**, **3u**) and aryl groups (**3g-3j**, **3r**) with electronically diverse substituents were compatible with this protocol despite the presence of reactive methyl and aryl C–H bonds. Functionalities such as methoxy (**3f**), OTBS (**3q**), and phenyl (**3d-3e**, **3p**) were well tolerated. Bicyclic aliphatic acid (**1l**) was converted to the corresponding fused 6-5-4-6 ring (**3l**) in 32% yield. Cyclohexanecarboxylic acid–bearing substitutions at the 3 position (**1s**) consistently provided the corresponding product (**3s**) in moderate yield. In addition to the aliphatic acids containing an α-quaternary center, aliphatic acids containing an α-hydrogen (**1k**, **1t**, **1v-1z**) afforded the products (**3k**, **3t**, **3v-3z**) in 20 to 52% yield. Quinoline-pyridone ligand **L28** was more suitable for 8- to 15-membered rings (**1w-1z**). The structures of **3h** and *trans*-**3y** were confirmed by single-crystal x-ray diffraction analysis. We also investigated the scalability of this [2+2] annulation reaction, in which 1.05 g of the desired product **3a** could be obtained in 88% isolated yield (see supplementary materials for experimental details).

Acyclic acid substrate scope

Acyclic aliphatic acid **4a** only gave low yield with **L7** (2%) (table S6). We found that six-membered chelating amide-pyridone ligand **L12** emerged as the most effective ligand for these linear substrates, affording diverse *trans*-BCBs (Fig. 3). A wide range of functional groups, such as methoxy (**5d**), chloro (**5e-5f**), phosphonate (**5g**), and substituted aryl groups (**5h-5l**) were compatible. Aliphatic acids bearing α-*gem*-dimethyl groups (**4a-4l**) or a single methyl group (**4n-4o**) were also compatible despite the presence of the typically more accessible α-methyl C–H bonds. In addition to the substrates containing an α-quaternary center (**4a-4o**), aliphatic acids containing an α-hydrogen (**4p-4q**) were also found to be viable, as exemplified with products **5p** and **5q**, generated in moderate yields. A number of synthetic elaborations of benzocyclobutyl acid products were also demonstrated. Heating a solution of **5a** as an *o*-quinodimethane precursor with *N*-methylmaleimide in *o*-dichlorobenzene at 180°C provided Diels-Alder adduct **6** in 60% yield. **5a** could be readily transformed to medicinally important amine **7**. The benzocyclobutyl amide **11**, a patented bioactive molecule for crop protection (*29*), could be rapidly synthesized by [2+2] annulation of aliphatic acid **4a** and *N*-dihaloaryl amide **10** in a two-step sequence.

Dihaloarene substrate scope

The broad scope of bromoiodoarenes further expands the diversity of BCBs (**12b-12u**) (Fig. 4A). Substituents at various positions were compatible with this catalytic system (**12c-12f**). Halogens such as fluorine (**12g**), chlorine (**12h**), and bromine (**12i-12j**) were compatible, providing the desired BCBs in good yields. Arenes bearing electron-donating (**12k-12l**) and electron-withdrawing (**12m-12t**) groups were also tested, affording the products in good to excellent yields. The amide (**12s-12t**) and carboxylic acid (**12u**) groups, widely used in C–H activation as directing groups, were also viable for this protocol. An aryl triflate, a highly reactive site in Pd cross-coupling chemistry, was also tolerated (**12m**). Moreover, dibromoarenes ($\mathbf{2b^{Br}}$, **2v-2y**) and diiodoarene ($\mathbf{2b^{I}}$) were amenable to the standard conditions, generating the desired BCBs in moderate to good yields (Fig. 4B). As expected, reaction of **1b** with 2,3-dibromotoluene ($\mathbf{2c^{Br}}$) afforded a mixture of two regioisomers (**12c** and **12d**, regioisomeric ratio rr = 2.5: 1), indicating that bromoiodoarenes with two different active sites are vital for controlling regioselectivity.

Heteroaromatics and complex molecules

Modular synthesis of hetero-BCBs through various methods has been a long-standing challenge (*22*, *23*). Gratifyingly, pyridine (**13a-13b**, **13j-13l**), quinoline (**13c-13d**), quinoxaline (**13e**), and indole (**13f**) derivatives were also competent in this process, affording the desired hetero-BCBs in 25 to 59% yield (Fig. 4C). Heterocycles such as pyrrole (**13g**), thiophene (**13h**), and furan (**13i**) were well tolerated, providing the desired products in good to excellent yields. Dihaloarenes bearing bioactive structures such as celecoxib (**13m**), pregnenolone (**13n**), galactose (**13o**), and menthol (**13p**) were also compatible. This protocol could also be applied in late-stage modification of bioactive molecules, such as gemfibrozil and isosteviol, leading to products **13q** and **13r**.

Mechanistic study

The sequential activation of two C–H and two aryl halide bonds with unexpectedly exclusive regioselectivity calls for mechanistic investigation into this complex catalytic cycle. Reactivity of various possible intermediates (figs. S2 and S3) was consistent with a highly complex Pd(II)/Pd(0)/Pd(II)/Pd(IV) catalytic reaction pathways (fig. S4). β,γ-Dehydrogenation likely occurred via ligand-enabled β- or γ-C–H bond activation accompanied by the formation of Pd(0) species. Oxidative addition of the more reactive aryl iodide with Pd(0) species followed by directed carbopalladation would then afford a key intermediate bearing the aryl on the γ position with all regioselectivity defined. Next, intramolecular oxidative addition of the aryl bromide with the alkyl Pd(II) would lead to a Pd(IV) intermediate, which could undergo reductive elimination to afford the desired BCB product and Pd(II) catalyst. The plausibility of the dehydrogenation pathway is supported by the catalytic formation of the desired BCB product using β,γ-unsaturated aliphatic acid **S5a** as substrate (for various control experiments, see fig. S4, A to G). Notably, the following two elementary steps must be exclusively regioselective to form a single regioisomer of BCB: oxidative addition with Pd(0) of the more reactive aryl iodide bond; and subsequent carbopalladation governed by a directing effect to place the aryl group at the γ position and Pd(II) at the β position.

Beyond the realization of regioselective functionalization of two adjacent methylene C–H bonds, the modular synthesis of diverse BCBs with exclusive regioselectivity is applicable to pharmaceuticals, materials science, and mechanochemistry, as well as natural product synthesis. Further studies on the reaction mechanism are underway.

REFERENCES AND NOTES

1. A. K. Sadana, R. K. Saini, W. E. Billups, *Chem. Rev.* **103**, 1539–1602 (2003).
2. M. S. Elnaggar *et al*., *RSC Advances* **7**, 30640–30649 (2017).
3. M. A. Psotka, J. R. Teerlink, *Circulation* **133**, 2066–2075 (2016).
4. A. Tsotinis, P. A. Afroudakis, P. J. Garratt, A. Bocianowska-Zbrog, D. Sugden, *ChemMedChem* **9**, 2238–2243 (2014).
5. J. L. Segura, N. Martín, *Chem. Rev.* **99**, 3199–3246 (1999).
6. J. Wang *et al*., *Nat. Chem.* **7**, 323–327 (2015).
7. C. R. Hickenboth *et al*., *Nature* **446**, 423–427 (2007).
8. T. H. McLean *et al*., *J. Med. Chem.* **49**, 5794–5803 (2006).
9. S. Kotha, K. Lahiri, Y. Tangella, *Asian J. Org. Chem.* **10**, 3166–3185 (2021).
10. P. Schiess, M. Heitzmann, S. Rutschmann, R. Stäheli, *Tetrahedron Lett.* **19**, 4569–4572 (1978).
11. M. Chaumontet *et al*., *J. Am. Chem. Soc.* **130**, 15157–15166 (2008).
12. J. Ye *et al*., *Nat. Chem.* **9**, 361–368 (2017).
13. P. A. Provencher *et al*., *J. Am. Chem. Soc.* **143**, 20035–20041 (2021).
14. W.-X. Wei *et al*., *J. Am. Chem. Soc.* **143**, 7868–7875 (2021).
15. J. Liu, T. Hao, L. Qian, M. Shi, Y. Wei, *Angew. Chem. Int. Ed.* **61**, e202204515 (2022).
16. F. J. T. Talbot *et al*., *ACS Catal.* **11**, 14448–14455 (2021).
17. T. Fujii, S. Gallarati, C. Corminboeuf, Q. Wang, J. Zhu, *J. Am. Chem. Soc.* **144**, 8920–8926 (2022).
18. J. M. Riemann, W. S. Trahanovsky, *Tetrahedron Lett.* **18**, 1867–1870 (1977).
19. A. Naiman, K. P. C. Vollhardt, *Angew. Chem.* **91**, 440–441 (1979).
20. H. S. S. Chan, J.-M. Yang, J.-Q. Yu, *Science* **376**, 1481–1487 (2022).
21. L. Hu, G. Meng, J.-Q. Yu, *J. Am. Chem. Soc.* **144**, 20550–20553 (2022).
22. Z. Wang *et al*., *Science* **374**, 1281–1285 (2021).
23. E. M. Stang, M. C. White, *J. Am. Chem. Soc.* **133**, 14892–14895 (2011).
24. A. F. Prusinowski, R. K. Twumasi, E. A. Wappes, D. A. Nagib, *J. Am. Chem. Soc.* **142**, 5429–5438 (2020).
25. T. Shen, T. H. Lambert, *Science* **371**, 620–626 (2021).
26. T. Shen, Y.-L. Li, K.-Y. Ye, T. H. Lambert, *Nature* **614**, 275–280 (2023).
27. Z. Li *et al*., *Science* **372**, 1452–1457 (2021).
28. M. Wasa *et al*., *J. Am. Chem. Soc.* **134**, 18570–18572 (2012).
29. C. Dubost *et al*., "Benzocyclobutane(Thio) Carboxamides," U.S. Patent 9878985 B2 (2018).

ACKNOWLEDGMENTS

We thank K. Wu for proofreading and providing helpful suggestions in preparing the manuscript. We thank Z. Li for proofreading supplementary information and repeating reactions. We thank M. E. Hoque for sharing substrates. We also thank the X-ray Crystallography Facility (UCSD) for x-ray crystallography. Y.-K. L. thanks Q. N. Wong for help with purification and high-resolution mass spectrometry.

Funding: We acknowledge The Scripps Research Institute and NIH (NIGMS, 2R01GM084019) for financial support. **Author contributions:** J.-Q.Y. conceived the study. J.-M.Y. discovered and developed the [2+2] benzannulation reaction. Y.-K.L. developed the amide-pyridone ligands, completed the benzannulation of the acyclic acids, and proposed the mechanism. J.-M.Y. and Y.-K.L. conducted preliminary mechanistic studies and the transformations of BCB products. T.S., L.H., and X.-P.C. participated in the substrate scope survey. J.-M.Y., Y.-K.L., and J.-Q.Y. wrote the manuscript. J.-Q.Y. directed the project. **Competing interests:** J.-Q.Y., J.-M.Y., and Y.-K.L. are inventors on a patent application related to this work (US Patent application 63/487,381) filed by The Scripps Research Institute. The authors declare no other competing interests. **Data and materials availability:** All data are in the supplementary materials. Crystallographic data for compounds **3h** and trans-**3y** are available from the Cambridge Crystallographic Data Center under reference numbers CCDC 2220709 and 2246693. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg5282
Materials and Methods
Figs. S1 to S4
Tables S1 to S8
References (*30*–*38*)
X-ray Crystallographic Data
NMR Spectra

Submitted 3 January 2023; accepted 3 April 2023
10.1126/science.adg5282

CATALYSIS

Restructuring of titanium oxide overlayers over nickel nanoparticles during catalysis

Matteo Monai[1]†, Kellie Jenkinson[2]†, Angela E. M. Melcherts[1]†, Jaap N. Louwen[1], Ece A. Irmak[2], Sandra Van Aert[2], Thomas Altantzis[3], Charlotte Vogt[1]‡, Ward van der Stam[1], Tomáš Duchoň[4], Břetislav Šmíd[5], Esther Groeneveld[6], Peter Berben[6], Sara Bals[2]*, Bert M. Weckhuysen[1]*

Reducible supports can affect the performance of metal catalysts by the formation of suboxide overlayers upon reduction, a process referred to as the strong metal–support interaction (SMSI). A combination of operando electron microscopy and vibrational spectroscopy revealed that thin TiO_x overlayers formed on nickel/titanium dioxide catalysts during 400°C reduction were completely removed under carbon dioxide hydrogenation conditions. Conversely, after 600°C reduction, exposure to carbon dioxide hydrogenation reaction conditions led to only partial reexposure of nickel, forming interfacial sites in contact with TiO_x and favoring carbon–carbon coupling by providing a carbon species reservoir. Our findings challenge the conventional understanding of SMSIs and call for more-detailed operando investigations of nanocatalysts at the single-particle level to revisit static models of structure-activity relationships.

Supported metal catalysts respond to pretreatment and reaction conditions by restructuring (*1*), phase changes (*2*), and chemical and structural oscillations (*3*), which can lead to catalyst activation, deactivation, or changes in selectivity. A good example of catalyst restructuring is the so-called strong metal–support interaction (SMSI), in which metal nanoparticles (NPs) are encapsulated by overlayers of reducible supports, such as TiO_2, after high-temperature reduction (*4*) or oxidation (*5*). SMSI can result in a substantial change in product selectivity (*6*), and it has been shown to be correlated with high activity of mixed metal oxides in the oxygen evolution reaction (*7*) and with activation of industrial $Cu/ZnO_x/Al_2O_3$ catalysts (*8*).

Although there is substantial evidence for overlayer formation during catalyst pretreatments (*9*–*11*), including from atomic-resolution in situ electron microscopy studies of overlayer formation under H_2 (*10*, *12*–*14*), O_2 (*5*), and CO_2 environments (*14*), very little is known about the structure of the overlayer under reaction conditions. Recently, the complete removal of a TiO_x ($x \leq 2$) overlayer was shown for Pt/TiO_2 using in situ transmission electron microscopy (TEM) under O_2/H_2 mixtures at 600°C (*12*). This result raises the question of whether overlayers persist during reaction conditions or if the pretreatments used to induce SMSI only indirectly affect the catalyst performance, for example, by particle growth or phase changes, given that the geometric and electronic structure of prereduced catalysts can change substantially under reaction conditions (*15*, *16*). Multiscale operando methodologies (from single nanoparticle to ensemble level) are crucial to develop meaningful structure–performance relations for supported metal NPs. Recent advances in high-resolution TEM methods allow the observation of materials under operando conditions at ambient pressure and at high temperatures relevant to catalysis (*17*–*19*).

Here, we unravel the evolution of TiO_x overlayers on industrially relevant Ni/TiO_2 catalysts during catalysis, for the CO and CO_2 hydrogenation reaction, after low-temperature (400°C) and high-temperature (600°C) reduction. We combined quantitative operando scanning TEM (STEM) and infrared (IR) spectroscopy, near-ambient pressure x-ray photoelectron spectroscopy (NAP-XPS), and density functional theory (DFT) calculations to observe and understand overlayer restructuring and its effect on catalysis. The combination of both nano- and bulk-scale operando analytical techniques reveals information with unprecedented resolution on the single-atom level and also provides new insights into the properties of the ensemble of supported metal nanoparticles.

We show that TiO_x overlayers, formed during reduction, restructure in CO_2:H_2 mixtures at a temperature as low as 200°C, (partially) reexposing the Ni surface and introducing interfacial sites responsible for higher C–C coupling activity. We tailored suboxide overlayers to be stable under reaction conditions using different reduction pretreatment temperatures. This capability adds a parameter for catalyst optimization. These results could enable an understanding of the performance of a number of supported metals in reactions for sustainable technologies, such as biomass upgrading (*20*) and CO_2 valorization (*21*, *22*).

Ni partial encapsulation during 400°C reduction

To induce oxide overlayer coverage by SMSI and to investigate its role in CO_2 hydrogenation, we reduced 6 wt % Ni/TiO_2 catalysts at 400°C (400-Ni/TiO_2). The formation of TiO_x overlayers upon the supported Ni NPs was investigated on the single-particle level by in situ electron microscopy experiments in a windowed gas cell (climate G+, gas supply system, DENSsolutions) (Fig. 1A; see the supplementary materials for details). In situ reduction of Ni/TiO_2 at 400°C yielded supported metallic Ni NPs with an average diameter of 7.4 nm, and a size distribution of ±1.5 nm SD (Fig. 1B and figs. S4 and S5) that were consistent with STEM and x-ray diffraction (XRD) observations after ex situ reduction (figs. S8 and S9 and tables S1 and S2). The metallic state of the Ni NPs was confirmed by the Fourier transforms derived from the atomic-resolution high-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) images (figs. S10 and S11).

A TiO_x overlayer was detected on Ni NPs by in situ HAADF-STEM imaging for 400-Ni/TiO_2 that corresponded to a partially encapsulating bilayer (Fig. 1C and figs. S12 and S13). This bilayer formed preferentially over (111) Ni surfaces for each of the five Ni NPs observed in a crystallographic zone axis, with no evidence of overlayer formation on other Ni facets (figs. S12 and S13). Because we aimed to investigate the catalyst dynamics in H_2 and during CO/CO_2 methanation conditions, it was important that structural changes that we observed were not an artifact of the electron beam. Therefore, we acquired multiple frames of the encapsulated Ni NP with low screen currents and fast image acquisition parameters to minimize potential electron beam–induced artifacts (*23*). After applying neural network restoration of the HAADF-STEM data and frame averaging, we could estimate the positions of the atomic columns with high accuracy and precision. We applied statistical parameter estimation theory to determine and refine the atomic column positions in the Ni NP and in the two atomic layers of Ti (Fig. 1D and figs. S14 to S16) (*24*). In addition, we applied the more-advanced maximum a posteriori probability rule to detect the Ti overlayer atomic columns that yield low contrast-to-noise ratio (*25*, *26*).

Next, the displacements of all atomic columns with respect to the ideal atomic column

[1]Inorganic Chemistry and Catalysis Group, Institute for Sustainable and Circular Chemistry and Debye Institute for Nanomaterials Science, Utrecht University, 3584 CG Utrecht, Netherlands. [2]EMAT and NANOlab Center of Excellence, University of Antwerp, 2020 Antwerp, Belgium. [3]ELCAT, University of Antwerp, 2610 Wilrijk, Belgium. [4]Peter-Grünberg-Institut 6, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany. [5]Department of Surface and Plasma Science, Faculty of Mathematics and Physics, Charles University, 180 00 Prague, Czech Republic. [6]BASF Nederland B.V., 3454 PK De Meern, Netherlands.
*Corresponding author. Email: Sara.Bals@uantwerpen.be (S.B.); B.M.Weckhuysen@uu.nl (B.M.W.)
†These authors contributed equally to this work.
‡Present address: Schulich Faculty of Chemistry, Technion, Israel Institute of Technology, Technion City, Haifa 32000, Israel.

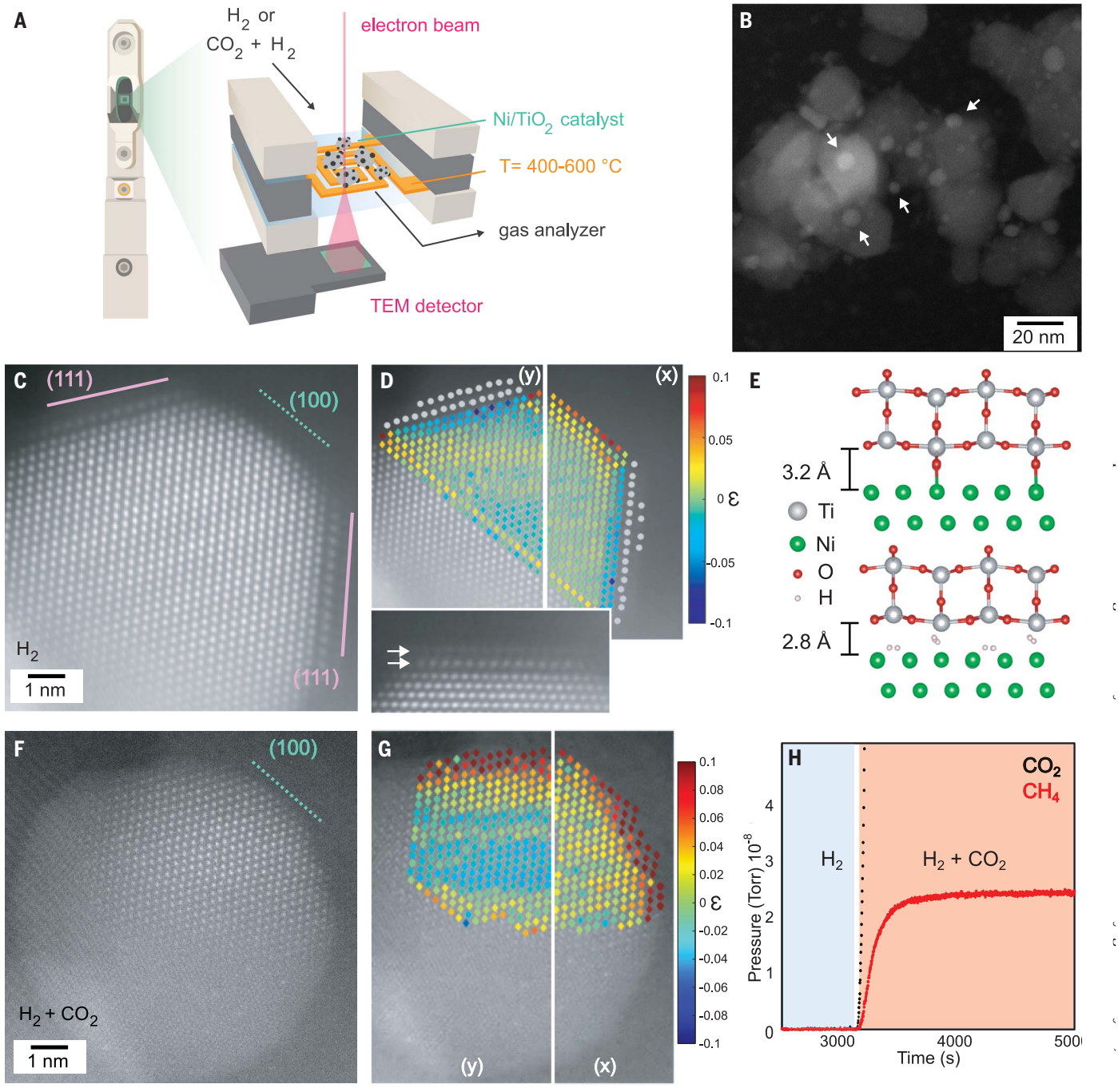


Fig. 1. TiO_x overlayer formation by low-temperature H_2 reduction and its restructuring during CO_2 methanation on a Ni/TiO_2 hydrogenation catalyst. (**A**) Schematic of the operando electron microscopy setup and windowed gas cell (climate G+, DENSsolutions) used in this study. (**B**) HAADF-STEM overview image of a Ni/TiO_2 catalyst prepared in situ at 400°C (400-Ni/TiO_2), where white arrows highlight the Ni NPs upon the TiO_2 support. (**C**) Representative atomic-resolution HAADF-STEM image of a Ni NP in a Ni/TiO_2 catalyst in H_2 at 400°C. Solid lines indicate TiO_x-covered Ni atomic planes, and dashed lines indicate unoccupied facets. Ni(111) and Ni(100) facets are indicated by pink and green, respectively. (**D**) Estimated positions of Ti (gray) and Ni (color scale) atomic columns. Ni atomic columns were identified through statistical parameter estimation theory and neural network image restoration, and Ti atomic positions were identified using a combination of statistical parameter estimation theory and MAP probability rule because of Ti's limited visibility (*25*, *26*). Strain maps (resulting from the displacements with respect to the ideal atomic column positions) in the *x* and *y* directions. Inset: Close-up of the TiO_x bilayer, where white arrows highlight each Ti layer. (**E**) Model of rutile TiO_2 (110) overlayer on a Ni(111) slab with oxygen (top) or hydrogen (bottom) atoms occupying the Ni–Ti interface calculated by DFT, suggesting that the titania overlayer is partially reduced at the interface. (**F**) HAADF-STEM image of the same particle as in (D) upon exposure to a CO_2:H_2 (0.25 bar: 0.75 bar) mixture at 400°C showing complete reexposure of Ni and NP restructuring. (**G**) Estimated atomic column positions of Ni from the same particle shown in (F) identified through statistical parameter estimation theory and neural network image restoration. Strain maps in the *x* and *y* directions resulting from the displacements with respect to the ideal atomic column positions. (**H**) MS data measuring the partial pressure of CO_2 (black) and CH_4 (red) for the Ni/TiO_2 catalysts at 400°C. Regions highlighted in light blue and pink correspond to H_2 and CO_2:H_2 environments, respectively.

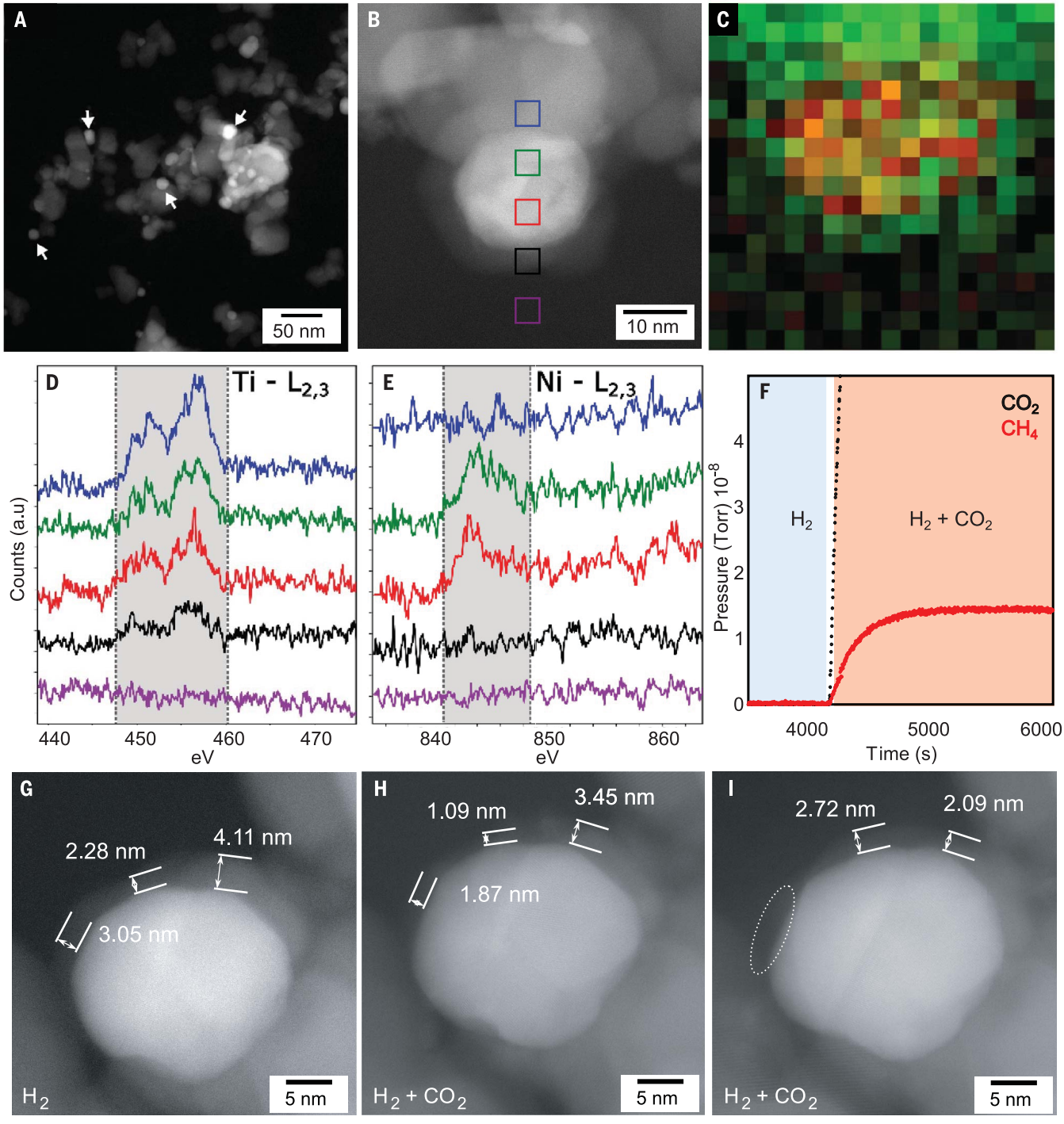


Fig. 2. Formation of stable TiO_x/Ni overlayers during CO_2 methanation on a Ni/TiO_2 hydrogenation catalyst first reduced at high temperature. (**A**) HAADF-STEM overview images of Ni/TiO_2 catalyst prepared in situ at 600°C (600-Ni/TiO_2), where white arrows highlight the Ni NPs upon the TiO_2 support. (**B**) High-resolution HAADF-STEM image of Ni/TiO_2 catalyst in H_2 at 600°C showing a Ni NP encapsulated in a thick TiO_x shell. Colored squares correspond to the location where EELS spectra in (D) and (E) were acquired. (**C**) Core loss EELS map of Ti $L_{2,3}$ (green) and Ni–$L_{2,3}$ (red) ionization edges. (**D** and **E**) Core loss Ti $L_{2,3}$ (D) and Ni $L_{2,3}$ EELS (E) spectra showing several regions of interests within the encapsulated NP, where spectrum color indicates the location of pixels shown in (B). (**F**) MS data acquired at 400°C showing the partial pressure of CO_2 (black) and CH_4 (red) for Ni/TiO_2 catalysts prepared at 600°C, with the regions highlighted in light blue and pink corresponding to the H_2 and CO_2:H_2 environments, respectively. (**G**) High-resolution HAADF-STEM image of an individual Ni NP in H_2 at 600°C before CO_2 hydrogenation. (**H**) High-resolution HAADF-STEM image of the same particle as in (G) upon exposure to a CO_2:H_2 (0.25 bar CO_2: 0.75 bar H_2) mixture at 400°C showing only partial restructuring of the thick TiO_x shell. (**I**) HAADF-STEM image of the same particle as in (G) and (H) upon longer exposure to a CO_2:H_2 (0.25 bar CO_2: 0.75 bar H_2) mixture at 400°C showing further restructuring of the thick TiO_x shell and reexposure of the Ni surface, highlighted by the white circle.

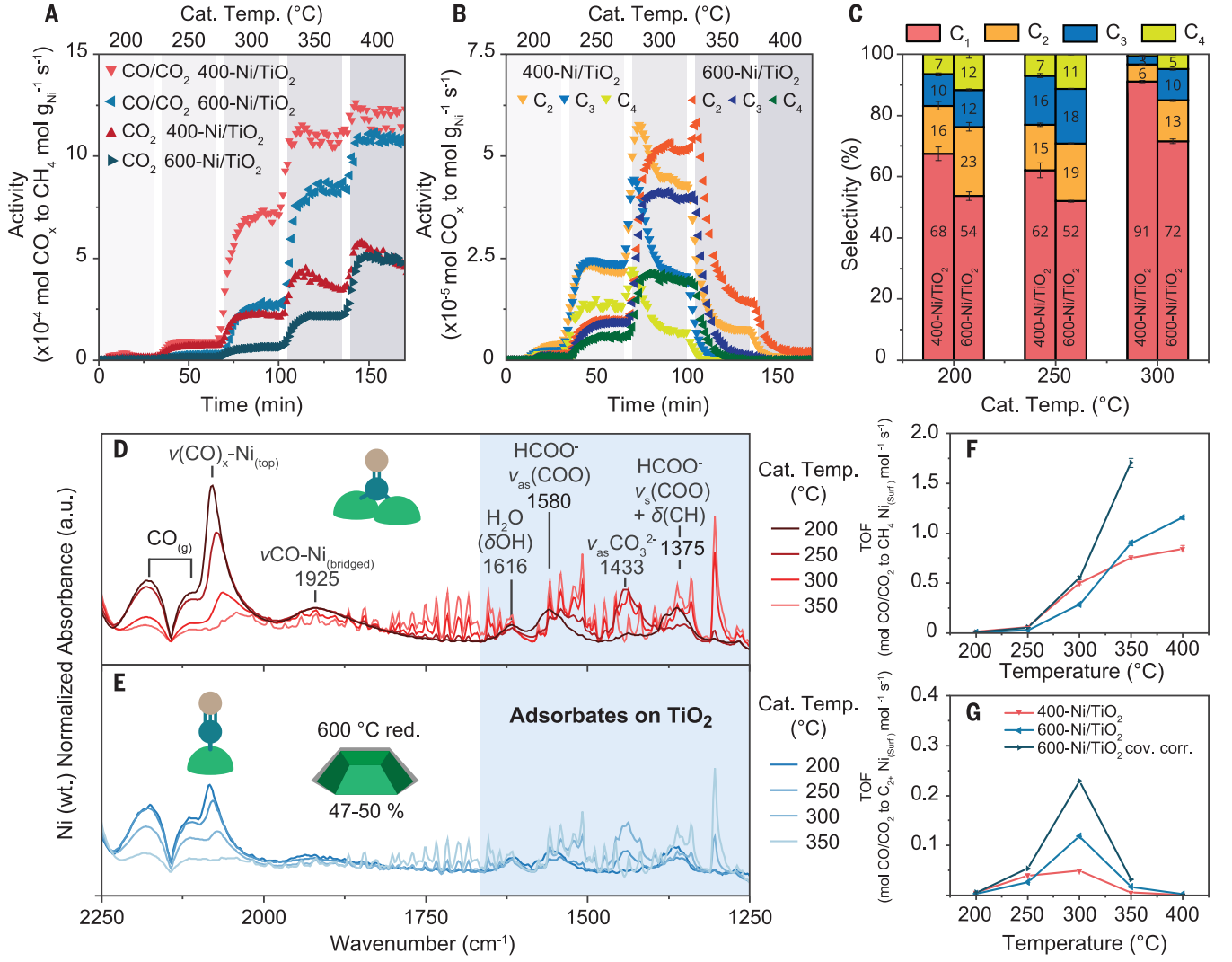


Fig. 3. Ni/TiO_2 reexposure under CO and CO_2 hydrogenation conditions revealed by operando vibrational spectroscopy. (**A**) Catalytic activity of 400-Ni/TiO_2 (red) and 600-Ni/TiO_2 (blue), normalized per gram of Ni, for CO_2 (dark shade) and CO/CO_2 methanation (light shade). (**B**) C–C coupling activity of 400-Ni/TiO_2 and 600-Ni/TiO_2 normalized per gram of Ni, during CO/CO_2 hydrogenation. (**C**) Selectivity to hydrocarbons in the CO/CO_2 hydrogenation reaction over 400-Ni/TiO_2 and 600-Ni/TiO_2. (**D** and **E**) Operando FTIR spectra collected at 200° to 300°C under CO/CO_2 hydrogenation conditions over 400-Ni/TiO_2 (D) and 600-Ni/TiO_2 (E) showing Ni exposure in both catalysts. The estimate for Ni coverage for 600-Ni/TiO_2 and the majority CO_{ads} species at 300°C is schematically depicted for the two catalysts (see the supplementary materials for details). (**F** and **G**) TOF for methane production (F) and C–C coupling products (G) over 400-Ni/TiO_2 (red) and 600-Ni/TiO_2 (light blue). The dark blue line shows the TOF corrected for the most conservative estimate of exposed Ni surface in 600-Ni/TiO_2 (see the supplementary materials for details). Conditions for CO_2 methanation: CO_2:H_2:He = 1:4:5, 5 bar, gas hourly space velocity (GHSV) = 80,000 $hours^{-1}$. Conditions for CO/CO_2 methanation: CO_2:CO:H_2:He = 3:1:7:9, 5 bar, GHSV = 80,000 $hours^{-1}$.

positions in the Ni lattice were determined, from which we calculated a strain map (Fig. 1D and figs. S15 and S16). This quantitative analysis revealed an increase in tensile strain at the exposed (100) surface facet, as expected from the increased mobility of surface adatoms. By contrast, both (111) facets covered by a TiO_x bilayer displayed a small negative (compressive) strain over the surface and the second atomic subsurface layer as Ni accommodated the Ni–Ti interface. A compressive strain of Ni atoms at the TiO_x/Ni(111) interface was also predicted by DFT calculations for different TiO_x/Ni models (tables S3 and S4).

The interatomic Ti–Ti and Ni–Ti distances were measured using the atomic positions deduced by our statistics-based approaches and found to be 2.95 ± 0.01 and 2.93 ± 0.09 Å, respectively. The Ti–Ti distances showed a range of 2.77 to 2.95 Å for several measured particles using this technique (fig. S17), and of the possible TiO_x polymorphs, our measurements are consistent with a compressed rutile (110) *d* spacing (Fig. 1D and fig. S10). This Ti–Ti distance is also consistent with continuous TiO_x overlayers previously reported (*13*), and we anticipate that the observed compression was partially caused by the discrete and minimal secondary layer of Ti atoms that was too small to arrange perfectly, as would be expected for a bulk phase–idealized crystal structure. The identity of the TiO_x overlayer was further distinguished from a NiO_x overlayer through in situ electron energy-loss spectroscopy (EELS) analysis (fig. S18). Similar observations were made regardless of which crystal structure (rutile or anatase) the Ni nanoparticle resided on (fig. S19).

To better understand our finding of the selective coverage of Ni(111) facets by bilayers of TiO_2 rutile (110), we calculated the relative stability of TiO_2 overlayers on the most observed

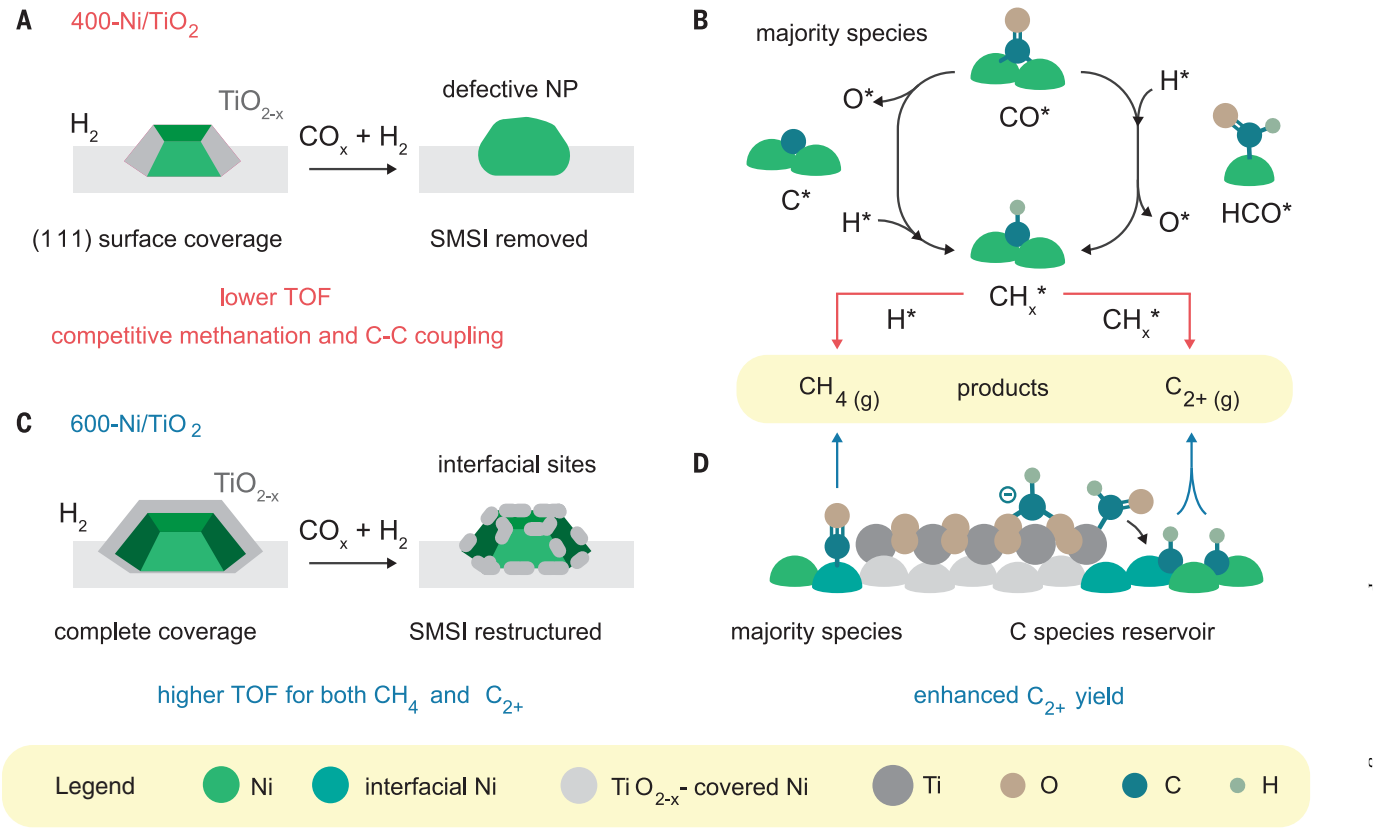


Fig. 4. New understanding of Ni/TiO_2 catalyst performance and evolution during reduction and reaction. Model for TiO_x overlayer formation after 400 and 600°C reduction and restructuring under reaction conditions, which explains the observed catalytic and spectroscopic results. (**A**) 400°C reduction leads to the formation of a TiO_x bilayer that is completely removed under reaction conditions, resulting in defective Ni particles. (**B**) For these particles, CO can adsorb on bridge sites (majority species). The adsorbed CO* is converted to CH_x* through either HCO* (H-assisted dissociation) or C* (direct dissociation). This species is in turn converted to methane by hydrogenation (reaction with H*) or to C–C coupling products (C_{2+}) by reaction with CH_x* adsorbates. (**C**) 600°C reduction leads to complete Ni coverage by TiO_x, which restructures under reaction conditions. (**D**) On these particles, interfacial Ni-TiO_x sites form, upon which C species from the support are brought in contact with C species on Ni, favoring C–C coupling. Furthermore, CO is preferentially bound linearly, which explains the higher TOF per active site in the catalytic methanation of CO and CO_2.

Ni facets using DFT (figs. S20 and S21). Based on the quantitative structural analysis of the atomic resolution in situ HAADF-STEM data, we built TiO_2/Ni models using a Ni slab of four atomic layers oriented along the (111), (100), and (110) facets covered by two monolayers of rutile (110) (see the "Computational details" section in the supplementary materials). The lattice mismatch between the TiO_2 rutile structure and the Ni surfaces would induce strain. After energy minimization, we found the strain energies for the Ni(111) and Ni(100) surfaces to be 0.5 and 3.1 kJ/mole of TiO_2, respectively. This difference was consistent with the distortion and irregularity of the modeled TiO_2 overlayer on Ni(100) (Fig. 1E and fig. S20). The lower energy toll calculated for the formation of TiO_2 overlayers on Ni(111) thus explains the experimentally observed facet-selective coverage of Ni NPs.

To further investigate the structure of the TiO_x/Ni(111) interface, we used different configurations as input for computational calculations, varying the type and number of atoms between the interfacial Ni and TiO_x layers (figs. S20 and S21). When the Ni surface was covered with a TiO_2 rutile (110), the computed Ni–Ti interlayer distance was 3.20 Å, which is larger than what was experimentally observed (2.80 to 3.00 Å; Fig. 1E, top). Conversely, a Ni–Ti distance of 2.79 Å was calculated for a partially reduced overlayer, where interfacial oxygen atoms were replaced by H atoms, and H* atoms were adsorbed on the Ni surface. On the basis of these calculations, the overlayers are most likely a suboxide phase.

Because in situ HAADF-STEM imaging provided only local information on single NPs, we performed H_2 chemisorption to estimate the bulk average percentage of exposed Ni surface for the ensemble of Ni NPs after the reduction of Ni/TiO_2 (fig. S23). The amount of chemisorbed H_2 for 400-Ni/TiO_2 was 32 μmol/g, which corresponds to 6.7% of the total number of Ni atoms, or an apparent average NP size of 15 nm, assuming a 1:1 H:Ni stoichiometry. The apparent Ni NP size calculated from H_2 chemisorption was substantially larger than that measured with TEM (5.3 to 7.5 nm; Fig. 1B and figs. S4, S5, and S7), indicating that part of the surface of Ni was blocked and not accessible to the gas phase. If no chemisorption took place on the TiO_x overlayers, then there was a 25 to 46% surface coverage of Ni by TiO_x in 400-Ni/TiO_2. This value is in agreement with the fraction of Ni(111) surface sites (35 to 57%) predicted for Wulff-constructed Ni NPs of such sizes, (*27*) as well as with the in situ HAADF-STEM evidence of selective encapsulation of Ni(111) facets.

Ni reexposure during CO_2 hydrogenation for Ni/TiO_2 after 400°C reduction

We followed changes in the TiO_x overlayer under CO_2 hydrogenation reaction conditions by tracking a single Ni NP with HAADF-STEM while switching from H_2 to CO_2:H_2 at 400°C and atmospheric pressure. Reactant and product gases from the windowed gas cell were tracked in real time during the reaction using mass spectrometry (MS) (Fig. 1H and fig. S24). Upon introduction of the CO_2:H_2 mixture, there

was detectable formation of CH_4, confirming that the catalyst actively performed CO_2 hydrogenation during observation within the electron microscope. As observed for other NPs (figs. S12 and S13), the Ni NP shown in Fig. 1C developed a TiO_x overlayer that selectively occupied Ni's (111) surface facets before hydrogenation.

Upon CO_2 exposure at 400°C, the same Ni NP restructured and lost a large proportion of its highly faceted surface, with only a small percentage of the Ni(100) facet remaining. The NP exhibited an overall rounded morphology and retained its Ni metallic phase (Fig. 1F and figs. S25 and S26). Comparative strain maps calculated for the same NP during CO_2 hydrogenation showed an increase in tensile strain over the entire NP surface in both the x- and y-directions and a loss in compressive strain (Fig. 1G). We attributed these changes to the loss of the Ni–Ti interface during hydrogenation, which resulted in lower atomic coordination and higher mobility of surface atoms compared with its TiO_2-encapsulated counterpart (400-Ni/TiO_2; Fig. 1, C and D). The supported Ni catalyst dynamically evolved to express its more-active facets during CO/CO_2 hydrogenation (*28*). Thermodynamic calculations by the Gibbs free energy minimization method showed that TiO_x became unstable under CO_2 hydrogenation conditions (table S6; see the supplementary materials for details), which suggests that the removal of the TiO_x overlayer was induced by changes in chemical potential and not by the Ni NP restructuring.

To show that such removal of TiO_x overlayer upon exposure to CO_2 hydrogenation conditions is representative of the bulk of the sample, we measured the attenuation of Ni 2p signal relatively to Ti 2p signals in NAP-XPS, which provided indirect evidence for encapsulation (fig. S27). A fresh Ni/TiO_2 catalyst was reduced at 400°C in situ at 0.3 mbar H_2 and then exposed to cycles of 0.3 mbar H_2 + 0.3 mbar CO_2 and pure H_2. This sequence resulted in a reversible change of the Ni/Ti signal ratio that was consistent with the exposure and coverage of Ni under CO_2 hydrogenation reaction conditions and H_2, respectively (fig. S29; see the supplementary materials for details).

TiO_x overlayer formation and restructuring for 600°C reduced Ni/TiO_2

With the aim of forming a more-stable TiO_x overlayer that would survive CO_2 hydrogenation conditions, we reduced the 6 wt % Ni/TiO_2 catalyst in situ at 600°C (600-Ni/TiO_2). Compared with the Ni/TiO_2 catalyst prepared at 400°C (400-Ni/TiO_2), the NP diameter approximately doubled to 14.0 nm (±4.0 nm size distribution) for the 600-Ni/TiO_2 (Fig. 2A and figs. S6 and S7). The larger NP diameter is the result of the higher reduction temperature and is consistent with TEM and XRD observations after ex situ reduction (tables S1 and S2 and figs. S6 to S8). 600-Ni/TiO_2 displayed a completely encapsulating and relatively thick (1 to 2 nm) TiO_x overlayer (Fig. 2B and fig. S30). The greater degree of encapsulation occurred with the higher temperature treatment because the partial reduction of TiO_2 that is needed to form the TiO_x overlayers is favored at higher temperatures (figs. S2 and S3). Because of the lack of a single crystalline layer, atomic-resolution HAADF-STEM was not applicable. Therefore, the formation of a TiO_x overlayer was corroborated by in situ EELS analysis, which showed a Ti $L_{2,3}$ ionization edge signal throughout the entire Ni NP (Fig. 2, B to E, and figs. S31 to S33). Consistent with the complete encapsulation by TiO_x observed by electron microscopy, H_2 chemisorption was almost completely suppressed for the 600-Ni/TiO_2 catalysts (fig. S23).

After reduction, the in situ–prepared 600-Ni/TiO_2 catalyst was exposed to CO_2 hydrogenation reaction conditions at 400°C. Similar to the 400-Ni/TiO_2, CO_2 methanation was confirmed by tracking reactant and product gases by MS (Fig. 2F). Figure 2, G to I, shows the same Ni NP under H_2 (Fig. 2G) and under a CO_2:H_2 mixture (Fig. 2, H and I) at atmospheric pressure. The TiO_x-suboxide overlayer was still visible but became more inhomogeneous after CO_2 exposure compared with the pristine catalyst, indicating its partial degradation. Shell thickness measurements in Fig. 2, G to I, show the regions in which most changes in coverage were observed, with a circle identifying an area of the Ni surface that no longer appeared to be encapsulated by the overlayer (fig. S34; see the supplementary materials for details).

Such overlayers survived even exposure to pure CO_2 at 400°C (fig. S34), in contrast to the easily removed TiO_x bilayer observed for 400-Ni/TiO_2 catalysts. We hypothesize that the different stability is caused by the greater thickness of the shell, which contained sufficient Ti atoms to form stable patches of TiO_x on Ni. Despite the extended encapsulation, the catalyst was active for the CO_2 hydrogenation reaction to methane, as shown by online MS analysis (Fig. 2F and fig. S24).

Our operando STEM results suggest that 600-Ni/TiO_2 remained predominantly encapsulated by TiO_x under CO_2 hydrogenation reaction conditions. Therefore, we expected to observe an almost complete loss of activity for such catalysts in reactions promoted by Ni. However, both the 400- and 600-Ni/TiO_2 catalysts were active for CO_2 hydrogenation during operando STEM experiments (Figs. 1H and 2F).

Comparison of catalytic performance

To investigate the effect of the residual TiO_x overlayer on the catalytic CO/CO_2 hydrogenation performance of Ni/TiO_2 catalysts, we performed catalytic tests on 20 mg of catalysts in operando Fourier transform infrared (FTIR) spectroscopy experiments under hydrogenation conditions. These experiments showed that both catalysts were active for CO_2 hydrogenation at 200° to 400°C and 5 bar (Fig. 3A and figs. S35 to S38), in agreement with operando TEM MS measurements. Compared with 400-Ni/TiO_2, 600-Ni/TiO_2 showed a decrease in overall catalytic activity per gram of Ni, but a higher stability (Fig. 3A), and higher C_{2+} selectivity (figs. S36 and S37). All operando FTIR spectroscopy and catalytic results reported herein were repeated on two separate batches of Ni/TiO_2 catalysts, showing comparable figures and trends (figs. S45 to S50).

The enhancement in C_{2+} selectivity for the 600-Ni/TiO_2 catalyst became even more evident in CO/CO_2 cohydrogenation experiments, with a C_{2+} selectivity of 28 and 9% for 600- and 400-Ni/TiO_2 catalysts, respectively, at 300°C and 5 bar pressure (Fig. 3, A to C). These results further indicate that under reaction conditions, the Ni surface in 600-Ni/TiO_2 was partially reexposed, and that this affected the product selectivity of the reaction. Accordingly, ex situ Ni 2p XPS spectra of the used 600-Ni/TiO_2 catalyst (after one reaction cycle, as in Fig. 3A) showed a higher Ni signal intensity than the reduced 600-Ni/TiO_2, further supporting Ni reexposure under reaction conditions (fig. S27).

To provide direct evidence for the reexposure of the Ni surface under reaction conditions and to quantify the number of exposed sites, we performed operando FTIR spectroscopy experiments during CO/CO_2 cohydrogenation (Fig. 3, D and E) and CO_2 hydrogenation (fig. S38). The 400- and 600-Ni/TiO_2 catalysts showed similar CO_x hydrogenation intermediate signals, but with different relative intensities: $CO_{(g)}$ bands were observed in the region 2250 to 2100 cm^{-1}, followed by relatively sharp bands at 2078 to 2062 cm^{-1} (peak positions depended on temperature; figs. S40 and S41), which we ascribed to adsorbed sub-carbonyl Ni $(CO)_x$ (with x = 2,3) species (*29*–*31*). $Ni(CO)_4$ formation under the conditions of this study was ruled out by thermodynamic calculations (table S7). A component corresponding to Ni-CO_{top} was observed in the spectral region between 2060 and 2032 cm^{-1}, whereas bridged CO (CO_{br}) adsorbed on Ni was observed in the spectral region between 1920 and 1940 cm^{-1}. A series of bands at lower wave numbers were ascribed to adsorbates on TiO_2, because they were also present over the pure support {fig. S42: adsorbed water peaks at 1616 cm^{-1} [δ (OH)]; formate [$HCOO^-$] peaks at 1580 and 1375 cm^{-1}; and carbonates [CO_3^{2-}] peak at 1433 cm^{-1} (*32*)}.

Because the bands of adsorbed CO species (2078 to 1750 cm^{-1}) were absent over pure TiO_2 under the same reaction conditions (fig.

S42), these spectral features were taken as a measure of the extent of Ni surface exposure under reaction conditions. Assuming complete reexposure of 400-Ni/TiO_2 under reaction conditions, as evidenced by TEM (Fig. 1F and fig. S26), the integration of the CO bands for 400- and 600-Ni/TiO_2 revealed a TiO_x coverage of ~ 50% for 600-Ni/TiO_2, decreasing with increasing reaction temperature (table S5). These results, combined with the STEM results shown in Fig. 2, G to I, provided evidence for a reexposed Ni surface, partially covered by TiO_x agglomerates, that remained on the Ni NP surfaces.

Taking into account such a fraction of exposed Ni surface derived by operando FTIR spectroscopy, we calculated exposure-corrected turnover frequencies (TOFs) for methanation and C–C coupling (Fig. 3, F and G, and figs. S43 and S44; see the supplementary materials for details). Although the geometrical (i.e., not corrected by coverage) TOFs of 400- and 600-Ni/TiO_2 were comparable, the exposure-corrected TOFs of 600-Ni/TiO_2 were higher, indicating that the reexposed Ni sites had a higher specific activity for both methanation and C–C coupling.

Mechanistic insights and structure sensitivity

On the basis of our results, we propose a model of the restructured Ni/TiO_2 catalysts under reaction conditions to explain changes in surface chemistry and, ultimately, in catalyst performance. The higher methanation TOF for 600-Ni/TiO_2 may be explained by a change in most of the CO_{ads} species, from less-active, bridged CO_{ads} for 400-Ni/TiO_2 (Fig. 4, A and B) to more active, linear CO_{ads} for 600-Ni/TiO_2 (Fig. 4, C and D, and fig. S40B) (*28*). Because Ni NPs were larger in the 600-Ni/TiO_2 catalyst, and more extended Ni surfaces should result in more bridged CO_{ads} (*33*), this observation can best be explained by the formation of a patchy TiO_x overlayer that increased the extent of the Ni–TiO_x interface and led to fewer available sites for CO_{bridge} formation, even on extended surfaces. Similarly, the enhancement in CO hydrogenation TOF on Pt/TiO_2 catalysts was caused by cooperative adsorption of CO at the interface, with the C atom coordinated to Pt and oxygen on the oxygen-deficient TiO_x (*34*). The formation of such structures was shown by the combination of operando FTIR spectroscopy and HAADF-STEM results (Figs. 2 and 3). In the case of the 400-Ni/TiO_2, the interface among Ni, TiO_2, and the gas phase is instead limited to the perimeter of the NP in contact with the support because of the complete removal of the TiO_x overlayers under reaction conditions.

The higher C–C coupling TOF for 600-Ni/TiO_2 was also in contrast with previous observations on Ni/SiO_2 catalysts, where C–C coupling was anticorrelated with methanation activity, because it was limited by the high hydrogenation rate of CH_x* (x = 0 to 3) adsorbates on Ni. Both C–C coupling and methanation having higher TOF for 600-Ni/TiO_2 suggested that an alternative pathway was available for C–C coupling in the presence of a patchy TiO_x overlayer. We propose that the increased Ni–TiO_x interface aided C–C coupling by providing a reservoir of C species in close contact with adsorbates on Ni and by stabilizing intermediates and transition states (Fig. 4D). Accordingly, calculations on TiO_x/Ni catalysts showed that CO activation and CH–CH coupling were both more favorable at interfacial Ni-TiO_x sites (*35*). An alternative explanation for enhanced C–C coupling would involve reverse H spillover from Ni to TiO_x, hindering hydrogenation of CH_x* adsorbates by H*. However, this mechanism is not consistent with the observed higher methanation TOF, which requires H*. Ex situ XRD showed no extended TiO_2 phase transformation, such as from anatase to rutile, at 600°C (fig. S9) that could have affected the catalytic activity (*36*), although further studies are required to substantiate this explanation.

Discussion

The observed restructuring of TiO_x overlayer under reaction conditions was consistent with recent observations on Pt/TiO_2 by in situ TEM during reduction-oxidation-reduction cycles at 600°C, in which oxidative treatments cause amorphous TiO_2 islands to form, reexposing part of the metal surface (*10*). CO/CO_2 hydrogenation conditions were sufficiently oxidative to shift the thermodynamic equilibrium to TiO_2 (table S6). Our results show that, kinetically, a (partial) removal and restructuring of TiO_x overlayers could occur even at temperatures as low as 200°C, reexposing the active Ni surface.

Both CO and CO_2 hydrogenations are structure-sensitive reactions (*33*, *37*), meaning that the TOF in the reactions changes as a function of NP size (*38*). Nonetheless, structure sensitivity only resulted in lower TOF for NPs with diameters below 2 to 4 nm, well below the range considered in this study (*28*, *37*). Conversely, NPs with larger sizes have been reported to have similar, or even smaller, TOF, which would make the effect of SMSI in TOF enhancement even greater (*33*). Finally, C–C coupling was not strongly structure sensitive, according to theoretical calculations on Ni surfaces (*28*). Thus, the current observations that the TOF of both methanation and C–C coupling increase with Ni NP size could not possibly be explained by classical structure sensitivity concept, which only accounts for Ni surface sites variation with NP size (*39*). Similar deviations from classical structure sensitivity were observed for Co-based catalysts in CO_2 hydrogenation, where partially reduced cobalt oxide covered with metallic clusters of a few Co atoms showed substantially higher intrinsic activity caused by interfacial effects (*40*).

In summary, here, we directly observed for the first time the restructuring of metal oxide overlayers under operating conditions and rationalized their effects on the activity and selectivity of Ni in CO/CO_2 hydrogenation reactions depending on the reduction temperature. A similar, although rudimental, model of highly active interfacial sites was proposed by Burch *et al.* in the 1980s to explain the higher activity of Ni/TiO_2 compared with Ni/SiO_2 on the basis of chemisorption results alone (*41*). Evidence that overlayers may be stabilized under reaction conditions was reported for Rh/TiO_2 catalysts, where pretreatments in CO_2-rich atmospheres lead to the formation of a more-persistent, "adsorbate-induced" SMSI (*14*). Recently, the enhancement of activity in methanol steam–reforming reaction was also correlated with the amount of Cu–ZnO_x interfacial sites (albeit observed ex situ), exhibiting faster kinetics for *CH_3O dehydrogenation and H_2O dissociation (*8*). We now provide operando evidence for the formation of Ni–TiO_x interfacial sites and show their effect on catalytic performance. A similar operando approach could be applied to understand many other chemical reactions in which the proximity of the active phase and support or promoters has been found to be essential for higher activity or better selectivity, such as olefins production over Fe-based catalysts promoted by K (*42*), CO oxidation over Pd/CeO_2 (*43*), or propane dehydrogenation over Pt/Sn/CeO_2 (*44*). Crucial in this endeavor will be the simultaneous measurement of catalytic performance and the monitoring and control of the support, such as crystal phase, porosity, oxidation state, and exposed facets, and of the supported active phase nanostructure, such as NP size, shape, and composition.

REFERENCES AND NOTES

1. C. Vogt *et al.*, *Nat. Commun.* **12**, 7096 (2021).
2. A. V. Puga, *Catal. Sci. Technol.* **8**, 5681–5707 (2018).
3. S. B. Vendelbo *et al.*, *Nat. Mater.* **13**, 884–890 (2014).
4. S. J. Tauster, S. C. Fung, R. L. Garten, *J. Am. Chem. Soc.* **100**, 170–175 (1978).
5. M. Tang *et al.*, *Angew. Chem. Int. Ed.* **60**, 22339–22344 (2021).
6. C. Wu *et al.*, *Nat. Commun.* **11**, 5767 (2020).
7. S. Samira *et al.*, *JACS Au* **1**, 2224–2241 (2021).
8. D. Li *et al.*, *Nat. Catal.* **5**, 99–108 (2022).
9. M. Macino *et al.*, *Nat. Catal.* **2**, 873–881 (2019).
10. A. Beck *et al.*, *Nat. Commun.* **11**, 3220 (2020).
11. C. Hernández Mejía, T. W. van Deelen, K. P. de Jong, *Nat. Commun.* **9**, 4459 (2018).
12. H. Frey, A. Beck, X. Huang, J. A. van Bokhoven, M. G. Willinger, *Science* **376**, 982–987 (2022).
13. S. Zhang *et al.*, *Nano Lett.* **16**, 4528–4534 (2016).
14. J. C. Matsubu *et al.*, *Nat. Chem.* **9**, 120–127 (2017).
15. A. Bergmann, B. Roldan Cuenya, *ACS Catal.* **9**, 10020–10043 (2019).
16. X. Zhang *et al.*, *Nat. Catal.* **3**, 411–417 (2020).
17. S. W. Chee, T. Lunkenbein, R. Schlögl, B. R. Cuenya, *J. Phys. Condens. Matter* **33**, 153001 (2021).
18. A. S. Kashin, V. P. Ananikov, *Nat. Rev. Chem.* **3**, 624–637 (2019).

19. S. Hwang, X. Chen, G. Zhou, D. Su, *Adv. Energy Mater.* **10**, 1902105 (2020).
20. X. Zhang *et al*., *ACS Catal.* **9**, 3551–3563 (2019).
21. A. Dokania, A. Ramirez, A. Bavykina, J. Gascon, *ACS Energy Lett.* **4**, 167–176 (2019).
22. M. Li *et al*., *ACS Appl. Energy Mater.* **4**, 12326–12335 (2021).
23. T. Altantzis *et al*., *Nano Lett.* **19**, 477–481 (2019).
24. A. De Backer, G. T. Martinez, A. Rosenauer, S. Van Aert, *Ultramicroscopy* **134**, 23–33 (2013).
25. J. Fatermans, S. Van Aert, A. J. den Dekker, *Ultramicroscopy* **201**, 81–91 (2019).
26. J. Fatermans *et al*., *Phys. Rev. Lett.* **121**, 056101 (2018).
27. E. B. Sterk *et al*., *JACS Au* **2**, 2714–2730 (2022).
28. C. Vogt *et al*., *Nat. Commun.* **10**, 5330 (2019).
29. K. Hadjiivanov *et al*., *J. Catal.* **185**, 314–323 (1999).
30. M. Agnelli, H. M. Swaan, C. Marquez-Alvarez, G. A. Martin, C. Mirodatos, *J. Catal.* **175**, 117–128 (1998).
31. C. Hernández Mejía, C. Vogt, B. M. Weckhuysen, K. P. de Jong, *Catal. Today* **343**, 56–62 (2020).
32. A. Davydov, *Molecular Spectroscopy of Oxide Catalyst Surfaces* (Wiley, 2003).
33. C. Vogt *et al*., *Nat. Catal.* **1**, 127–134 (2018).
34. B. Cornils, W. A. Herrmann, J.-H. Xu, H.-W. Zanthoff, *Catalysis from A to Z: A Concise Encyclopedia* (Wiley, 5th ed., 2020).
35. M. Xu *et al*., *Nat. Commun.* **13**, 6720 (2022).
36. D. A. H. Hanaor, C. C. Sorrell, *J. Mater. Sci.* **46**, 855–874 (2011).
37. G. L. Bezemer *et al*., *J. Am. Chem. Soc.* **128**, 3956–3964 (2006).
38. R. A. Van Santen, *Acc. Chem. Res.* **42**, 57–66 (2009).
39. C. Vogt, B. M. Weckhuysen, *Nat. Rev. Chem.* **6**, 89–111 (2022).
40. A. Parastaev *et al*., *Nat. Catal.* **5**, 1051–1060 (2022).
41. R. Burch, A. R. Flambard, *J. Catal.* **78**, 389–405 (1982).
42. Y. Han *et al*., *ACS Catal.* **10**, 12098–12108 (2020).
43. M. Cargnello *et al*., *Science* **341**, 771–773 (2013).
44. H. Xiong *et al*., *Angew. Chem. Int. Ed.* **56**, 8986–8991 (2017).

ACKNOWLEDGMENTS

B.M.W. thanks NWO for a CHIPP Research Grant, a Gravitation Program [Netherlands Center for Multiscale Catalytic Energy Conversion (MCEC)] grant, and an Advanced Research Center Chemical Building Blocks Consortium (ARC CBBC) grant. We thank S. Turner for performing the STEM measurements, L. Smulders for H_2 chemisorption experiments, C. Hernández Mejía for performing the TPR-TPO experiments, and J. Harmel and K. Cheng for performing the H_2 chemisorption measurements (all from Utrecht University). This work made use of the Dutch national e-infrastructure with the support of the SURF Cooperative using grant no. EINF-2959. **Funding:** This work was supported by BASF and NWO CHIPP (research grant to B.M.W.); the MCEC NWO Gravitation Program (B.M.W.); the ARC-CBBC NWO Program (B.M.W.); the European Research Council (grant 770887 PICOMETRICS to S.V.A.); and the European Research Council (grant 815128 REALNANO to S.B.). **Author contributions:** Conceptualization: M.M., K.J., A.E.M.M., S.B., B.M.W.; Funding acquisition: B.M.W., S.B.; Investigation: M.M., K.J., A.E.M.M., C.V., T.D., E.G., J.N.L., W.v.d.S., B.Š.; Methodology: M.M., K.J., A.E.M.M., C.V., T.D., B.S., E.G., J.N.L., E.A.I., S.V.A., T.A.; Project administration: E.G., P.B.; Supervision: M.M., B.M.W., S.B.; Visualization: K.J., M.M., A.E.M.M., T.D., J.N.L., E.A.I.; Writing: M.M., K.J., A.E.M.M., S.B., B.M.W. **Competing interests:** The authors declare that no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in this study are present in the main text or the supplementary materials.
License information: Copyright © 2023 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.adf6984
Materials and Methods
Figs. S1 to S50
Tables S1 to S11
References (*45*–*72*)

Submitted 7 November 2022; accepted 11 April 2023
10.1126/science.adf6984

THERMAL REGULATION

"Breath holding" as a thermoregulation strategy in the deep-diving scalloped hammerhead shark

Mark Royer[1]*, Carl Meyer[1], John Royer[2], Kelsey Maloney[1], Edward Cardona[1], Chloé Blandino[1], Guilherme Fernandes da Silva[3,4], Kate Whittingham[5], Kim N. Holland[1]

Fish moving between different thermal environments experience heat exchange via conduction through the body wall and convection from blood flow across the gills. We report a strategy of preventing convective heat loss at the gills during excursions into deep, cold water by the tropical scalloped hammerhead shark (*Sphryna lewini*). Adult scalloped hammerhead sharks dive rapidly and repeatedly from warm (~26°C) surface waters to depths exceeding 800 meters with temperatures as low as 5°C. Biologgers attached to adult sharks show that warm muscle temperatures were maintained throughout the deepest portion of each dive. Substantive cooling only occurred during the latter stages of the ascent phase and, once initiated, was rapid. Heat transfer coefficient modeling indicated that convective heat transfer was suspended, probably by suppressing gill function during deep dives. This previously unobserved strategy has broad similarities to marine mammal "breath hold" diving.

As fish move from warm to cold water, metabolic heat generated in muscle tissue is carried away by the blood and is rapidly lost to the environment at the gills (*1*, *2*). Regionally endothermic fishes (those that warm part but not all of their body), such as lamnid sharks and tunas (families Lamnidae and Scombridae), retain heat by using vascular countercurrent heat exchangers to keep specific organs or tissues warm during deep dives into cold water (*3*–*7*). Although large body sizes can passively buffer the rate of change of body temperature, most fishes, including the scalloped hammerhead shark (*Sphyrna lewini*), are ectotherms that lack morphological and vascular adaptations to actively conserve heat (fig. S1).

Scalloped hammerhead sharks occupy warm surface waters in tropical and warm temperate oceans but make repeated nocturnal dives to depths exceeding 800 m, where temperatures are as low as 5°C (*8*–*11*). These dives provide access to the slow-moving mesopelagic and bathypelagic prey that dominate their diets (*12*–*18*). Such deep diving is risky because body cooling can reduce visual acuity, cardiac function, and swimming muscle power, which could be detrimental for obligate ram ventilators—fish that rely on their forward movement to force water across their gills for respiration—such as scalloped hammerhead sharks (*4*, *19*).

To determine how scalloped hammerhead sharks function in frigid water temperatures experienced during deep dives, we equipped adult individuals with instrument packages that measured depth, ambient water temperature, activity rates, and body orientation using triaxial accelerometers and measured core muscle temperature with a thermistor probe inserted ~8 cm into the dorsal musculature near the dorsal fin (table S1). Our specific objectives were to (i) characterize swimming performance during repetitive deep dives and (ii) measure core body temperature during dives to determine whether core body temperature is a result of simple thermal inertia or active thermoregulation [distinct from simply selecting optimal thermal environments to achieve optimal body temperature (*20*–*22*)].

Deep-diving behavior

We defined deep dives as those exceeding 400 m with starting and ending depths <100 m. All deep dives (n = 106) were conducted at night (Fig. 1 and figs. S2 to S4) and consisted of five distinctive phases (Fig. 2, figs. S5 and S6, and tables S2 and S3)—two with low activity and three with high activity.

Slow descent to ~110 m (duration 21.27 ± 22.7 min)

Beginning their dives at depths varying from the surface to 60 m, the sharks descended to ~110 m with a gradually decreasing pitch angle (i.e., pointing increasingly downward) from 0° to −15° with an average activity intensity [as measured by overall dynamic body acceleration (*23*)] of 0.7 ± 0.2 m/s^2, an average tailbeat frequency of 0.4 ± 0.1 Hz, and an average tailbeat sway amplitude of 0.3 ± 0.2 m/s^2.

Fast descent to bottom (duration 6.45 ± 1.82 min)

At ~110 m, the sharks abruptly oriented at a steep pitch angle between −70° and −80° and increased swimming intensity (with an average overall dynamic body acceleration of 1.3 ± 0.4 m/s^2). At depths between 300 and 500 m and while still at a steep dive angle, they conducted sporadic bursts of intense swimming activity lasting between 5 and 20 s, with overall dynamic body acceleration increasing by more than an order of magnitude (with peaks averaging 22.2 ± 12.0 m/s^2 and peaking as high as 49.0 m/s^2). The frequency of these burst events increased as the sharks approached the bottom of their dives.

Bottom time (duration 4.3 ± 3.02 min)

The sharks reached an average maximum depth of 635 ± 90 m with a range of 418 to 825 m. The bottom portions of the dives exhibited V-, U-, vU-, Uv-, and W-shaped depth profiles that are similar to those observed in deep-diving marine mammals such as elephant seals and other fish species that bounce dive, such as sicklefin devil rays (*Mobula tarapacana*) (*11*, *24*, *25*). Tagged sharks would either hold a steady depth within a 30-m range (V- and U-shaped dives) or move between depths varying by >100 m (vU-, Uv-, and W-shaped dives). Intense swimming activity occurred either in frequent bursts or was sustained throughout the duration of the bottom time (average overall dynamic body

Table 1. Results from representative whole-body heat coefficient model 4 for scalloped hammerheads HH7, HH8, and HH9. Parameter estimates include ±95% confidence intervals. k_1, whole-body heat transfer coefficient during normal swimming; k_2, whole-body heat transfer coefficient during the high-activity phase of a deep dive; $\dot{T}_{o1}$, rate of temperature change because of internal (metabolic) heat production during normal swimming; $\dot{T}_{o2}$, rate of temperature change because of internal (metabolic) heat production during the high-activity phase of a deep dive.

Shark	k_1 normal	k_2 diving	$\dot{T}_{o1}$ normal	$\dot{T}_{o2}$ diving
HH7	0.0175 ± 0.0026	0.0050 ± 0.0031	0.0040 ± 0.0024	0.0522 ± 0.0424
HH8	0.1129 ± 0.0046	0.0158 ± 0.0060	0.0339 ± 0.0022	0.1057 ± 0.0437
HH9	0.0311 ± 0.0007	0.0020 ± 0.0006	0.0052 ± 0.0003	0.0427 ± 0.0090

[1]Hawai'i Institute of Marine Biology, University of Hawai'i at Mānoa, Kāne'ohe, HI 96744, USA. [2]School of Physics and Astronomy, The University of Edinburgh, Edinburgh EH9 3FD, UK. [3]Department of Marine Sciences, Federal University of São Paulo, 11070-102 Santos, Brazil. [4]Ocean and Resources Engineering, University of Hawai'i at Mānoa, Honolulu, HI 96822, USA. [5]Biology Department, Whitman College, Walla Walla, WA 99362, USA.
*Corresponding author. Email: royerm@hawaii.edu

acceleration values of 1.9 ± 0.5 m/s^2 with bursts to 17.0 m/s^2). Tailbeat frequency during this stage of the dive was difficult to distinguish because of the erratic nature of the movement, as indicated by rapid fluctuations in pitch and roll angles.

Fast ascent to inflection point (duration 6.45 ± 1.9 min)

Ascents began with a steep, abrupt increase in pitch angle to 70° to 80° and sudden increases in swimming intensity (tailbeat frequency = 1.2 ± 0.6 Hz; tailbeat sway amplitude up to 4.5 m/s^2, averaging 2.0 ± 0.6 m/s^2; and overall dynamic body acceleration averaging 2.5 ± 0.3 m/s^2, with average peaks at 10.5 ± 6.6 m/s^2 and reaching as high as 41.1 m/s^2). Sharks sustained this intense swimming effort and steep pitch angle until they reached depths of between 350 and 250 m (average 290 ± 19 m). At this point, there was a consistent inflection in the profile characterized by an abrupt decrease in ascent rate associated with a change to a shallow pitch angle and reduction in swimming intensity (indicated by drops in overall dynamic body acceleration, tailbeat frequency, and tailbeat sway amplitude).

Slow ascent to upper 50 m (18.12 ± 3.55 min)

After the inflection point, the sharks continued their ascent at a slower rate to roughly 50-m depth. Swimming activity was greatly reduced compared with the fast ascent phase (overall dynamic body acceleration averaging 1.2 m/s^2, tailbeat frequency 0.5 Hz, and tailbeat sway amplitude 0.4 ± 0.2 m/s^2). During interdive intervals, sharks stayed within the top 50 m of the water column and resumed their characteristic nighttime side-swimming behavior of maintaining a rolled angle of ~60° during steady swimming (*26*) (fig. S5). Interdive intervals lasted an average of 43 ± 28 min and ranged from 18 min to >3 hours.

Overall, dives lasted an average of 56.0 ± 22.0 min (from the start of the slow descent to the end of the slow ascent). Bottom times (end of the fast descent to the start of the fast ascent) lasted an average of 4.0 ± 3.0 min. The average high-activity phase of the dives (from the start of the fast descent to the end of the fast ascent at the inflection point) lasted an average of 17.0 ± 3.5 min (table S3). Between deep dives, sharks spent most (98%) of their time between 175 m and the surface, with ambient water temperatures averaging 25.6°C (±1.6°C) and shark intramuscular temperatures averaging 26°C (±0.4°C) with a narrow variation (minimum 21.3°C, maximum 28.2°C) (table S4).

Body temperature profiles

At the bottom of deep dives, sharks experienced ambient temperatures of 5° to 11°C [mean 6.8°C (±1.2°C)]. Thus, during the fast descent phase of a dive, they experienced an ~20°C drop in water temperature over a period of 5 to 7 min (tables S3 and S4). However, the body temperature of the sharks did not change appreciably until rapidly cooling as the shark approached the surface waters during the final ascent stage of the dive (Fig. 2B and figs. S7 to S9).

During their fast descents into cold water, shark body temperatures initially cooled very slightly (0.1°C) but then either held constant or warmed again slightly (0.1° to 0.25°C) during the rest of the descent and through at least half of their bottom time. After an average of 4 min of bottom time and during the initial fast ascent, body temperatures began to drop slowly (0.03° ± 0.02°C/min). Near the inflection point of the ascent (290-m average depth), heat loss increased by an order of magnitude to 0.23° ± 0.15°C/min. Body temperatures continued to drop at this faster rate until the final phase of the dive, when sharks leveled off in the surface mixed layer (Fig. 2 and figs. S7 to S9). Thus, sharks experienced an average drop in body temperature of 2.8°C from the bottom of their dive to the start of their surface interdive interval, with the greatest heat loss occurring around the ascent inflection point (290-m average depth) that is concurrent with an abrupt decline in swimming intensity (Fig. 2, figs. S7 to S9, and tables S2 and S3).

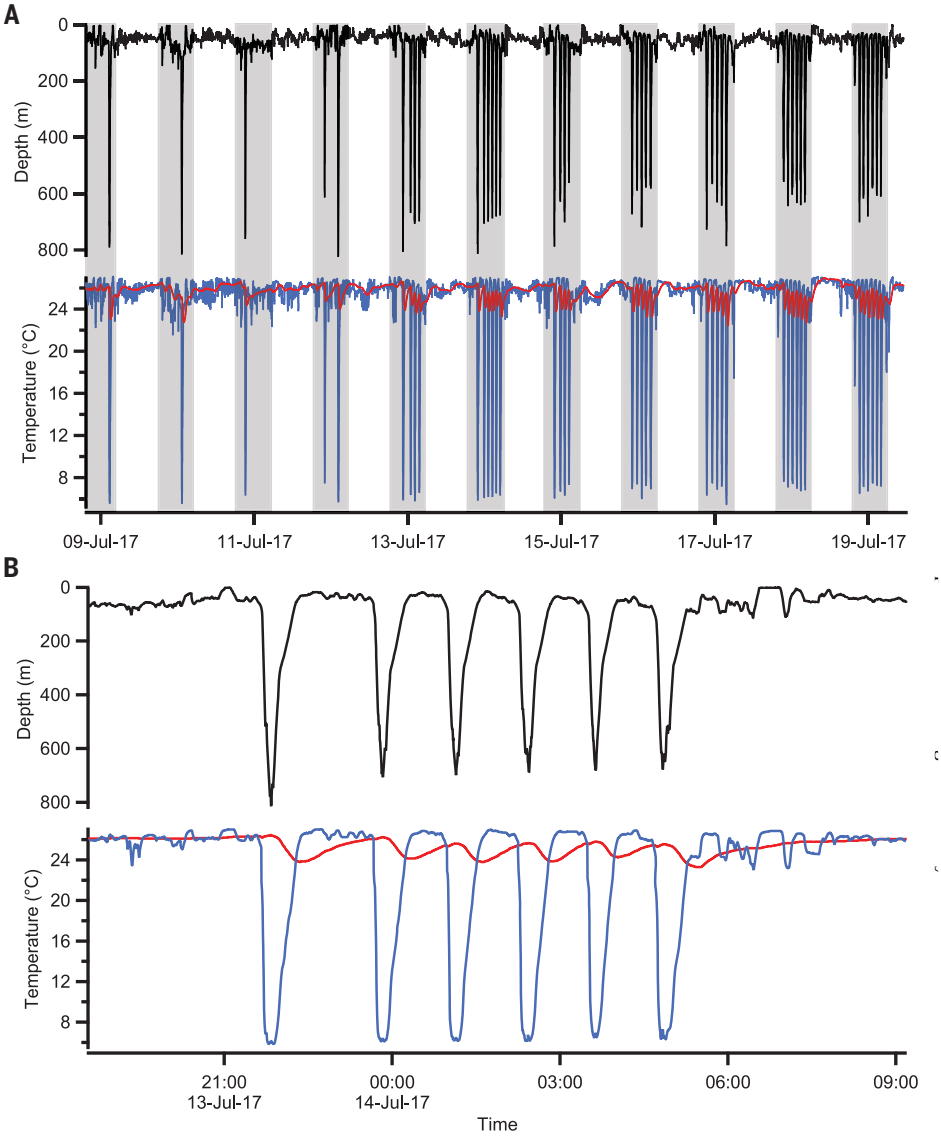


Fig. 1. Scalloped hammerhead shark diving behavior and temperature. (**A**) Eleven days of depth (top), ambient (blue), and intramuscular (red) temperature profiles from scalloped hammerhead shark HH7. All deep dives were conducted at night (gray shading). (**B**) Depth (top), ambient (blue), and intramuscular (red) temperature profiles from scalloped hammerhead shark HH7 during six deep dives conducted in a single evening.

Whole-body heat transfer coefficient (*k*) modeling

Heat exchange is proportional to the difference between body temperature and ambient water temperature

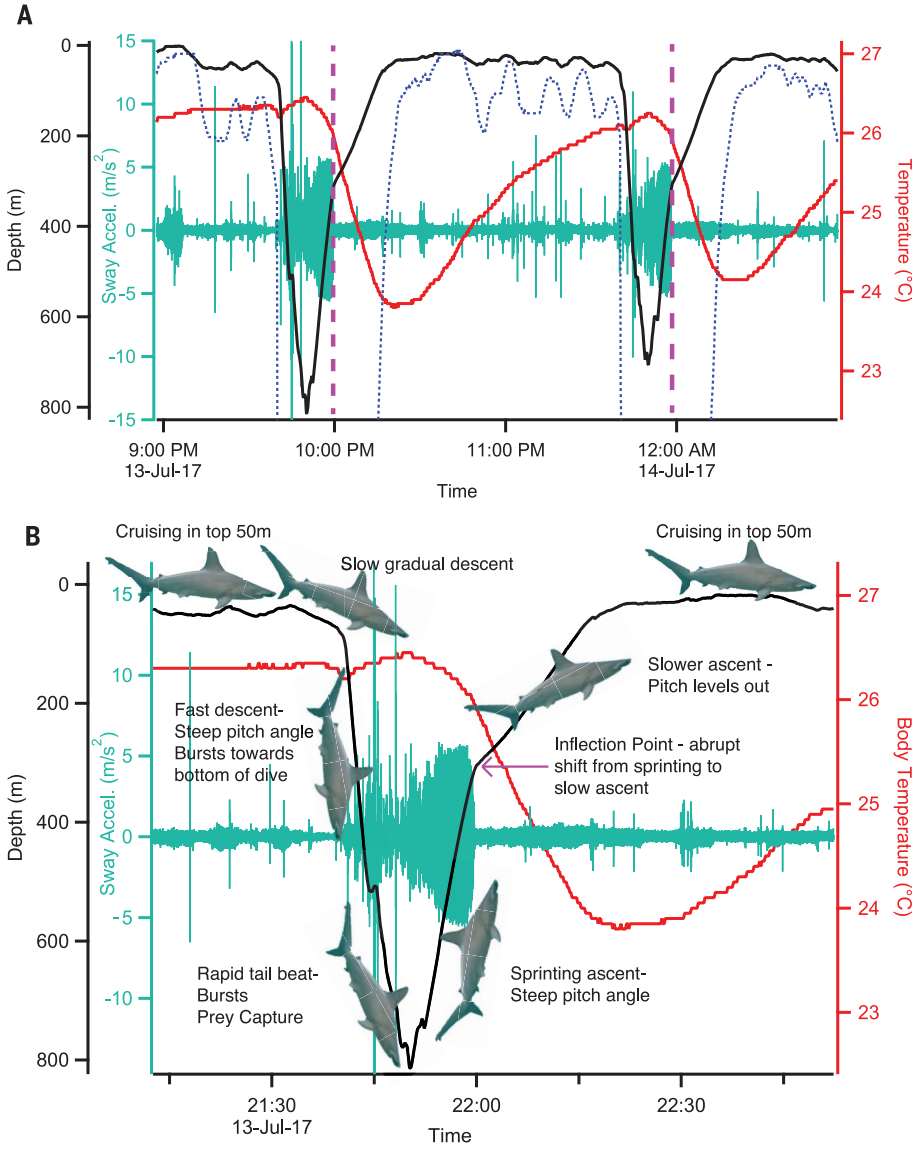


Fig. 2. Scalloped hammerhead shark swimming behavior and body temperature during deep dives. (**A**) Depth profiles (black), body temperature (red), ambient water temperature (blue dashed), and swimming activity [tailbeat sway acceleration (green)] during two successive deep dives by shark HH7. Dotted lines indicate the inflection point of the ascents when the swimming activity and pitch angle abruptly decrease. (**B**) Distinct phases of a deep dive using a representative dive. Shown are depth (black), intramuscular temperature (red), and tailbeat sway acceleration (green).

$$\frac{dT_b(t)}{dt} = k[T_a(t) - T_b(t)] + \dot{T}_o \qquad (1)$$

where k is the whole-body thermal rate coefficient (degrees Celsius per minute per degree Celsius), $T_a(t)$ is the ambient water temperature (degrees Celsius) as a function of time t, $T_b(t)$ is the core muscle temperature (degrees Celsius) as a function of time t, and $\dot{T}_o$ is the rate of temperature change because of metabolic heat production (degrees Celsius per minute) by the swimming muscles (*27*–*32*). The whole-body heat transfer coefficient (k) accounts for both convective heat transfer (i.e., heat exchange between blood flowing through the gills and ambient seawater) and conductive heat exchange across the body wall (*33*–*35*). Predicted values of the whole-body heat transfer coefficient (k) and rate of temperature change because of metabolic heat production ($\dot{T}_o$) were modeled to match the observed rates of body warming and cooling (*27*, *28*).

Modeling (k) for three sharks revealed that more than one k value is needed to accurately describe the rate of heat transfer required to produce body temperatures observed during deep dives. Model 4, which used two k and two $\dot{T}_o$ (heat production) values based on the different phases of the deep dives, was the most accurate in replicating the observed body temperatures (Table 1, table S5, Fig. 3, and figs. S10 and S11).

Values for k_2 estimated for the intense swimming phases (fast descent, bottom time, and fast ascent) of deep dives were an order of magnitude less than k_1 values estimated for the slow descent and slow ascent phases of the dive (Table 1). This indicates a substantially slower rate of heat transfer during the intense swimming phases of deep dives and suggests the existence of a mechanism enabling them to reduce heat loss and conserve body temperature until they reach the inflection point during ascent. The modeled metabolic heat production during the intense swimming phases of deep dives, $\dot{T}_{o2}$, was greater than the heat production rate during slow swimming, $\dot{T}_{o1}$, by a factor of ~3 to 13 (Table 1).

Scalloped hammerhead k_2:k_1 ratios were exceptional compared with those of other ectothermic sharks and fishes of similar body sizes [e.g., the blue shark (*Prionace glauca*)] and were comparable to those of regionally endothermic species, such as the bigeye tuna (*Thunnus obesus*) and swordfish (*Xiphias gladius*), both of which have specialized anatomical adaptations for conserving body heat (*32*) (fig. S12). The large differences in the heating and cooling coefficients of the scalloped hammerhead sharks in our study exceed those of the ectothermic ocean sunfish (*Mola mola*), which exhibits similarly large differences in these coefficients. This is surprising because the disk-like anatomy of the ocean sunfish is well suited for surface basking for rewarming, whereas scalloped hammerhead sharks have a more typical fusiform shape similar to that of blue sharks or makos (*Isurus oxyrynchus*). It is worth noting that although the coefficient differences and ratios of the scalloped hammerhead sharks are comparable to those of regionally endothermic species, scalloped hammerhead sharks lack the specialized anatomy characteristic of these species (*32*) (figs. S1 and S12).

The whole-body heat transfer coefficient was also estimated for two dead adult scalloped hammerheads that were obtained as freshly dead incidental fisheries mortalities (table S6). The single–k value model 1 was used because heat flux can only be attributed to conductive heat transfer across the body wall (there is no blood flow to the gills to support convective heat transfer), and physiological processes are absent. The whole-body heat transfer coefficients (k) from the dead sharks (0.0039° and 0.0038°C min^{-1} $°C^{-1}$) were similar (between 0.0011° and 0.0019°C min^{-1} $°C^{-1}$ difference for comparably sized sharks HH7 and HH9, 0.0121°C min^{-1} $°C^{-1}$ for the smaller HH8) to the k_2 values estimated for the tagged scalloped hammerheads during the intense phases of their deep dives using model 4 (Table 1 and

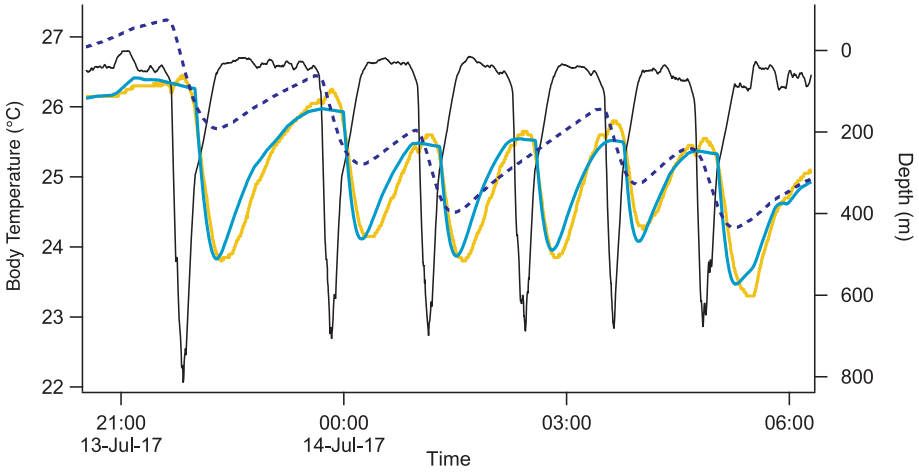


Fig. 3. Body temperature modeling across repetitive deep dives. Observed (yellow) and modeled intramuscular temperature during six deep dives from scalloped hammerhead shark HH7 with depth (black). Model 1 (blue dashed) assumes that the rate of heat exchange between the shark and the environment (k) and the rate of temperature change because of metabolic heat production ($\dot{T}_o$) is constant and not altered by the shark. Model 4 (cyan solid) assumes k_1 and $\dot{T}_{o1}$ when the shark is swimming normally and k_2 and $\dot{T}_{o2}$ during the high activity phases of a deep dive. The optimized parameters for each model were estimated on the basis of the coefficient of determination (R^2) between the observed and predicted body temperature. Models were run for each night that consisted of more than one deep dive for each shark. Start and stop times for each model duration were arbitrarily set to include all deep dives within a night duration and to include the durations where body temperature reaches the ambient temperature after the conclusion of a series of deep dives.

table S6). The similarity of the heat transfer coefficients of the dead sharks and the k_2 values of similarly sized live sharks indicates that scalloped hammerheads do not lose body heat through convective heat transfer at the gills during the intense swimming phases of their dives.

Discussion

If convective heat transfer occurs at the gills throughout the dive, body cooling should routinely begin and persist during the deeper stages of the dive; yet, this is not the case. Instead, these sharks maintained an elevated body temperature (up to 20°C above ambient) during most of each dive and only rapidly lost heat near the inflection point in the ascent. Sharp changes in body temperature at different points of the dive cannot be explained by simple passive buffering, and the apparent lack of convective heat transfer suggests that scalloped hammerhead sharks have an active mechanism to prevent heat loss from their gills during excursions into cold water. Possible mechanisms include shunting blood away from the gills or reducing the flow of water across the gills by reducing ram-ventilation by closing the mouth, the gill slits, or both. Either mechanism will inhibit the shark's ability to absorb oxygen from the environment. Essentially, the sharks seem to be holding their breath. We believe that the breath holding mechanism is the closing of the gill slits, and video evidence supports this contention. For example, remotely operated vehicle (ROV) footage (*36*) of an adult scalloped hammerhead shark swimming at a depth of 1043 m off Tanzania appeared to have its gill slits closed, whereas videos of scalloped hammerhead sharks in surface waters shows their gill slits open (figs. S13 and S14). No anatomical evidence of true vascular shunts at the gills has been observed in sharks (*37*); therefore, the most parsimonious explanation for the observed heat retention is closing of the gill slits. Further research is needed to confirm the gill closing hypothesis, for instance by attaching cameras to the pectoral fins of sharks that would capture the opening or closing of the gills. The faster decline in body temperature at or near the ascent inflection point (250- to 300-m depth) probably reflects reopening of the gill slits and the resumption of convective heat transfer. The interdive period might facilitate repayment of oxygen debt incurred from the recruitment of the white anaerobic musculature during the sprinting portion of each dive. This highly active swimming combined with apparent breath holding should be reflected in the activity levels of key muscle enzymes involved with locomotion and energy mobilization and should be examined further.

The physiological advantages conferred by maintaining a warm body temperature at depth enable scalloped hammerhead sharks to exploit resources in the cold mesopelagic depths lying below warm surface waters (*38*). These advantages are similar to those ascribed to so-called high-performance, regionally endothermic fishes, such as tuna and lamnid sharks, and include faster swimming capability because of enhanced muscle power output and enhanced enzyme and cardiac performance and neural processing (*39*). These benefits are consistent with our observations of burst swimming and high activity rates at the bottom of dives that may be feeding events. Feeding events are probably very brief. Hence, even if the sharks fleetingly open their gills while feeding at depth, it would be unlikely to be reflected in the core muscle temperature. For example, recorded feeding events at depth in blue sharks (*Prionace glauca*) (*32*) and ocean sunfish (*Mola mola*) (*30*) did not result in detectable changes in muscle temperature.

There are distinct physiological differences between scalloped hammerhead sharks and high-performance fishes. The latter maintain warm muscles and sensory organs during dives, but they are nevertheless affected by reductions in heart function because the heart is supplied by cold blood flowing from the gills (*40*–*45*). By reducing convective heat loss at the gills, scalloped hammerhead sharks maintain both muscle and heart temperature, thereby possibly preserving cardiac function during deep dives (*45*).

REFERENCES AND NOTES

1. E. D. Stevens, A. M. Sutterlin, *J. Exp. Biol.* **65**, 131–145 (1976).
2. F. G. Carey, Q. H. Gibson, *Physiol. Zool.* **60**, 138–148 (1987).
3. R. W. Brill, H. Dewar, J. B. Graham, *Environ. Biol. Fishes* **40**, 109–124 (1994).
4. B. A. Block, J. R. Finnerty, *Environ. Biol. Fishes* **40**, 283–302 (1994).
5. C. A. Sepulveda, S. Kohin, C. Chan, R. Vetter, J. B. Graham, *Mar. Biol.* **145**, 191–199 (2004).
6. J. C. Patterson, C. A. Sepulveda, D. Bernal, *J. Morphol.* **272**, 1353–1364 (2011).
7. D. Bernal, J. K. Carlson, K. J. Goldman, C. G. Lowe, in *Biology of Sharks and Their Relatives*, J. C. Carrier, J. A. Musick, M. R. Heithaus, Eds. (CRC Press, ed. 2, 2012), pp. 211–237.
8. S. J. Jorgensen, A. P. Klimley, A. F. Muhlia-Melo, *J. Fish Biol.* **74**, 1682–1687 (2009).
9. S. Bessudo *et al.*, *Bol. Investig. Mar. Costeras* **40**, 91–106 (2011).
10. E. R. Hoffmayer, J. S. Franks, W. B. Driggers, P. W. Howey, *Bull. Mar. Sci.* **89**, 551–557 (2013).
11. J. L. Y. Spaet, C. H. Lam, C. D. Braun, M. L. Berumen, *Anim. Biotelem.* **5**, 20 (2017).
12. B. A. Seibel, E. V. Thuesen, J. J. Childress, L. A. Gorodezky, *Biol. Bull.* **192**, 262–278 (1997).
13. J. C. Drazen, B. A. Seibel, *Limnol. Oceanogr.* **52**, 2306–2316 (2007).
14. T. A. Clarke, *Pac. Sci.* **25**, 133–144 (1971).
15. A. P. Klimley, D. R. Nelson, *Behav. Ecol. Sociobiol.* **15**, 45–54 (1984).
16. J. D. Stevens, J. M. Lyle, *Mar. Freshw. Res.* **40**, 46–129 (1989).
17. M. J. Smale, G. Cliff, *S. Afr. J. Mar. Sci.* **7615**, 241–253 (1998).
18. T. Vaske Júnior, C. M. Vooren, R. P. Lessa, *Braz. J. Oceanogr.* **57**, 97–104 (2009).
19. A. F. Bennett, *Am. J. Physiol.* **259**, R253–R258 (1990).
20. W. H. Neill, *Amer. Zool.* **19**, 305–317 (1979).
21. S. A. Matern, J. J. Cech, T. E. Hopkins, *Environ. Biol. Fishes* **58**, 173–182 (2000).
22. D. W. Sims *et al.*, *J. Anim. Ecol.* **75**, 176–190 (2006).
23. A. C. Gleiss, R. P. Wilson, E. L. C. Shepard, *Methods Ecol. Evol.* **2**, 23–33 (2011).
24. B. J. Le Boeuf, Y. Naito, T. Asaga, D. Crocker, D. P. Costa, *Can. J. Zool.* **70**, 786–795 (1992).
25. C. D. Braun *et al.*, *Annu. Rev. Mar. Sci.* **14**, 129–159 (2022).
26. M. Royer *et al.*, *Anim. Biotelem.* **8**, 11 (2020).

27. K. N. Holland, R. W. Brill, R. K. C. Chang, J. R. Sibert, D. A. Fournier, *Nature* **358**, 410–412 (1992).
28. K. N. Holland, J. R. Sibert, *Environ. Biol. Fishes* **40**, 319–327 (1994).
29. H. Dewar, J. B. Graham, R. W. Brill, *J. Exp. Biol.* **192**, 33–44 (1994).
30. I. Nakamura, Y. Goto, K. Sato, *J. Anim. Ecol.* **84**, 590–603 (2015).
31. A. Stoehr, J. St. Martin, S. Aalbers, C. Sepulveda, D. Bernal, *ICES J. Mar. Sci.* **75**, 858–870 (2018).
32. Y. Y. Watanabe, I. Nakamura, W. C. Chiang, *Mar. Biol.* **168**, 161 (2021).
33. W. H. Neill, E. D. Stevens, *Science* **184**, 1008–1010 (1974).
34. R. G. Fechhelm, W. H. Neill, *Physiol. Zool.* **55**, 229–239 (1982).
35. M. Pépino, K. Goyer, P. Magnan, *J. Exp. Biol.* **218**, 3461–3467 (2015).
36. A. B. M. Moore, A. R. Gates, *Mar. Biodivers. Rec.* **8**, 1–4 (2015).
37. N. C. Wegner, in *Physiology of Elasmobranch Fishes: Structure and Interaction with Environment*, R. E. Shadwick, A. P. Farrell, C. J. Brauner, Eds., vol. 34 of *Fish Physiology* (Academic Press, 2015), pp. 101–151.
38. J. B. Graham, K. A. Dickson, *Fish Physiol.* **19**, 121–165 (2001).
39. K. A. Dickson, J. B. Graham, *Physiol. Biochem. Zool.* **77**, 998–1018 (2004).
40. K. E. Korsmeyer, N. C. Lai, R. E. Shadwick, J. B. Graham, *J. Exp. Biol.* **200**, 1975–1986 (1997).
41. R. W. Brill, T. E. Lowe, K. L. Cousins, in *Cardiovascular Function in Fishes*, K. Gamperl, D. MacKinlay, A. Farrel, Eds. (American Fisheries Society, 1998), pp. 57–62.
42. J. M. Blank *et al*., *J. Exp. Biol.* **207**, 881–890 (2004).
43. K. C. Weng *et al*., *Science* **310**, 104–106 (2005).
44. H. A. Shiels, G. L. J. Galli, B. A. Block, *Proc. R. Soc. B* **282**, 20141989 (2015).
45. N. C. Wegner, O. E. Snodgrass, H. Dewar, J. R. Hyde, *Science* **348**, 786–789 (2015).
46. M. Royer *et al*., "Breath-holding" as a thermoregulation strategy in the scalloped hammerhead, a deep-diving tropical ectothermic shark, dataset, Dryad (2023); https://doi.org/10.5061/dryad.p8cz8w9v0.
47. M. Royer *et al*., "Breath-holding" as a thermoregulation strategy in the scalloped hammerhead, a deep-diving tropical ectothermic shark, Zenodo (2023); https://doi.org/10.5281/zenodo.7697788.

ACKNOWLEDGMENTS

We thank K. Carlson, K. Haiat, N. Hu, B. Rackliffe, J. Lazor, P. Mino, T. Clegg, M. Maxwell, and S. Rucker for assisting with fieldwork and J. Muir for providing crucial boat support for tag recovery. The lead author thanks C. Meyer, B. Bowen, A. Seale, and M. Yoshizawa for providing feedback on this project as dissertation committee members and K. Holland for being the chair of this committee. We thank L. Robinson, the Robinson family, and the people of Niʻihau for finding our tag package and J. TenBruggencate for coordinating communication and shipping of the recovered tag to us. We thank the anonymous reviewers for their insightful feedback. This is contribution no. 1918 from the Hawaiʻi Institute of Marine Biology and no. 11660 from the School of Ocean and Earth Science and Technology at the University of Hawaiʻi. Data for this study are available on Dryad (*46*). The Matlab code used to for thermal coefficient analysis is available on Zenodo (*47*). **Funding:** This study received support from the Jessie D. Kay Memorial Research Grant, UH Mānoa Department of Biology (M.R.); a Hawaiʻi Institute of Marine Biology Lord Endowed Scholarship (M.R.); a Robinson Family Ocean Studies Assistantship (M.R.); and Pacific Islands Ocean Observation Systems (K.N.H. and C.M.). **Ethics statement:** Scalloped hammerhead shark handling and tagging procedures were approved by the ethics committee at the University of Hawaiʻi (Institutional Animal Care and Use Committee Protocol no. 05-053). **Author contributions:** M.R. designed the study. M.R., K.M., E.C., K.W., C.B., and G.F.d.S. designed, constructed, and tested biologging packages. M.R. led fieldwork operations and attached biologging packages to all sharks. K.M., E.C., K.W., C.B., and G.F.d.S. assisted in fieldwork operations, including field preparation and planning, fishing effort, shark tagging, and package recovery. J.R. developed the Matlab code used for thermal coefficient modeling and analysis. C.M. provided detailed instruction for analyzing accelerometry data. M.R., K.M., E.C., G.F.d.S., and K.W. analyzed telemetry data. M.R. and J.R. conducted analyses on thermal coefficient modeling. M.R. drafted the figures. M.R. drafted the main body, and K.N.H. and C.M. substantially revised it. M.R. wrote the supplementary materials section. C.M. and K.N.H. provided substantial contributions to all written material and revisions during initial draft writing. All authors gave substantive feedback and suggested edits during the final drafting of the manuscript. M.R., C.M., and K.N.H. secured funding. **Competing interests:** The authors declare that there are no competing interests. **Data and materials availability:** The following data are available on Dryad (*46*): results of thermal coefficient modeling (Matlab code) for each night of deep diving for each shark; deep-diving telemetry metrics for each individual dive from each shark and summary statistics for each metric; and depth, ambient temperature, body temperature, and triaxial acceleration telemetry for each shark for each night of deep diving analyzed. Thermal coefficient modeling was performed using Matlab code that is available through Zenodo (*47*). Telemetry data were analyzed with Igor Pro 9 (Wavemetrics, Portland, OR, USA). Correspondence should be addressed to M.R.

License information: Copyright © 2023 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.add4445
Materials and Methods
Figs. S1 to S14
Tables S1 to S6
References (*48*–*55*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 5 July 2022; accepted 15 March 2023
10.1126/science.add4445

BLACK HOLE PHYSICS

Optical polarization from colliding stellar stream shocks in a tidal disruption event

I. Liodakis[1]*, K. I. I. Koljonen[1,2,3], D. Blinov[4,5], E. Lindfors[1], K. D. Alexander[6,7], T. Hovatta[1,2], M. Berton[1,2,8], A. Hajela[6,9], J. Jormanainen[1,10], K. Kouroumpatzakis[4,5,11], N. Mandarakas[4,5], K. Nilsson[1]

A tidal disruption event (TDE) occurs when a supermassive black hole rips apart a passing star. Part of the stellar material falls toward the black hole, forming an accretion disk that in some cases launches a relativistic jet. We performed optical polarimetry observations of a TDE, AT 2020mot. We find a peak linear polarization degree of 25 ± 4%, consistent with highly polarized synchrotron radiation, as is typically observed from relativistic jets. However, our radio observations, taken up to 8 months after the optical peak, do not detect the corresponding radio emission expected from a relativistic jet. We suggest that the linearly polarized optical emission instead arises from shocks that occur during accretion disk formation, as the stream of stellar material collides with itself.

A tidal disruption event (TDE) occurs when a star passes close enough to a supermassive black hole (BH) for the star's self-gravity to be overpowered by tidal forces produced by the BH's gravitational potential (*1*). As the star is torn apart, roughly half of the gas forms an accretion disk before falling into the BH. The disruption leads to an outburst of emission which peaks in the ultraviolet (UV) and soft x-rays, with a spectrum corresponding to a black body at temperature ~10^5 K. This is followed by a steady decline, regulated by the rate of gas fall toward the black hole, with emissions proportional to ~$t^{-5/3}$ (where t is the time since peak brightness) (*1*). Multiwavelength observations of TDEs have shown that some lead to the BH launching a relativistic jet of plasma (*2*). TDEs can therefore provide information on the early stages of formation for both accretion disks and relativistic jets.

Some TDEs, known as optical TDEs (*3*), have emissions that instead peak in the optical-UV range, having a lower black body temperature than typical TDEs. Optical TDEs are typically faint or undetected in radio and x-rays and do not show evidence of a jet. The emission profile of such events is not consistent with the standard theoretical model (*1*) of non-optical TDEs (also known as x-ray TDEs); several alternative models have been proposed to explain this divergence. Possible explanations include x-ray emission reprocessed to UV-optical wavelengths by surrounding gas (*4*), or emission powered by tidal shocks that form as a result of the star's gas stream colliding with itself as it flows around the black hole during accretion disk formation (known as the outer shock model) (*5*). There is observational evidence in favor of both models. For example, the brightness evolution and spectrum of the TDE AT 2018fyk have been interpreted as evidence for reprocessed emission (*6*). Similar conclusions have been reached for two other TDEs, AT 2018hyz and AT 2019qiz, based on their spectral line properties (*7*, *8*). In contrast, the late-time x-ray brightening of the TDEs ASASSN-15oi and AT 2019azh have been interpreted as due to the late accretion disk formation expected in the outer shock model (*9*, *10*).

The reprocessing and outer shock models make similar predictions for the total emission intensity, light curve, and late-time emission (e.g., x-ray brightening). However, their polarization properties are expected to differ substantially. Previous polarization studies have reported a single detection of ~7 to 8% polarized near-peak emission for TDEs with jets and ~1 to 2% for optical TDEs (*11*–*14*). Other observations have yielded only upper limits. Polarization measurements could potentially be used to determine the orientation of the accretion disk or the jet magnetic field.

Observations of AT 2020mot

We measured the optical polarization of AT 2020mot (also known as ZTF 20abfcszi or Gaia 20ead), an optical TDE identified in June 2020 that occurred in a nearby galaxy, at redshift z = 0.07 (*15*). It was classified as a TDE because it showed broad He II and hydrogen (Balmer) emission lines in the optical spectrum (*16*). We supplemented our polarization observations with archival radio, x-ray, photometric, and spectroscopic observations taken up to eight months after the optical peak (*17*). We also performed radio observations with the Karl G. Jansky Very Large Array (VLA) which did not detect the transient source, and no[t] archival x-ray observations by the Neil Gehrels Swift Observatory (Swift); both non-detections are not uncommon for an optical TDE. Near the optical peak [at Modified Julian Date (MJD) 59077.8] the 3 σ upper limit on radio emission at 15 GHz was < 27 microjansky (μJy).

Radio observations at lower frequencies (1 to 7 GHz), taken more than six months after the optical peak, detect a point source with a power-law spectrum. This is consistent with optically thin emissions and the non-detection at 15 GHz and indicates synchrotron emission from star formation in the host galaxy. The star formation rate (SFR) calculated from the radio observations differs by 2.3 σ from the star formation rate estimated using infrared data (*17*). During the decay phase, an x-ray source was detected with flux of 0.00160.0004 counts·s^{-1} from the direction of AT 2020mot (*17*). The Swift and XMM-Newton observations, taken ~6 to 8 months (MJDs 59270.5, 59306.4, and 59308.4) after the peak, did not find any increase in the x-ray emission (*17*). We therefore attribute all the x-ray emissions to a contaminating source, likely a background active galactic nucleus (AGN). The measured UV and optical fluxes, six to eight months after the optical peak, do not indicate any elevated ultraviolet activity above the background level from the host galaxy. Additional spectroscopic observations using the Nordic Optical Telescope (NOT), taken about six months after the outburst peak (MJD 59264), did not detect any optical emission lines associated with the TDE (*17*). All these observed properties are consistent with AT 2020mot being a typical optical TDE (*17*).

We began optical polarization observations of AT 2020mot ~25 days after the outburst peak, using the RoboPol polarimeter mounted on the 1.3-m telescope of the Skinakas Observatory, Crete (*18*). We used RoboPol to observe the source 3 times, separated by ~20 days each. We performed a fourth polarization observation using NOT on MJD 59164, which was before the *R*-band optical flux reached the host galaxy level. Figure 1 shows the *r*-band light curve from the Zwicky Transient Facility (ZTF) (*19*), and our measured evolution of the polarization degree (Π) and the polarization angle (PA) (*17*). We find a low polarization degree (Π ~2%) for the first observation, similar to the values previously reported from other optical TDEs (*12*). The second observation showed a higher level of polarization (Π ~7%) similar to TDEs with a jet, accompanied by a change in the PA. The last two observations showed a decline in Π with a stable PA (within the uncertainties). After removing the unpolarized contribution of the host galaxy, we calculate that the source reached an intrinsic Π = 25 ± 4% at the time of the second observation (*17*).

[1]Finnish Center for Astronomy with the European Southern Observatory, University of Turku, FI-20014 Turku, Finland. [2]Metsähovi Radio Observatory, Aalto University, FI-02540 Kylmälä, Finland. [3]Institutt for Fysikk, Norwegian University of Science and Technology, 7491 Trondheim, Norway. [4]Institute of Astrophysics, Foundation for Research and Technology - Hellas, GR-71110 Heraklion, Greece. [5]Department of Physics, University of Crete, GR-70013 Heraklion, Greece. [6]Center for Interdisciplinary Exploration and Research in Astrophysics, Northwestern University, Evanston, IL 60208, USA. [7]Steward Observatory, University of Arizona, Tucson, AZ 85721-0065, USA. [8]European Southern Observatory, Santiago, 19001, Chile. [9]Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, 2200 Copenhagen, Denmark. [10]Department of Physics and Astronomy, University of Turku, FI-20014 Turku, Finland. [11]Astronomical Institute of the Czech Academy of Sciences, CZ-14131 Prague, Czech Republic.
*Corresponding author. Email: yannis.liodakis@utu.fi

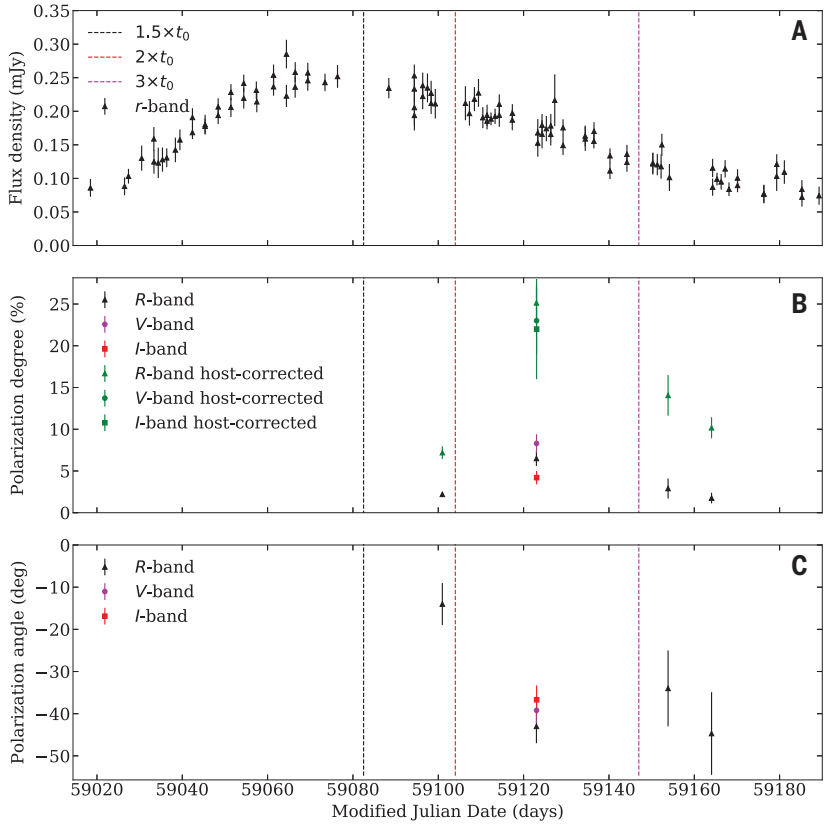


Fig. 1. Optical light curve and polarization measurements of AT 2020mot. (**A**) *r*-band brightness [black triangles, data from (*19*)], showing the temporal evolution. (**B**) Observed polarization degree in the *V*-band (purple circles), *R*-band (black triangles), and *I*-band (red squares). Green symbols indicate the values after correction for the host galaxy, in the *V*-band (green circle), *R*-band (green triangles), and *I*-band (green square). (**C**) Same as (B) but for the polarization angle. In all panels, error bars indicate the 68% confidence interval, and the vertical dashed lines show the characteristic times calculated using the outer shock model for 1.5 × t_0 (black), 2 × t_0 (red), and 3 × t_0 (magenta). We infer that the disruption happened before the rise of the *r*-band light curve, at <59026 MJD.

Interpretation of the optical polarization

The high polarization degree observed in AT 2020mot is most likely to have been produced by synchrotron radiation. We consider whether this could be due to the formation of a relativistic jet, which could explain both the observed polarization level and its variability. The lack of both radio and x-ray detections near the peak of the TDE allows us to reject a collimated jet viewed at a small angle to the line of sight. This is consistent with the observed rarity of TDEs with jets (*2*). On-axis gamma-ray burst jets—the only other transient short-lived jet source—have been observed to have high Π near the γ-ray peak, which steadily declines over time with either a stable PA or exhibiting 90° jumps (*20*–*22*), which we do not observe. It is possible that an off-axis jet could produce radio and x-ray emission that IS delayed with respect to the outburst peak (*23*). However, we would then expect that the jet opening angle would increase over time (*24*), causing a late-time brightening in both radio and x-ray emission. There was no such increase in brightness at the time of our observations. An off-axis jet would not contribute substantially to the emissions at early times, so could not produce the observed Π = 25% even if the jet was maximally polarized.

Scattering from an accretion disk can polarize the optical emission. The theoretical maximum degree of polarization that can be produced by scattering is Π = 11.7% for an exactly edge-on disk, and rapidly declines toward a pole-on oriented disk (*25*). In a disk scattering scenario, we would expect the PA to remain at a constant value, set by the scattering geometry. Both the higher value of Π and the PA variability lead us to reject accretion disk scattering as the origin of the polarized emission.

In an alternative reprocessing scenario, polarized emissions would arise from radiation emitted by an accretion disk that is scattered by the optically thick obscuring material left behind from the disrupted star. In this case, a clumpy medium could cause polarization angle variability, but the polarization degree is expected to be <10% (*26*). Reprocessing models often invoke an accretion disk wind for the obscuring material. The quasi-spherical geometry of the disk wind is expected to produce low polarization as a result of symmetry. The disk wind is also expected to produce detectable radio emissions (*27*), which we do not find. Nor was there any UV or x-ray enhancement or spectral lines associated with reprocessing. We therefore reject the reprocessing scenario as the origin of the optical emission from AT 2020mot.

Another possibility is stellar stream shocks (*5*, *28*). In this scenario, the gas from the disrupted star does not immediately form an accretion disk, as in the standard TDE model (*1*). Instead, the stellar stream flowing around the BH collides with itself. The collision forms shocks that dissipate some of the energy in the flow of stellar material, reducing the eccentricity of the stream (shifting it closer to the circular orbits of an accretion disk). The energy dissipated could account for all the optical emissions.

Colliding stellar streams model

Figure 2 shows a schematic of our proposed scenario for AT 2020mot. The UV and optical spectral energy distributions of optical TDEs are usually consistent with a nearly constant temperature black body model (*29*). However, we found that at some times a power-law model provides a better fit to the spectral energy distributions of AT 2020mot. This indicates that the UV-optical emission arises from either a modified black body or a combination of several emission components. To obtain an estimate for the bolometric luminosity of AT 2020mot at the peak of the outburst, we fitted the UV-optical spectrum of the peak with a black body model, finding a luminosity of $(1 \pm 0.2) \times 10^{44}$ erg × s^{-1} and a temperature of 21,800 ± 1600 K (*17*).

Applying these values to a previously-published outer shock model (*30*), we estimate a BH mass of 3.6×10^6 solar masses ($M_\odot$), disrupted star mass $M_* = 1.3 M_\odot$, and orbital timescale of the innermost material $t_0 \approx 43$ days. Figure 1 indicates the model's predicted peak position at 1.5 × t_0, which is consistent with the observed light curve. In the outer shock model, the emission during the initial rise is dominated by the shock formed at the pericenter (also known as the nozzle shock, labeled shock 1 in Fig. 2A) of the stellar stream orbit (*28*). At the outburst peak (Fig. 2B), the reverse and forward shocks (the outer shocks, labeled shocks 2 and 3) have formed. By 2 × t_0, the outer shocks account for a larger fraction of the optical emission, which coincides with our measurement of the highest Π value and the change in PA (Fig. 1).

The degree of polarization produced by each shock depends on the shock compression and its orientation with respect to the line of sight (*17*). The observed polarization behavior can then be produced as the sum of multiple shocks with differing PAs and varying contributions to the total emission. Changes in orientation of the shocks, as well as changes in their relative contributions, could produce large variations in Π, similar to the behavior observed due to the presence of multiple shocks in AGN jets (*31*).

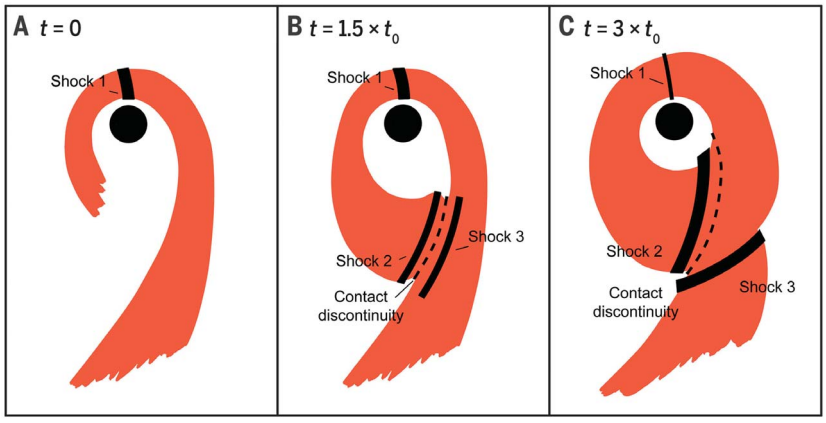


Fig. 2. Schematic diagram of the qualitative tidal stream shock model. Each panel shows a different time. The black circle indicates the supermassive black hole, the red shading is the stream of material from the disrupted star, and black lines indicate the locations of shocks. (**A**) Shock 1 forms after the stellar disruption and dominates the optical emission during the rise of the light curve. (**B**) By $1.5 \times t_0$, two additional shocks have formed, and the light curve reaches its peak brightness (Fig. 1A). As time passes, shocks 2 and 3 account for increasing fractions of the emission. (**C**) By $3 \times t_0$ shocks 2 and 3 dominate the total emission. Between $1.5 \times t_0$ and $3 \times t_0$, the changing shock orientation and relative contributions to the optical emission cause the polarization degree and polarization angle to vary.

By $3 \times t_0$ (Fig. 2C), the emission is dominated by the outer shocks. Subsequent weakening of the shocks, an increase in turbulence, or the breakup of shock 2 could then produce the observed decrease in the observed Π. The evolution of the optical-to-UV spectrum (*17*) is also consistent with a varying contribution to the emissions from different shocks.

The match between this model and the observations of AT 2020mot lead us to favor tidal stream shocks as the origin of the optical emission. In this scenario, the lack of late time UV or x-ray brightening can be interpreted as demonstrating that the system had not formed an accretion disk eight months after the optical peak. However, a weak accretion disk emitting at a lower flux than the host galaxy cannot be excluded by the observations. AT 2020mot is similar to other optical TDEs, so we suggest that the tidal stream shock model might explain the difference between optical and x-ray TDEs.

REFERENCES

1. M. J. Rees, *Nature* **333**, 523–528 (1988).
2. K. D. Alexander, S. van Velzen, A. Horesh, B. A. Zauderer, *Space Sci. Rev.* **216**, 81 (2020).
3. S. van Velzen, T. W. S. Holoien, F. Onori, T. Hung, I. Arcavi, *Space Sci. Rev.* **216**, 124 (2020).
4. B. D. Metzger, N. C. Stone, *Mon. Not. R. Astron. Soc.* **461**, 948–966 (2016).
5. H. Shiokawa, J. H. Krolik, R. M. Cheng, T. Piran, S. C. Noble, *Astrophys. J.* **804**, 85 (2015).
6. T. Wevers *et al.*, *Mon. Not. R. Astron. Soc.* **488**, 4816–4830 (2019).
7. T. Hung *et al.*, *Astrophys. J.* **903**, 31 (2020).
8. M. Nicholl *et al.*, *Mon. Not. R. Astron. Soc.* **499**, 482–504 (2020).
9. S. Gezari, S. B. Cenko, I. Arcavi, *Astrophys. J. Lett.* **851**, L47 (2017).
10. X.-L. Liu, L.-M. Dou, R.-F. Shen, J.-H. Chen, *Astrophys. J.* **925**, 67 (2022).
11. K. Wiersema *et al.*, *Mon. Not. R. Astron. Soc.* **421**, 1942–1948 (2012).
12. A. B. Higgins *et al.*, *Mon. Not. R. Astron. Soc.* **482**, 5023–5040 (2019).
13. C.-H. Lee *et al.*, *Astrophys. J. Lett.* **892**, L1 (2020).
14. K. Wiersema *et al.*, *Mon. Not. R. Astron. Soc.* **491**, 1771–1776 (2020).
15. E. C. Bellm *et al.*, *Publ. Astron. Soc. Pac.* **131**, 018002 (2019).
16. G. Hosseinzadeh *et al.*, Global SN Project Transient Classification Report for 2020-08-13 (Global SN Project, Report No. 2020-2478, 2020).
17. Materials and methods are available as supplementary materials.
18. A. N. Ramaprakash *et al.*, *Mon. Not. R. Astron. Soc.* **485**, 2355–2366 (2019).
19. K. W. Smith *et al.*, *Res. Notes AAS* **3**, 26 (2019).
20. R. Sari, *Astrophys. J.* **524**, L43–L46 (1999).
21. C. G. Mundell *et al.*, *Nature* **504**, 119–121 (2013).
22. K. Wiersema *et al.*, *Nature* **509**, 201–204 (2014).
23. S. Mattila *et al.*, *Science* **361**, 482–485 (2018).
24. T. Piran, *Rev. Mod. Phys.* **76**, 4 (2004).
25. S. Chandrasekhar, *Radiative Transfer* (Dover Publications, 1960).
26. F. Marin, M. Stalevski, in *SF2A-2015: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*, F. Martins, S. Boissier, V. Buat, L. Cambrésy, P. Petit, Eds. pp.167–170 (2015).
27. T. Matsumoto, T. Piran, *Mon. Not. R. Astron. Soc.* **507**, 4196–4210 (2021).
28. T. Piran, G. Svirski, J. Krolik, R. M. Cheng, H. Shiokawa, *Astrophys. J.* **806**, 164 (2015).
29. T. Hung *et al.*, *Astrophys. J.* **842**, 29 (2017).
30. T. Ryu, J. Krolik, T. Piran, *Astrophys. J.* **904**, 73 (2020).
31. I. Liodakis *et al.*, *Astrophys. J.* **902**, 61 (2020).
32. F. Bonnarel *et al.*, *Astron.Astrophys.* **143**, 1–7 (2000).
33. T. Boch, P. Fernique, in *Astronomical Data Analysis Software and Systems XXIII*, N. Manset, P. Forshay, eds. (2014), vol. 485 of *Astronomical Society of the Pacific Conference Series*, p. 277.
34. D. Blinov, N. Mandarakas, Optical polarimetric observations of the TDE AT 2020mot with RoboPol, Version 2, Zenodo (2023); https://doi.org/10.5281/zenodo.7716054.
35. E. Lindfors, J. Jormanainen, K. Nilsson, M. Berton, Optical polarimetric, photometric and spectroscopic observations of the TDE AT 2020mot with Nordic Optical Telescope, Version 1, Zenodo (2023); https://zenodo.org/record/7744446#.ZD2UYnbMI2w.

ACKNOWLEDGMENTS

We thank N. Globus, N. Kylafis, T. Piran, and R. Romani for comments and suggestions that helped improve this work. I.L. thanks the University of Crete and Institute of Astrophysics for their hospitality while this paper was written. Based on observations obtained with XMM-Newton, an ESA science mission with instruments and contributions directly funded by ESA Member States and the National Aeronautics and Space Administration (NASA). Lasair is supported by the UKRI Science and Technology Facilities Council and is a collaboration between the University of Edinburgh (grant ST/N002512/1) and Queen's University Belfast (grant ST/N002520/1) within the LSST: UK Science Consortium. Z.T.F. is supported by National Science Foundation grant AST-1440341 and a collaboration including Caltech, IPAC, the Weizmann Institute for Science, the Oskar Klein Center at Stockholm University, the University of Maryland, the University of Washington, Deutsches Elektronen-Synchrotron and Humboldt University, Los Alamos National Laboratories, the TANGO Consortium of Taiwan, the University of Wisconsin at Milwaukee, and Lawrence Berkeley National Laboratories. Operations are conducted by COO, IPAC, and UW. This research has made use of the Aladin sky atlas developed at CDS, Strasbourg Observatory, France (*32*, *33*). We acknowledge the use of public data from the Swift data archive. Based on observations made with the Nordic Optical Telescope, owned in collaboration by the University of Turku and Aarhus University, and operated jointly by Aarhus University, the University of Turku and the University of Oslo, representing Denmark, Finland, and Norway, the University of Iceland and Stockholm University at the Observatorio del Roque de los Muchachos, La Palma, Spain, of the Instituto de Astrofisica de Canarias. Some of the Nordic Optical Telescope observations were performed by the 2020 Nordic Optical Telescope observing school for upper secondary school students, organised by the Science Centre of the University of Turku and the Finnish Centre for Astronomy with ESO, University of Turku. This research has made use of data and/or software provided by the High Energy Astrophysics Science Archive Research Center (HEASARC), which is a service of the Astrophysics Science Division at NASA/GSFC. The Pan-STARRS1 Surveys (PS1) and the PS1 public science archive have been made possible through contributions by the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max-Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, the Queen's University Belfast, the Harvard-Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the National Central University of Taiwan, the Space Telescope Science Institute, the National Aeronautics and Space Administration under grant NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation grant AST-1238877, the University of Maryland, Eotvos Lorand University (ELTE), the Los Alamos National Laboratory, and the Gordon and Betty Moore Foundation. This publication makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/ California Institute of Technology, funded by NASA and the National Science Foundation. **Funding:** E.L. was supported by Academy of Finland projects 317636 and 320045. K.I.I.K. was supported by Academy of Finland project 320045 and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 101002352). T.H. was supported by Academy of Finland projects 317383, 320085, and 322535. J.J. was supported by Academy of Finland project 320085. D.B. and N.M. acknowledge funding from the ERC under the European Unions Horizon 2020 research and innovation programme under grant agreement 771282. K.D.A. was supported by NASA through the NASA Hubble Fellowship grant HST-HF2-51403.001-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555. Funding to support our NOT observations was provided by the Finnish Centre for Astronomy with ESO, under Academy of Finland grant 306531. **Author contributions:** D.B., T.H., J.J., E.L., I.L., and N.M. performed the polarization observations, data analysis and modeling. K.I.I.K. performed the UV and x-ray analysis. K.D.A. and A.H. performed the radio observations and data reduction. M.B., E.L., K.K.. and K.N. performed the host-galaxy characterization, optical spectroscopy, and star-formation rate estimation. K.K. performed the spectral energy distribution fitting. I.L., K.I.I.K. and K.N. estimated the black hole mass. All authors contributed writing the paper and the interpretation of the results. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Our RoboPol and NOT observations are available on Zenodo (*34*, *35*). Our radio observations are available at https://data.nrao.edu/ under project code '20A-372'. The archival optical photometry is available at https://lasair-ztf.lsst.ac.uk/objects/ZTF20abfcszi/, and the archival UV & x-ray data at https://heasarc.gsfc.nasa.gov/cgi-bin/W3Browse/w3browse.pl by searching for object name 'AT 2020mot' and setting the mission to 'Swift' and 'XMM-Newton'.

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj9570
Materials and Methods
Figs. S1 to S8
Tables S1 to S4
References (*36*–*82*)

Submitted 12 June 2021; accepted 3 April 2023
10.1126/science.abj9570

TECHNICAL COMMENT

EVOLUTION

Comment on "Exceptional preservation of organs in Devonian placoderms from the Gogo lagerstätte"

Bjarke Jensen[1]*, Antoon Fransiscus Maria Moorman[1], Tobias Wang[2], Peter Rask Møller[3,4], José Manuel Icardo[5], Henrik Lauridsen[6]

Trinajstic et al., (*Science*, 16 September 2022, p. 1311–1314) describe exceptionally well-preserved organs in fossilized Devonian placoderms to infer the early evolution of the vertebrate heart. We argue that the report has numerous shortcomings and examples of mixed specimen codes. Further, we question whether there indeed is any evidence for a mineralized chambered heart in these placoderms.

Trinajstic and colleagues recently reported on exceptional preservation of organs in Devonian placoderms from the Gogo lagerstätte (*1*). The specimens are exquisite and the generated datasets can be of great value to future studies of the heart if their presentation is accurate. As the main finding, the authors describe that the atrium is positioned dorsally to the ventricle as in more recent lineages of fish. However, the atrium is actually ventral to the ventricle in the one specimen where the dorsoventral axis is unambiguous (Fig. 1, A to C) (nominally specimen MV P230859 but actually WAM 18.4.2, see below). The alleged outflow tract shows an unprecedented sharp left-ward bend (Fig. 1B). Proportionally, the heart is extremely small, even considering the possibility of soft tissue shrinkage during fossilization (*1*), and it is difficult to envision how it could have supported the metabolic demands of a living placoderm. We argue that the described heart does not resemble any known vertebrate heart.

The morphology of the cardiac chambers of specimen WAM 2020.2.1 is also inconsistent with current knowledge. Fossilized fish heart chambers have previously been shown to have the thin-walled atrium and thick-walled ventricle that is well established in hearts of all extant vertebrates, from hagfish to humans (Fig. 1, D and E) (*2*–*4*). Accordingly, the thin-walled atrium only has one-sixth the mass of the ventricle (*4*) which reflects it only has to fill the relaxing ventricular cavity, whereas the ventricle has to provide enough pressure to circulate blood through the entire organism. Yet, the cavities of WAM 2020.2.1 do not show the expected differences in wall thickness (Fig. 1, F to I). From the publicly available image stack, we reproduced the labelling of the alleged atrium and ventricle (Fig. 1, J to L). We argue that the walls surround a single cavity, labelled in orthogonal planes, meaning there is no evidence for two chambers (Fig. [illegible]). How the dorsoventral axis was established is not clear.

We believe that the correct official acronym for Museums Victoria specimens is NMV and not MV as used in (*1*). This is an important detail if specimens are to be reexamined. We could not reproduce the published views of the alleged heart of specimen "MV P230859". The authors could then clarify the heart is in a different specimen (WAM 18.4.2) in which, when the image stack became available, we could reproduce the published views. The contents of this specimen, however, has no description (*1*). Also, we believe that the spiral intestine of two specimens of apparently different sizes are documented by photographs of what is likely the part-counterpart pair of the same fossil (Fig. 2). One specimen is extraordinarily large compared to previously reported specimens of the same species (*5*), if the scale bar is correct that relates to its atrium (Fig. 1F). Analyses of ontogenetic changes could be severely hampered if the published scales are inaccurate.

In summary, for the main finding of the paper to stand, it requires 1) the presence of fossilized heart, 2) distinct chambers, and 3) a dorsally positioned atrium. We can neither confirm nor disprove whether there are any fossilized remains of the heart, but criteria 2) and 3) have not been met in the paper in its current state. Exquisite fossils have been found (*1*), but their full value can only fully be exploited if major corrections are made to their presentation.

[1]Department of Medical Biology, Amsterdam Cardiovascular Sciences, Amsterdam University Medical Centres, Amsterdam, The Netherlands. [2]Department of Biology, Zoophysiology, Aarhus University, Aarhus, Denmark. [3]Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark. [4]Norwegian College of Fishery Science, UiT - The Arctic University of Norway, Tromsø, Norway. [5]Department of Anatomy and Cell Biology, University of Cantabria, Santander, Spain. [6]Department of Clinical Medicine, Aarhus University, Aarhus N, Denmark.
*Corresponding author. Email:

REFERENCES AND NOTES

1. K. Trinajstic *et al.*, *Science* **377**, 1311–1314 (2022).
2. L. Maldanis *et al.*, *eLife* **5**, e14698 (2016).
3. J. M. Icardo, *Fish Physiol.* **36**, 1–54 (2017).
4. B. Jensen *et al.*, *PLOS ONE* **8**, e63651 (2013).
5. M. Rücklin *et al.*, *Nature* **491**, 748–751 (2012).
6. F. Tessadori *et al.*, *PLOS ONE* **7**, e47644 (2012).

Submitted 14 December 2022; accepted 28 February 2023
10.1126/science.adg2748

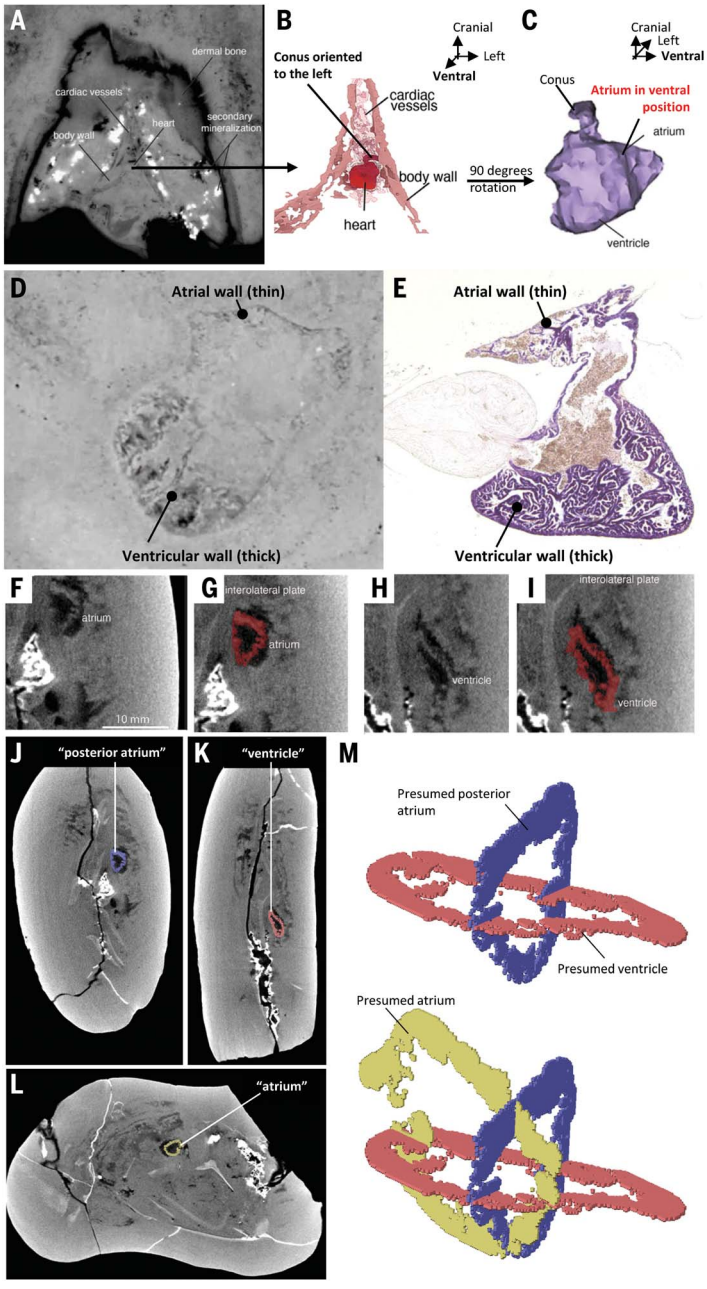


Fig. 1. (**A**) X-ray of nominally specimen MV P230859 but actually WAM 18.4.2 [from figure S4K of (*1*)]. (**B**) Adaptation of figure 1G of (*1*) showing the reconstructed heart in ventral view. (**C**) Lateral view of the reconstructed heart, showing the atrium is ventral. Adapted from figure S4H of (*1*), the label "atrium" is from the original figure. (**D** and **E**) X-ray of fossilized fish heart (D), adapted from (*2*), and histological section of the zebrafish heart (E), from (*6*). (**F** to **I**) Heart chamber labelling of specimen WAM 2020.2.1 from (*1*). (**J** to **L**) Our reproduction of heart chamber labelling of (*1*), specifically their figure 1B (blue), figure 1D (red), and figure S4B (yellow). Image viewing and segmentation was performed in Amira (version 2020.2, Thermo Scientific). (**M**) The three labels surround the same cavity.

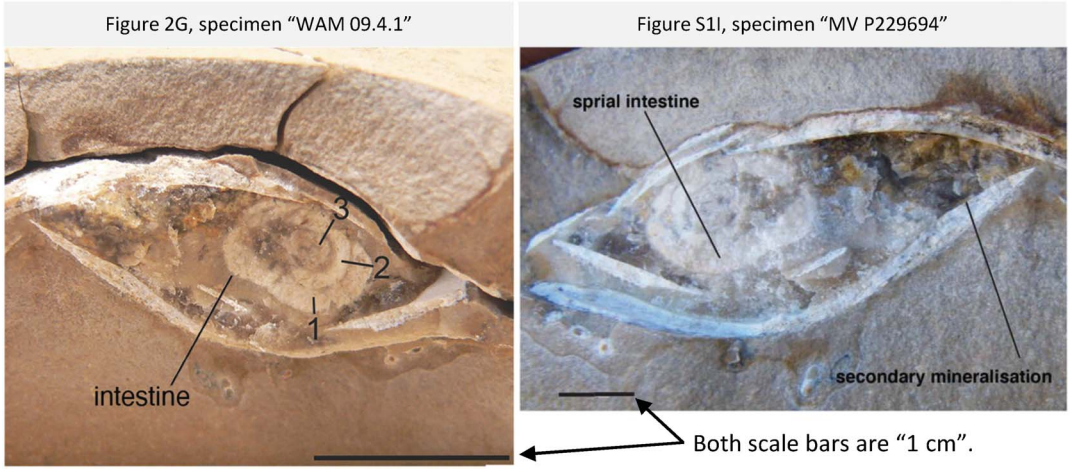


Fig. 2. Two specimens are documented with what we argue is part-counterpart of the same fossil and with very different scale bars. Images reproduced from (*1*).

TECHNICAL RESPONSE

EVOLUTION

Response to comment on "Exceptional preservation of organs in Devonian placoderms from the Gogo largerstätte"

Kate Trinajstic[1,2]*, John Long[3,4], Sophie Sanchez[5,6], Catherine A. Boisvert[1], Daniel Snitting[5], Paul Tafforeau[6], Vincent Dupret[5], Alice M. Clement[3], Peter D. Currie[7], Brett Roelofs[1], Joseph J. Bevitt[8], Michael S.Y. Lee[3,9], Per E. Ahlberg[5]

Jensen *et al*. (*1*) question evidence presented of a chambered heart within placoderms, citing its small size and apparently ventral atrium. However, they fail to note the belly-up orientation of the placoderm within one nodule, and the variability of heart morphology within extant taxa. Thus, we remain confident in our interpretation of the mineralized organ as the heart.

Arthrodires conform to the same patterns of anatomy as living gnathostomes, particularly in their general bauplan and tissue development (*2*, *3*). However, like all fossil soft tissues, the organs of these placoderms (*2*) are modified by taphonomy (*4*–*7*). They are emphatically not "fossil dissections" that can straightforwardly be compared with fresh animal specimens (*5*). To take an obvious example, the best-preserved liver is fractured into two parts and the associated stomach is ruptured ventrally [figure 2, C and I, in (*2*)].

As for the heart, we begin by noting that it lies exactly where we would expect a heart [figure 1F in (*2*)]—and where we would not expect any other organ (*3*, *8*). Furthermore, it is the right size and shape for a heart and contains an internal cavity [figure 1, A to D, in (*2*)]. Jensen *et al*. (*1*) claim that the atrium is ventral to the ventricle but this is not true. WAM 20.2.1 is lying belly-up within the sediment, as dead fish often do; when the arthrodire is orientated to life position the atrium is dorsal. As with our fossils (*2*), *Rhacolepis buccalis* (*9*) shows imperfect alignment, and compression of the heart in relation to the body axis. We therefore regard our identification as secure; there is no plausible alternative to it being the heart.

All fossils with soft tissue preservation show evidence of decomposition which in fact is required for preservation (*2*, *4*–*7*). However the amount of decay and associated morphological loss is variable (*4*–*9*). Thus, contra to Jensen *et al*. (*1*) the preservation of anatomical components in one specimen (*9*) cannot be taken as evidence that the same feature is preserved in all specimens. They (*1*) are incorrect in stating cardiac chamber morphology is inconsistent with current knowledge. Although the chambers of the heart are visible in WAM 2020.2.1 decay, causing muscle slippage, precludes determining the wall thickness or chamber volume (*2*). In WAM 18.4.2 (mislabeled as MV P230859), no internal structure of the heart is preserved [figure S4, G to H, in (*2*)]. Therefore, the information is unknown and not inconsistent.

Heart size in extant elasmobranchs shows considerable variation (*10*, *11*). The size of the placoderm heart (*2*), as preserved, is comparable to the heart of extant sharks [supplementary text in (*2*)] and certainly sufficient for a bottom-dwelling fish. Furthermore, as other organs (stomach, liver) show clear signs of shrinkage [figure S3 in (*2*)], the preserved size of the heart is only a minimum; it may well have been larger in life.

As Jensen *et al*. note (*1*), the hearts of all living vertebrates have at least two chambers, and phylogenetic bracketing strongly implies that this was also the case in placoderms. In the Gogo material, the chambers as preserved are confluent but can be distinguished as discrete shapes. We are thus left with three possibilities: 1) that the fossilised structure is an originally two-chambered heart but decay has caused the connection between the chambers to widen; or 2) that it is a single-chambered heart, a unique structure among vertebra[illegible] or 3) that it isn't a heart at all but anot[illegible] unique organ occupying the place where a gnathostome heart would normally sit. Possibilities 2 and 3 can both be excluded because they are implausible on grounds of comparative anatomy and functional morphology, whereas 1 is plausible from anatomical, functional, and taphonomic perspectives. We conclude that our interpretation of the structure as a two-chambered heart is correct.

We have used the correct official acronym MV-P when referring to specimens from the palaeontology collections housed in the Museums of Victoria (Fig. 1). In response to a reviewers request to provide a clearer image of the spiral gut, a photo of MV P229694 was used to replace one of WAM 09.4.1 in figure 2G (*2*), but unfortunately we omitted to update the specimen number and scale in the figure caption.

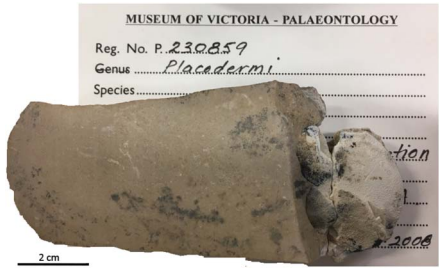


Fig. 1. Arthrodire specimen with accession card for the Museum of Victoria Palaeontology Collection (MV-P).

REFERENCES AND NOTES

1. B. Jensen, A. F. M. Moorman, T. Wang, P. Rask Møller, J. M. Icardo, H. Lauridsen, Comment on "Exceptional preservation of organs in Devonian placoderms from the Gogo lagerstätte". (Science, 2023); 10.1126/science.adg2748.
2. K. Trinajstic *et al*., *Science* **377**, 1311–1314 (2022).
3. P. Janvier, *Early Vertebrates* (Oxford Univ. Press, 1996).
4. R. S. Sansom, S. E. Gabbott, M. S. Purnell, *Palaeontology* **56**, 457–474 (2013).
5. P. C. J. Donoghue, M. A. Purnell, *BioEssays* **31**, 178–189 (2009).
6. L. Schnetz, R. J. Butler, M. I. Coates, I. J. Sansom, *Palaeontology* **65**, e12616 (2022).
7. P. R. Wilby, D. E. G. Briggs, *Geobios* **30**, 493–502 (1997).
8. R. G. Davidson, N. H. Trewin, *Scott. J. Geol.* **41**, 129–134 (2005).
9. L. Maldanis *et al*., *elife* **5**, e14698 (2016).
10. M. E. Larsen, D. C. Abel, D. P. Crane, R. D. Grubbs, *Mar. Biol.* **167**, 1–8 (2020).
11. J. M. Grim, A. A. Ding, W. A. Bennett, *Fish Physiol. Biochem.* **38**, 1409–1417 (2012).

ACKNOWLEDGMENTS

This work was funded by Australian Research Council grants DP1092870 and ARC-DP140104161 (to J.A.L. and K.T.), DP1096002 (to C.A.B. and P.D.C.), ARC-DP110101127 (to K.T.), and ARC-DP200103219 (to P.D.C. and K.T.); European Research Council Advanced Investigator grant 233111 (to P.E.A., V.D., and S.S.); the Knut and Alice Wallenberg Foundation (to P.E.A.); European Synchrotron proposal EC-203 (to P.E.A., S.S., and K.T.); and Australian Neutron proposal P5203 (to A.M.C., J.A.L., and K.T.). **Competing Interests Statement:** Authors declare no competing interests.

Submitted 2 January 2023; accepted 28 February 2023
10.1126/science.adg3748

[1]School of Molecular and Life Sciences, Curtin University, Kent Street, Bentley, 6102, WA, Australia. [2]Western Australian Museum, 49 Kew Street, Welshpool, 6106, Western Australia, Australia. [3]College of Science and Engineering, Flinders University, GPO Box 2100, Adelaide 5001, South Australia, Australia. [4]Museum Victoria, PO Box 666, Melbourne, Victoria, Australia 3001. [5]Department of Organismal Biology, Evolutionary Biology Center, Uppsala University, Norbyvägen 18A, 75236 Uppsala, Sweden. [6]European Synchrotron Radiation Facility, 71 Avenue des Martyrs, 38000 Grenoble, France. [7]Australian Regenerative Medicine Institute and EMBL Australia Building, Monash University, Clayton, 3800, Victoria, Australia. [8]Australian Nuclear Science and Technology Organisation (ANSTO), Lucas Heights, 2234, New South Wales, Australia. [9]Earth Sciences Section, South Australian Museum, Adelaide, 5000, South Australia.
***Corresponding author. Email: K.Trinajstic@curtin.edu.au**

By **Alex Hubert**

Space to meditate

I lay on the bench in the common area, put soothing music on my noise-canceling headphones, and closed my eyes. It was my usual routine—until I was interrupted by someone repeating "excuse me" with growing ferocity. It was a professor. I politely asked what the problem was. "What you are doing is incredibly unprofessional and highly inappropriate," they said sternly. "This is a *research* area." "I'm sorry," I replied hesitantly, "I am meditating." Without a pause, they retorted, "Well, if that is the case you should go somewhere quiet to do so." Their venom stunned me into silence. I apologized again and slunk out of the room.

I had taught myself meditation a year earlier, during the first year of my physics Ph.D. when I was struggling with severe anxiety. I knew I needed to get my symptoms under control, but taking action felt daunting. Meditation seemed a relatively easy option: no special equipment, no one else's schedule to work around, all I had to do was close my eyes for a few minutes a day. I wasn't confident it would make a difference. But my physics education taught me that the greater my skepticism, the more thoroughly I needed to challenge it. So, I decided to give meditation a try.

Every day, I would find 10 minutes to lie down, shut my eyes, and attempt to focus only on the rhythm of my breathing. At first it felt like I was just closing my eyes. But 3 months in, I started to experience short stretches of calming disconnection from the outside world. My anxiety symptoms decreased. And after another 3 months, it was as if the last cog fell into place. When I closed my eyes, I quickly found myself in a quiet internal space that lasted for the full 10 minutes and left me rejuvenated. From then on, I knew meditation would be part of my daily routine for a long time.

But that common area was the only place near my department where I could practice. Without it, how could I maintain my regimen?

I hoped the professor had simply not understood why I was meditating. If I explained, maybe they would soften and I would be able to resume. I knocked on their office door; the scowl was still there. I apologized again and tried to explain how important meditation was for managing my anxiety. They repeated I should go somewhere else, then shut the door in my face.

In tears of frustration, I went to a department administrator who had always been kind and helpful. She was keen to help me find a space to meditate, but the only available quiet room was locked when it wasn't being used. I had to borrow the technician's key whenever I wanted to meditate. I tried that for a couple of weeks, but it simply wasn't practical.



"Mindful breaks from academic life ... helped me actually enjoy most of my Ph.D."

Next, I went to an academic in my department who I knew was sympathetic to mental health issues. They promised to help. After a word with the purchasing manager, a new bench was bought and placed on the floor where I worked. If there was a problem, the academic told me, I was to refer any objectors to them.

For the remainder of my time working toward my degree, that meditation spot allowed me to take mindful breaks from academic life, which helped me actually enjoy most of my Ph.D. After each meditation session, I returned to my research refreshed. And when I hit a hurdle, in 1 minute flat I could shut my eyes, close myself off from everything around me, and find a more objective view. When I received a chapter of my thesis covered in red ink, for example, I once might have wallowed in an emotional response to what felt like a personal attack. Instead, through multiple meditations I was able to process my feelings and consider my supervisor's intentions and the faults in my writing he'd helped correct.

I was lucky to have colleagues who helped me. This is not the norm, however. Even as academics begin to have more open conversations about the importance of taking care of our mental health, harsh unwelcoming perspectives like that of the professor who kicked me out of the common room are all too common. It shouldn't be so hard to find—or make—a space for mental health. ■

Alex Hubert holds a Ph.D. from the University of Warwick. Send your career story to SciCareerEditor@aaas.org.